Personalized POI Groups Recommendation in Location-based social Networks

YU Fei, LI Zhijun, JIANG Shouxu, YANG XiaoFei

School of Compute Science and Technology,
Harbin Institute of Technology. 150001 Harbin, China
{yf,lizhijun_os,jsx}@hit.edu.cn
xiaofei.hitsz@gmail.com

Abstract. With development of urban modernization, there are a large number of hop spots covering the entire city, defined as Pionts-of-Interest (POIs) Group consist of POIs. POI Groups have a significant impact on people's lives and urban planning. Every person has her/his own personalized POI Groups (PPGs) based on preferences and friendship in location-based social networks (LBSNs). However, there are almost no researches on this aspect in recommendation systems. This paper proposes a novel PPGs Recommendation algorithm, and models the PPGs by expanding the model of DBSCAN. Our model considers the degree to each PPG covering the target users' POI preferences. The system recommends the target user with the PPGs which have the top-N largest scores, and it is one NP-hard problem. This paper proposes the greedy algorithm to solve it. Extensive experiments on the two LBSN datasets illustrate the effectiveness of our proposed algorithm.

Keywords: POI group recommendation · Personalization · Geo-Social Distance · Density-based clustering

1 Introduction

With the development of mobile community, persons' demands on location-based services become ever more. Wherein Point-of-Interest (POI) recommendation is one typical application for location-based social networks (LBSNs). Users check in many different POIs in LBSNs, and these checking-in historical trajectories can represent users' checking-in behaviour features and POI preferences. Meanwhile, with the rapid development of the modern city, a large number of hop spots cover the total city space. Each city has its own characteristics of urban culture, and generally they are represented by frequently checked-in points-of-interest (POIs). The hop spots recommendation is that the system recommends the tourists with these frequently checked-in points-of-interest (POIs), however, the tourists' POI preferences are different from that of the city. So, the existing urban hop spot recommendation cannot recommend users with locations which are in accordance with users' POI category preferences. Currently, POI recommendation methods mainly focused on the recommendation quality, that are the accuracy and recall ratio of the recommended POI list. Since the recommended list cannot cover the total city, and these POIs don't own the whole urban characters. So, existing POI recommendation methods[**?**][**?**][**?**] cannot provide the tourists with the city's hop spots.

In this paper, we mainly research the hop spots with the target user's POI category preferences. Since these hop spots consist of many POIs covering the city, our goal is to recommend users with POI groups sufficiently covering their POI preferences. Then we call these POI groups as the personalized POI groups. There are four steps in our researches: First, we propose the definition of POI groups and formula POI groups. Second, this paper studies the intra-cluster correlation in each POI group, and models personalized POI groups (PPGs). Third, we propose the measure of the degree to each POI group covering users' POI category preferences. Finally, this system recommends the target user with the top-K PPGs ranking of the degree of POI category preference coverage.

This paper proposes one novel recommendation problem: Personalized POI group recommendation problem. Our research can solve the following question: which hot spots in one city we may interest based on my own POI preferences? Which regions of a city my friends check in, and these regions are places I may be interested in

2 Problem Definition

In this section, we will give a formal definition of Personalized POI Group recommendation problem (**PPG-Rec**) in LBSN, and this paper formulates this problem.

Definition 1. (LBSN) An LBSN $< G, C >$ consists of a social network $G = < U, E >$, where U is the users set, E is the set of edges, and check-in records $C = \{(u, l, t)\}$, (u, l, t) represents one check-in record where user u checks in the location l at time t. A location l denotes: $l = (lon, lat, a)$, wherein lon is longitude, lat is latitude, a is one POI category.

Definition 2. (PPG-Rec) Given an LBSN $< G, C >$, a target user u_T and his/her friend set U_F, given check-in records (POI set) $POI = \{I_1, I_2, ...\}$, each item I_i is a triple in the form of $< lon, lat, time, I_i.a >$, wherein lon, lat and $time$ respectively denotes longitude, latitude and check-in time, and $I_i.a$ represents the POI I_i's category. Personalized POI group recommendation problem is to select the set of clustering $\mathcal{C}^* = \{C_1^*, C_2^*, ...\}$, wherein C_i^* is a set of POIs which satisfies u_T's preference demand and represents the semantics typical semantics and geography features corresponding to the clustering C_i^*.

3 PPGs Recommendation

In this section, we firstly model POI groups (**PG**). This paper extends the model of the DBSCAN [**?**] for modeling PG. For each POI I_i in the given LBSN, PG finds the geo-soical distance ϵ-neighborhood $N_\epsilon(I_i)$ of I_i, which includes all POIs I_j such that $p(I_i, I_j) \leq \epsilon$. If ϵ-neighborhood of I_i contains at least $MinPts$ POIs, then I_i is a *core* POI; in this case, I_i and all POIs in ϵ-neighborhood should belong to a POI group $C(I_i)$. If another core I_j belongs to $C(I_i)$, then $C(I_i) = C(I_j)$, in one words, the two POI group are merged. When the all core POIs are identified and merged to the corresponding POI groups (PG), PG ends up with a set of POI groups and a set of outliers. These outliers are the POIs who cannot belong to these POI groups, as Fig. 1 shown.

3.1 Modeling POI Groups

In this paper, we describe POI group based geo-social distance between locations. Wherein the geo-social distance is denoted as $D_{GS}(I_i, I_j)$, and it merges the geography distance $D_G(I_i, I_j)$ and $D_S(I_i, I_j)$. $D_{GS}(I_i, I_j)$ denotes as the following equation:

$$D_{GS}(I_i, I_j) = \lambda \cdot D_G(I_i, I_j) + (1 - \lambda) \cdot D_S(I_i, I_j) \tag{1}$$

The parameter $\lambda \in (0, 1)$ is the tradeoff of the geography distance and social distance, and it depends on the user's personal interests.

In this paper, the geography and social distance between POIs I_i, I_j respectively denotes as the followings:

Definition 3. (Geography Distance) Given LBSN $< G, C >$, two POIs I_i, I_j in given city, the geography distance between I_i and I_j is defined as the normalized Euclidean distance:

$$D_G(I_i, I_j) = \frac{E(I_i, I_j)}{maxD} \tag{2}$$

where $E(I_i, I_j)$is Euclidean distance, $D_G(I_i, I_j) \in [0, 1]$, in this paper, the researched users set consists of the target user and his/her friends $\{u_t \cup U_F(u_T)\}$. The social distance $D_S(I_i, I_j)$ between POIs I_i, I_j naturally depends on the social network relationships between the set $U(I_i)$ and $U(I_j)$ of users who checked in I_i, I_j, respectively. Our social distance $D_S(I_i, I_j)$ is based on the set of contributing users $CU(I_i, I_j)$ between POIs I_i and I_j.

Definition 4. (Contributing Users) Given two POIs I_i, I_j with checking in users set (concluding the target user u_T and $U_F(u_T)$), the set of contributing users $CU(I_i, I_j)$ for the POI pair (I_i, I_j) is defined as the following:

$$\begin{aligned} CU(I_i, I_j) = \{&u_a \in U(I_i) | u_a \in U(I_j) \ or \ u_b \in U(I_j), u_a, \\ &u_b \in \{u_T \cup U_F(u_T)\}\} \cup \{u_a \in U(I_j) | u_a \in U(I_i) \\ &or \ u_b \in U(I_i), u_a, u_b \in \{u_T \cup U_F(u_T)\}\} \end{aligned} \tag{3}$$

Definition 5. (Social Distance) Given LBSN $< G, C >$, two POIs I_i, I_j with visiting users U_{I_i}, U_{I_j}, the social distance between I_i and I_j is defined as:

$$D_S(I_i, I_j) = 1 - \frac{|CU(I_i, I_j)|}{|U(I_i) \cup U(I_j)|} \tag{4}$$

3.2 Intra-PG correlation analysis

In LBSN, each user has oneself own POI preference distributions. Generally, the distribution is the power-law distribution[**?**]. This paper focuses on the personalized POI groups (**PPGs**) recommendation, and the personal features are the target user u_T and his/her friends $U_F(u_T)$'s checking-in behaviours preferences. So, we synthesize u_T and $U_F(u_T)$'s POI preferences' features to formula the personalized POI groups (**PPGs**). In this paper, the system recommends the target user u_T with the PPGs, which cover the target user u_T's POI preferences as far as possible.

Preference Coverage Preference coverage considers how the POI group $C_i \in \mathcal{C}$ in Section 3.1, concludes different POI categories. In fact, each POI category may label many POIs, and one POI may be labelled with multiple POI categories. In this paper, we analyse the degree to the every POI group concluding the target user's POI preferences, then select the subset of POIs C_k^* in the POI group C_k instead of all POIs in the POI group. The larger degree to which C_k^* covering u_T's POI preferences $A_{u_T} = \{a_1, a_2, ..., a_{M_{u_T}}\}$, the more information C_k^* can provide for u_T. In this paper, we consider the intra-PG correlation based on the preference coverage.

Let $A_{u_T} = \{a_1, a_2, ..., a_{M_{u_T}}\}$ is the target user u_T's POI categories preferences, and this paper computes the preference coverage of POI group in Equation 5.

$$PreC(C) = \frac{1}{|C|} \sum_{a_k \in A_{u_T}} cov_{a_k}(C) \tag{5}$$

where $cov_{a_k}(C)$ measures the degree to a_k is covered by at least one POI in C. This section formulas $cov_{a_k}(C)$ with the following equation:

$$cov_{a_k}(C) = 1 - \prod_{I_i \in C} [1 - cov_{a_k}(I_i)] \tag{6}$$

where $cov_{a_k}(I_i)$ represents the degree to POI I_i covers a_k. The popular degree to the POI I_i labelled by the POI category a_k is described by the number of checking-in I_i with the label a_k. The popular degree $cov_{a_k}(I_i)$ denotes: $cov_{a_k}(I_i) = \frac{Num(I_i,a_k)}{\sum_{I_j \in C} Num(I_j,a_k)}$, $Num(I_i, a_k)$ represents the number of users checking-in the POI I_i with the label a_k, $\sum_{I_j \in C} Num(I_j, a_k)$ is the number of users checking-in the POIs with the label a_k in the POI group C.

Modeling Personalized POI Group In this section, we select a subset of POIs C^* in each POI group C, the POIs in the subset of POIs C^* cover the target user u_T as far as possible, as Fig. 4. This paper expects to optimize the subset of POIs. Given the target user and his/her friends $\{u_T \cup U_F(u_T)\}$, the POI groups $\mathcal{C} = \{C_1, C_2, \cdots\}$ from Section 4.1, the problem is to select a subset of POIs in each POI group $C_i \in \mathcal{C}$ that maximizing the preference coverage function $PreC(C^*)$ in the check-in data. We regard the subset of POIs C^* as the personalized POI group (PPG), and this problem can be called as PPG-Rec. Meanwhile, we model this problem as one multi-objective optimization problem:

$$\begin{aligned} \textbf{Given } &: \mathcal{C}, K, \{u_T \cup U_F(u_T)\} \\ \textbf{Objective Function } &: \max_{C^* \subset C} PreC(C^*) \\ & s.t. \quad |C^*| = K. \end{aligned}$$

Greedy Algorithm: Due to the objective function $PreC(C^*)$'s monotone and submodular property[**?**], we give a greedy algorithm to compute the problem. The algorithm is called as Select_PPGs algorithm, and its detail description is as the following Algorithm 1.

Algorithm 1 Select_PPGs algorithm

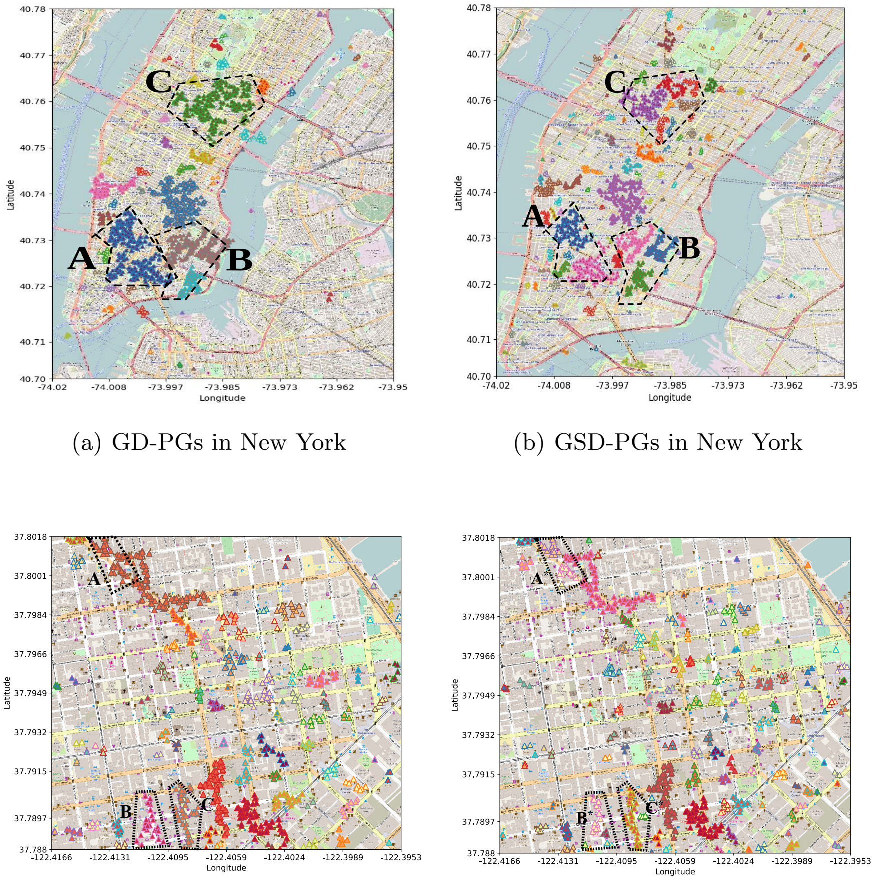
Input: POI Groups $\mathcal{C} = \{C_1, C_2, \cdots\}$, the target user u_T and $U_F(u_T)$, A_{u_T}, K.
Output: A subset of POIs $C^* \subseteq C$, $|C^*| = K$, the value $PreC(C^*)$.
1: Initialize $C^* \Leftarrow \phi$
2: **for** $j = 1$ **to** K **do**
3: $\quad I_j \leftarrow \arg\max_{I_j \in C}[PreC(C^* \cup I_j) - PreC(C^*)]$;
4: $\quad C^* \leftarrow C^* \cup I_j$
5: **return** C^*, $PreC(C^*)$.

Top-N PPGs Recommendation Algorithm This paper is in order to find some personalized POI groups $\{C_1^*, C_2^*, \cdots, C_N^*\}$ to recommend the target user u_T in LBSN. Comparing with the value $PreC(C_i^*), C_i^* \subseteq C_i$ in POI groups $\mathcal{C}$, we select the top-N personalized POI groups $C_{(i)}^*, (i = 1, 2, ..., N)$ as the recommended PPGs.

4 Experimental Evaluation

Dataset Description This paper utilizes the two real LBSN datasets, such as Foursquare, Gowalla datasets. They respectively consist of 36,907 users, 4,163 users; 26,907 locations, 121,142 locations; 1048,575 check-in times, 483.813 check-in times; the time span: 4/14/2010–1/17/2011, 1/18/2010–8/11/2011; friendship pairs: 23,148 pairs, 32,512 pairs; POI categories: 6,636 categories, 7,835 categories.

Comparative Approaches To illustrate the effectiveness of our method, we compare GSD-PPG against anther several methods. GD-based POI group **(GD-PGs)**



(a) GD-PGs in New York

(b) GSD-PGs in New York

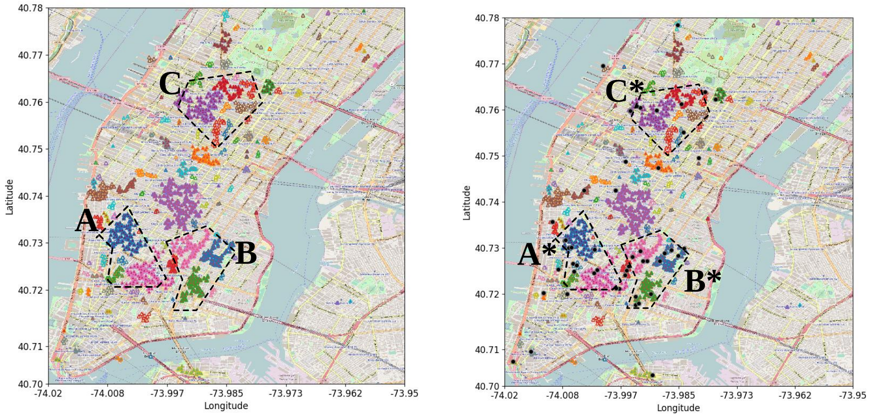
(c) GD-PGs in San Francisco

(d) GSD-PGs in San Francisco

Fig. 1. POI Groups of Check-in Datasets found in two Cities

recommendation method [?] utilizes the DBSCAN to model the POI group. This POI groups represent the space clustering only, and it is without considering POI category preference and social relationship information. GSD-based POI group**(GSD-PGs)** recommendation method[?] takes advantage of the DCPGS, and denotes the social and geography information. Due to the size of the PG is big, this situation leads to the smaller value of the personalization in each PG. Region-based POI group **(Region-PGs)** recommendation method [?] is in order to recommendation the target user with the POIs in a given query region. This recommendation and our approach are not from the same perspective. So, we do not compare this method and our method in this paper's experiment part.

Evaluation Metrics This paper utilizes the objective function is effective and reasonable, and the recommended Personalized POI Groups (PPGs) based on the function $PreC(\cdot)$ can describe the personalization of the PPGs. We utilize the recall ratio: Recall@K [?][?] to evaluate the quality of POI groups recommendation, and it is important to find out how many recommended POI categories actually belong to the target user POI category preferences,$Recall@K = \frac{|RC_{re}(\bigcup_{i=1}^{K} C^*_{(i)}) \cap A_{u_T}|}{|A_{u_T}|}$. Wherein A_{u_T} are user u_T's POI category preferences set. These metrics for the entire POI



(a) POI Groups in New York (b) Personalized POI Group in New York

Fig. 2. POI Groups and Personalized POI Groups. In Fig.**??**(b),the black dots are POI categories which belong to the target user's POI preferences

groups recommendation system are computed by averaging the above two metrics value for 2000 users (as the target users u_T) respectively.

Experimental Results Visualization-based Analysis Fig.**??**(a) shows the POI groups are spacial clusterings, but these POI goups disregards the social network behind POIs. These POI groups can not answer the above mentioned application issues. Fig.**??**(b) displays the POI groups can represent how close POIs are in the aspect of the spacial and social distance. By tuning the two parameters of spacial clustering method, this paper cannot find these POI groups found by GSD-based clustering method. As Fig.**??**(a)(b) shown, the POIs in the region **A** belong to one POI group. And its corresponding area is the region $\mathbf{A}^*$ in Fig.**??**(b), there are two POI groups in this region. Which reason is that these POIs in this region $\mathbf{A}^*$ are partitioned based on the social relationship. From Fig.**??**(c)(d), we can see the comparison results between two clustering methods is similar to that presented in Fig.**??**(a)(b).

Personalization Analysis for PG Each POI group has its own POI features for maximizing covering the u_T POI preferences. In this paper, we regard these POIs with the maximization score $PreC(C^*)$ in each POI group as the POI features in the POI group. Then the Personalized POI group denotes these POIs, $C^* = \{I_{(1)}, I_{(2)}, \cdots, I_{(k)}\}$, $C^* \subseteq C$. From Fig.**??**(b), we observe that there is black dots distribution in each POI group. These black dots are the features in POI groups. Some POI groups have many black dots, and some POI groups have less black dots.

As Fig.**??**(b) shown, there are some POIs in each POI group, and these POIs (as the black dots in the figure) belong to the target user's POI preferences, called as the personalized POI (**PP**) in this paper. From Fig.**??**(b), we see that some POI groups with many POIs have less personalized POIs. Hence, the size of POI groups do not determine the number of PPs in each POI group. In this section, we utilize the function $PreC(C^*) = \frac{1}{|C^*|}\sum_{a_k \in A_{u_T}} cov_{a_k}(C^*)$ to measure the degree to the personalization in each PPGs. Our recommend method is to recommend the target user with the top-N PPGs according to the personalization degree of PPGs. This method outperforms other competitor methods significantly as Fig.**??** shown.

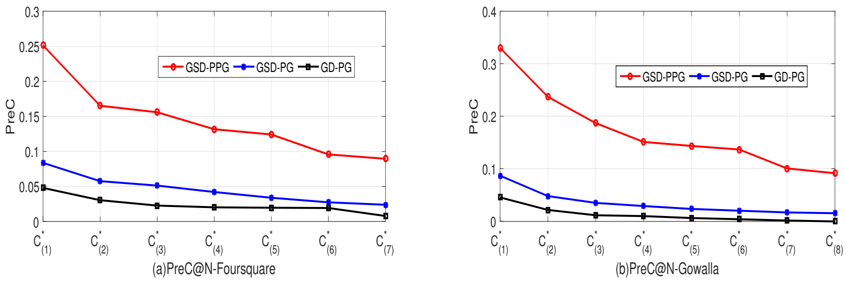


Fig. 3. Comparison of PreC of three mehtods

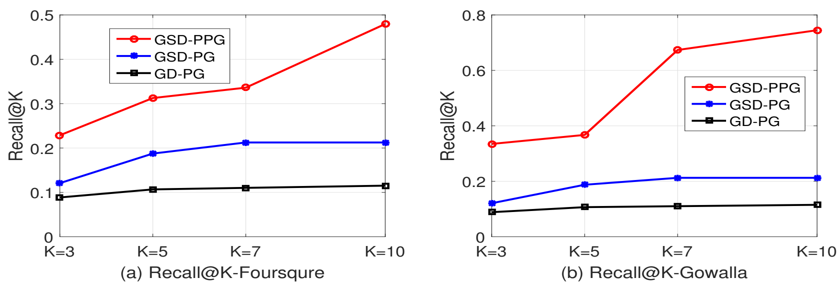


Fig. 4. Recall@K of Different Methods

Effectiveness of Methods This paper uses the Recall@K to evaluate the effectiveness between our method and other methods. Figure **??** reports the performance of our personalized POI groups recommendation method on Foursquare and Gowalla datasets. We show only the performance where K in the range $[3, 5, 7, 10]$. The recall of the three approaches (GSD-PPG, GSD-PG, GD-PG) are reported in Fig. **??**. Observe from Fig. **??**, GSD-PPG method achieves better recommendation accuracy than the other method. Due to the big size of each POI group, GSD-PG method provides the recommended POI groups with less personalization score than GSD-PPG. Without considering social distance and POI categories preferences, GD-PG method mainly focuses on the compactness of space between locations. And due to the big amount of POIs in each POI group, the recall ratio of this method is not as good as the above two methods in terms of recommendation precision.

5 Conclusions

In this paper, we propose a novel recommendation problem: personalized POI groups recommendation problem. The most important contribution of our work is to provide a definition of Personalized POI groups recommendation problem, and to formula the POI group with the geo-social clustering method Then this paper extracts the personalized features in each POI group, these features (PPG) consist of POIs in each POI group and replace the group's personalization. We design a metric to measure the degree to personalization of each POI group/ personalized POI group. To solve this new problem, we propose a greedy algorithm with $(1 - \frac{1}{e})$ theoretical bound. The enormous scale of LBSNs dataset verifies the degree to personalization and the effectiveness of our method. Especially, our approach shows the significant advantage in the degree to personalization. As the future work, we plan to utilize multi-source information (such as the temporal information), and model the periodic of check-in behaviours for improving the efficiency of the recommendation.

Acknowledgments

This work was supported in part by the National Science Foundation grants NSF-61672196, NSF-61370214, NSF-61300210.

References

1. Yu F, Che N, Li Z, et al. Friend Recommendation Considering Preference Coverage in Location-Based Social Networks[C]Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2017: 91-105.
2. Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. Kdd. 1996, 96(34): 226-231.
3. Shi J, Mamoulis N, Wu D, et al. Density-based place clustering in geo-social networks[C]. Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014: 99-110.
4. Li J P, Xu Y, Zhao L. OPGs-Rec: Organized-POI-Groups Based Recommendation[C]Asia-Pacific Web Conference. Springer International Publishing, 2016: 521-524.
5. Li Y, Wu D, Xu J, et al. Spatial-aware interest group queries in location-based social networks[C]ACM International Conference on Information and Knowledge Management. ACM, 2012:2643-2646.
6. Wang X, Donaldson R, Nell C, et al. Recommending groups to users using user-group engagement and time-dependent matrix factorization[C]Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016:1331-1337.
7. Ye M, Yin P, Lee W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]. Proceeding of the, International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July. 2011:325-334.
8. Cheng C, Yang H, King I, et al. Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks[C]. Aaai. 2012, 12: 17-23.