Exploring Geographic Information Effects for POI Recommendation in LBSNs

Shuang Liu¹, Wenguang Zheng¹* and Yingyuan Xiao¹
¹School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, 300384, China
*Corresponding author’s e-mail: wenguangz@tjut.edu.cn

Abstract. With the rapid development of global positioning technology, location-based social networks (LBSNs) have emerged. Massive check-in data provides development possibilities for mining user behavior preferences, and also brings a variety of location-based social services. For example, Point of Interest (POI) recommendation service can help users find their desired places. In this paper, we propose a novel POI recommendation model by integrating users’ preference and locations geographic information such that an efficient and accurate POI recommendation can be achieved. The model explores the spatial distribution of check-in points visited, the relevance of the category of check-in points visited, and the difference in the frequency of check-in points visited to generate a recommendation list for a particular user. Experimental results on the New York and Tokyo datasets from Foursquare show that this model has higher recommendation performance than other comparison models.

1. Introduction
With the development of information technology and the Internet, people have gradually entered the era of information overload from the era of lack of information. In order to solve the problem of information overload, the recommendation system came into being. Among them, Location based Social Networks (LBSNs) is considered to have broad prospects. LBSNs encourage users to share geographic locations in real time through check-ins. These massive check-in data provide opportunities for mining user preferences, which is conducive to providing location-based services, such as point of interest (POI) recommendations. Currently based on LBSNs, point of interest recommendation service is the focus of most scholars and researchers. It is also a hot issue in social networks [1,2]. However, it is very complicated to achieve accurate POI recommendations for the reason that high-dimensional data need to be explored (i.e., user-geographical location-category of POI-visiting frequency). Due to the addition of contextual information, such as the geographic coordinates of points of interest, the classification information of points of interest, and the user's interest preferences and other factors. How to use multiple types of contextual information to improve the recommendation accuracy of points of interest in LBSN is also a challenging task.

In this paper, we design a novel and effective personalized POI recommendation system. The main contribution of this paper can be summarized as follows:

- This paper proposes a new mechanism to estimate geographic correlation between different categories of POIs, such that more accurate recommendations can be achieved.
- This paper proposes a method to determine the activity areas of the users and combines with the geographical information to make personalized POI recommendations for users.
We conduct several experiments on New York dataset and Tokyo dataset from Foursquare, demonstrating the improvement of the proposed method in the accuracy of POI recommendation compared to three state-of-the-art approaches.

2. Related work
POI recommendation as one of the crucial tasks in LBSNs has become a popular research direction and has been widely concerned by the academic community [3]. Since the purpose of this article is to make reasonable location recommendations for users based on their check-in, we will start from the perspectives of space and user preferences and elaborate on the impact of different feature factors on location recommendations.

From spatial influence perspective, location recommendation tries to estimate the spatial influence of each check-in point to make POI recommendations to users. For example, Ye et al. [4] used the geographic influence to make points of interest recommendation, and used a power-law distribution to model clustering in space. In order to characterize the impact of spatial features on user activity preference, Ding et al. [5] proposed the notion of Personal Functional Region (PFR) to model user spatial activity preference. Zhang et al. [6] believed that the geographic impact of a user’s check-in should be personalized rather than the same distribution, so the IGSLR method used kernel density estimation (KDE), which used the personalized distance distribution of each user to model geographic distribution. Cheng et al. [7] modelled users’ check-ins via centre-based Gaussian distribution to capture users’ movement patterns based on users’ movements centre.

From user preference perspective, most research efforts focused on location recommendation. Location recommendation tries to suggest POIs to users by estimating user preference on individual POI [5]. For example, Yang et al. [8] estimated user preference on location using the hybrid preference extracted from user check-ins and text-based tips using statistics and sentiment analysis techniques. Feng et al. [9] develops a PRME-G model, which integrates sequential information, individual preference, and geographical influence, to enhance the recommendation performance. Ying et al. [10] proposed a set of features related to social factor, individual preference, and location popularity, and utilized a regression tree model to recommend POIs.

3. POI Recommendation Model Using Geographic Information

3.1. Geographic correlation modeling
Location category information as context information can effectively alleviate the problem of low recommendation accuracy. However, most traditional POI recommendation systems ignore the correlation between different categories, and separately estimate the users’ preference of a certain category only by investigating the check-in frequency on this category. Actually, the correlation between different categories can be reflected by their geographic locations due to the aggregation effect. For example, there are many supermarkets and entertainment venues near residential areas and users may visit supermarkets after visiting some nearby entertainment venues. In this paper, we first explore the geographic correlation between different categories, then calculate the popularity of each POI of a certain category, and finally integrate category correlation information and the popularity of each POI to obtain scores of user preference for each POI.

3.1.1. Category relevancy calculation. The distribution of geographic locations of each POIs reflects the correlation between different categories. Therefore, we first divide the whole check-in area into small grids (here 5×5 is used as an example, as shown in Figure 1), and then the number of locations of each category in each grid is counted. Finally, the category relevancy is calculated, as shown in Formula (1):

\[
rel(c_i, c_j) = \frac{\sum_{i \in Area_i} \cap \sum_{j \in Area_j}}{\sqrt{\sum_{i \in Area_i} q_i^2 \sum\sum_{a \in Area_j} q_a^2}}
\]
Where \( r_{el}(c_i, c_j) \) represents the correlation between category \( c_i \) and category \( c_j \), and \( Area_{c_i} \) is area matrix of category \( c_i \). Meanwhile, \( q \) is the number of locations belonging to category \( c_i \) in a small grid and \( a \) is the number of locations belonging to category \( c_j \) in a small grid.

After obtaining the category relevancy, according to user’s historical check-in records, we calculate the target user’s preference score \( s(u_i, c_j) \), as shown in formula (2):

\[
s(u_i, c_j) = \frac{\sum_{p=0}^{k} rel(c_p, c_j) freq_{i,p}}{\sum_{p=0}^{k} rel(c_p, c_j)} + \delta
\]  

(2)

Where \( k \) is the total number of location categories, and \( freq_{i,p} \) represents the frequency of the category \( c_p \) that the target user \( u_i \) visited. \( \delta \) is the user’s initial preference score value for the category, which effectively avoids the situation where \( s(u_i, c_j) = 0 \). In this paper, \( \delta \) is set to 0.000005.

3.1.2. Popularity of location calculation. After obtaining the user’s preference score for a certain category, we calculate the popularity of each POI to estimate the differences between POIs under the same category by using formula (3).

\[
\text{popular}(c_i, l_j) = \frac{1}{2} \left( \frac{\text{Num}_j}{\text{MAX}(\text{Num})} + \frac{\text{num}_j}{\text{MAX}(\text{num})} \right)
\]  

(3)

Where \( \text{popular}(c_i, l_j) \) is the popularity of the POI \( l_j \) belonging to the category \( c_i \). \( \text{Num}_j \) is the number of times all users visited POI \( l_j \) and \( \text{MAX}(\text{Num}) \) is the highest check-in frequency among all check-in locations; in the same way, \( \text{num}_j \) is the number of users who visited POI \( l_j \) and \( \text{MAX}(\text{num}) \) is the total number of users with the most visited POIs.

Finally, based on the combination of category correlation information and the popularity of each POI, we calculate user \( u_i \)’s score \( \text{Score}_p(u_i, l_j) \) for POI \( l_j \), as shown in formula (4):

\[
\text{score}_p(u_i, l_j) = s(u_i, c_x)\text{popular}(c_x, l_j)
\]  

(4)

Where \( s(u_i, c_x) \) is user \( u_i \)’s score for category \( c_x \).

3.2. User preference modeling

Generally, users have their own specific activity areas and such areas can be investigated to explore preference of POIs of users. Therefore, we construct a user’s personalized geographic distribution to obtain user’s POI preferences. In this paper, we first determine the activity areas for a particular user, then calculate visiting probability of each POIs in such areas, and finally the preference score of POIs for the user is calculated.

Given a target user \( u_i \), we construct a set \( L_i \) to store all the POIs visited by \( u_i \), and a set \( N_i \) to store the corresponding check-in frequencies of there POIs. Then, we take the POI with the highest check-in frequency in \( N_i \) as a center to draw a circle. Such circle is an activity area of the user and the number of POIs in the circle will not larger than the percentage \( d \) of the total number of the POIs visited by \( u_i \), such that the radius of the circle is determined. Remove the POIs visited by \( u_i \) in the current active area from \( L_i \) and continue to find the next active area of user \( u_i \). Finally, the active areas of the target user are obtained.
After constructing activity areas of a user, we use the power law distribution [4] to mine the influence of distance. Without loss of generality, user trends to visit near POIs. Thus, the smaller the visiting distance implies the higher visiting probability of the POI. We use calculate the probability of the user transferring to the next location from the current center by formula (5):

$$prob(l_i, l_j) = a \times \text{dist}(l_i, l_j)^b$$  (5)

Where $a$ and $b$ are the parameters of the power law distribution after training. And $\text{dist}(l_i, l_j)$ is the distance between two locations $l_i$ and $l_j$.

Based on the influence of distance and the limitation of active area, this paper uses formula (6) to calculate the user's preference score for each POI:

$$\text{score}_g(u, l) = \begin{cases} \frac{\sum_{p \in UA_{x}} \text{freq}_{lp}}{\sum_{b \in UA_{x}} \text{freq}_{lb}} \cdot \text{prob}(l_p, l_j), & l_j \text{ in } \text{Area}_{i,x} \\ \frac{1}{n} \sum_{p \in \text{Center}_i} \text{prob}(l_p, l_j), & \text{otherwise} \end{cases}$$  (6)

Where $UA_{i,x}$ is the list of check-in locations in the x-th active area of user $u_i$. And $\text{freq}_{lp}$ represents the frequency of user $u_i$ visiting the location $l_p$. Meanwhile, $\text{Area}_{i,x}$ is the x-th active area of user $u_i$, and $\text{Center}_i$ is the center location list of the active area of target user $u_i$. $n$ represents the number of active areas of the target user.

3.3. Model fusion

Multiple factors contribute to the improvement of the recommendation effect, such as location category information, popularity of location, and geographic influence. Finally, we use linear combination to integrate multiple influencing factors. Finally, user $u_i$'s score for the target location $l_j$ can be calculated by formula (7):

$$\text{score}(u, l) = \gamma \text{score}_p(u, l) + (1 - \gamma) \text{score}_g(u, l_j)$$  (7)

Where $\gamma$ is a weight parameter.

4. Experiment

4.1. Dataset description

We use the Foursquare dataset to verify the effectiveness of our proposed method. The Foursquare dataset contains data records of users checking in in two cities, New York and Tokyo. Each record consists of user id, category id, location id, longitude, latitude and other information. At the same time, the New York dataset under Foursquare is composed of 227,428 check-in records generated by 1083 users, while the Tokyo dataset under Foursquare is composed of 573,704 check-in records generated by 2293 users. In order to reduce the influence of noisy data and invalid data on the experimental results, we need to preprocess these two datasets. Finally, we randomly select 80% of the data as the training set, and the remaining data as the test set.

4.2. Evaluation metrics

We use three metrics to measure the recommendation effect of our experiment, such as Precision@K, Recall@K, NDCG@K. At the same time, we set K to {5, 10, 20}.

$$\text{Precision@K} = \frac{1}{|U|} \sum_{u \in U} \frac{|R_{u_i} \cap T_{u_i}|}{|R_{u_i}|}$$  (8)

$$\text{Recall@K} = \frac{1}{|U|} \sum_{u \in U} \frac{|R_{u_i} \cap T_{u_i}|}{|T_{u_i}|}$$  (9)

$$\text{NDCG@K} = \frac{1}{|U|} \sum_{u \in U} \frac{\text{DCG}_{u_i}}{\text{IDCG}_{u_i}}$$  (10)

Where $U$ is list of the user. $R_{u_i}$ and $T_{u_i}$ are the recommended list of $u_i$ and the set of location actually visited by $u_i$ in the test set, respectively. At the same time, $\text{DCG}_{u_i}$ is the cumulative discount gain of $u_i$, and $\text{IDCG}_{u_i}$ is the ideal state of $\text{DCG}_{u_i}$.
4.3. Recommendation performance

In order to evaluate the effectiveness of our proposed POI method, we use the following three POI recommendation methods to compare with the algorithm proposed in this paper. The baseline methods are PFMPD [11], PFMMGM [7] and LMF [12].

Next, we will introduce the recommendation performance of our method under two different datasets. In order to ensure the fairness of the experimental comparison, we need to adjust the parameters of the baseline method to the best. In this paper, in order to ensure the best recommendation effect of the method we propose, we set $d$ to 0.2, and $\gamma$ to 0.8. Figures 2 and 3 show the experimental results of our algorithm on the New York dataset and Tokyo dataset, respectively. Figures 2 and 3 show that our proposed algorithm is superior to the other three baseline methods in the New York dataset and Tokyo dataset. And compared with the best-performing baseline method LMF, in terms of the three metrics of Precision, Recall, and NDCG, our method has an average increase of 47.62%, 35.77%, and 60.65% in the New York dataset, respectively. Similarly, the methods we proposed in the Tokyo dataset improved by 40.8%, 39.08%, and 48.35%, respectively.

![Figure 2. New York (Foursquare).](image)

![Figure 3. Tokyo (Foursquare).](image)

5. Conclusion

This paper focuses on POI recommendation and studies two factors that affect recommendation performance: user preference and geographic information. Meanwhile, a personalized POI recommendation method that integrates user preference and geographic information is proposed. Finally, the feasibility of our method is verified through experiments and the experimental results show that our method has higher superiority compared with the current POI recommendation method.

Acknowledgments

This work was supported by National Nature Science Foundation of China (61702368), Natural Science Foundation of Tianjin, China (18JCQNJC0070).

References

[1] Aliannejadi, M., Rafailidis, D., Crestani, F.: A joint two-phase time-sensitive regularized collaborative ranking model for point of interest recommendation. IEEE Transactions on Knowledge and Data Engineering (2019).

[2] Aliannejadi, M., Mele, I., Crestani, F.: A cross-platform collection for contextual suggestion. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1269–1272. ACM (2017).

[3] X. Jiao, Y. Xiao, W. Zheng, L. Xu, and H. Wu, “Exploring spatial and mobility pattern’s effects for collaborative point-of-interest recommendation.” IEEE Access, vol. 7, pp. 158917–158930, 2019.

[4] Ye M, Yin P, Lee W, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]. international acm sigir conference on research and development in information retrieval, 2011: 325-334.

[5] Y. D, Z. D, and Z. V. W, “Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 45, no. 1, pp. 129–142, 2015.
[6] Zhang Jiadong, Chong C Y. IGSLR: Personalized geo-social location recommendation; A kernel density estimation approach [C]// Proc of 21th ACM SIGSPATIAL. Int Conf on Advances in Geographic Information Systems. New York: ACM, 2013: 334-343.

[7] Cheng, C., Yang, H., King, I., Lyu, M.R.: Fused matrix factorization with geo-graphical and social influence in location-based social networks. In: Twenty-Sixth AAAI Conference on Artificial Intelligence (2012).

[8] D.Yang, D.Zhang, Z.Yu, and Z.Wang, “A sentiment enhanced personalized location recommendation system,” In Proc. HT, 2013, pp. 119–128.

[9] S.Feng, X.Li, Y.Zeng, G.Cong, Y.M. Chee, Q. Yuan, Personalized ranking metric embedding for next new POI recommendation, in: International Conference on Artificial Intelligence, 2015, pp. 2069–2075.

[10] J. Ying, E. Lu, W. Kuo, and V. Tseng. Urban point-of-interest recommendation by mining user check-in behaviors. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, pages 63–70. ACM, 2012.

[11] Rahmani H A, Aliaannejadi M, Ahmadian S, et al. LGLMF: Local Geographical based Logistic Matrix Factorization Model for POI Recommendation[J]. 2019.

[12] Johnson, C.C.: Logistic matrix factorization for implicit feedback data. Advances in Neural Information Processing Systems 27 (2014).