An Adaptive Collaborative Filtering Based Approach for Point-of-Interest Recommendations

Yueyu Wang[1, a], Leyan Chen[1, b], Jiani Chen[1, c*]

[a]wyueyu2022@163.com, [b]chenleyan2022@163.com, [c]jianichen2022@163.com

[1]Meihua school, Suzhou, China

Abstract: LBSNs (Location-based Social Networks) provide abundant information for users to browse and explore the places where they are interested in, named POI (Point-of-Interest). However, such large amount of check-in records in LBSNs cause information overload problem and increase difficulty for users to find the really desire POIs. POI recommendation systems can be employed to solve this problem. Most traditional POI recommendation methods are CF (Collaborative Filtering) based and achieve recommendations for a particular user according to check-in records of his similar users. In this paper, we propose an adaptive CF based algorithm to achieve POI recommendations for users, considering their personalized activity regions. Compare with existing algorithm, our algorithm does not construct user-POI matrix by using the entire historical records. Instead, we first explore users' activity regions and construct more personalized user-POI matrix for each particular user according to corresponding activity regions. Besides, we propose a method to dynamically determine the number of similar users for a certain user, instead of using a fix number for all users, leading to more personalized recommendations. We have implemented our POI recommendation system and compared with state-of-the-art methods by using Foursquare dataset. The experimental results show that our POI recommendation system achieves better performance than all these compared approaches.

Keywords: Activity region, Point-of- Interest (POI), Recommendation system, Collaborative filtering (CF)

1 Introduction

LBSNs (Location-based Social Networks) have becoming increasingly popular, and provide users with a platform to share footprints (check-ins), images and comments. Many LBSN-based applications and services have been developed, which bring convenience for users to find the locations where they are interested in. Such locations are named POI (Point-of-Interest) in LBSN. However, with the rapidly increasing number of historical records in LBSN, it becomes more and more difficulty for users to find the really desired POIs among massive records. Therefore, POI recommendation systems have become necessary for corresponding LBSN-based applications and services to explore users' preferences and generate personalized visiting suggestions for users.

Most traditional POI recommendation systems are developed by using CF (Collaborative Filtering). Given the historical reck-in records provided by a LBSN, a user can be encoded to a

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322634

POI vising vector, and each element of this vector is the vising frequency of the user at a certain POI. Achieving such vectors of all the users, similarity between different users can be measured, and recommendations for a target user can be generated according to the visiting vectors of his similar users. K Neatest Neighbour (KNN) based approaches are usually used to construct recommendation list for a target user. The key idea of KNN-based approach is to select K most similar user of the target user, and construct a recommendation list according to the historical check-ins of these K users. However, these traditional POI recommendation systems suffer from two major drawbacks described as follows:

- High sparsity: the POI visiting vector of a user contains large amount of zero values, which means the user never visited these locations. Such sparse vectors cannot accurately measure similarities between different users;
- Hyperparameter K: considering variety of users, a fixed value K is inappropriate to determine the number of selected similar users for a target user. A flexible K is needed to determine really similar users according personalized check-ins of different users.

In this paper, we consider these two drawbacks and propose an adaptive CF-based POI recommendation algorithm. The major objective of our algorithm is to explore the personalized visiting preferences of users and recommend POIs which they are really interested in. The major contributions of this paper are summarized as follows:

- We propose a method to determine activity regions of users. Instead of considering the entire historical check-in records, we only use the check-in records in regions to generate recommendations for corresponding users;
- We propose a method to dynamically determine the number of similar users for a target user. Thus, more personalized POI recommendations can be achieved;
- We have implemented our algorithm and compared with four existing POI recommendation algorithm by using Foursquare dataset. Our algorithm achieves better performance for both precision and recall.

The rest of paper is organized as follow: section 2 summarizes related works; section 3 describes the adaptive CF based POI recommendation algorithm; section 4 shows the experimental results; section 5 concludes this paper.

2 Related works

POI recommendation is a challenging task and many researchers have proposed corresponding approaches to solve this problem. In this section, we briefly summarize related works.

Most previous methods are CF-based and a key problem of such methods is how to measure the similarity between different users. Ye et al. [1] considered both social factor and geographical factor to measure the similarities between different users, and a unified CF-based framework was proposed to integrate all these factors for POI recommendation. A novel model named GeoSoCa [2] proposed to explore geographical, social and categorical factors. Cheng et al. [3] also considered both geographical and social features to construct recommendations. Gao et al.

[4] pointed out that users usually show different preferences at different time, thus they explored temporal influences for POI recommendation.

Jia-Dong and Chi-Yin [5] pointed out that both temporal and spatial influences should be considered. Thus, a sequential influence which integrates both temporal and spatial factors should be employed to explore preferences of users and measure similarities. Jiao et al. [6] also considered sequential influence, and they constructed travelling trajectories of users and measured the similarities by comparing the trajectories of users.

The performances of these previous methods are unsatisfied. A main reason is that the check-in data is highly sparse. Thus, the similarity measurements between different users are lack of high accuracy, even though the method are relatively complex.

3 Adaptive CF-based poi recommendation approach

In this section, we will describe details of our adaptive CF-based approach for POI recommendation. The major objective of our algorithm is to explore visiting preferences of users and recommend POI which they are likely to visit. To facilitate description of our algorithm, we first introduce the following notations.

- U_N: the set of entire users in the dataset and the total number of users is N;
- L_M: the set of entire locations in the dataset and the total number of locations is M;
- R_i: the activity region of user $u_i \in U$;
- S: the partial user-POI matrix.

Our algorithm consists three major components: activity region determination, partial user-POI matrix completion and adaptive recommendation generation. In the following subsections, we will describe details of these components.

3.1 Activity Region Determination

Exploring check-in records of users, we find that users tend to visit locations in certain regions. Pervious paper (Smith and Yang, 2012) also pointed out that users trend to check in around several centres. In this paper, we named such region as activity region. We repeat the following procedures to determine the activity region for a particular user.

a) Given a target user u_i, we first construct a set Li to store all the locations visited by u_i. Each $l_{ij} \in L_i$ is described as <longitude, latitude, frequency>. Let f_{max} be the largest visiting frequency of locations in Li by u_i.

b) Set a threshold α, and update Li by removing the locations whose visiting frequency is less than $\alpha * f_{max}$;

c) Construct a circle to cover all the POIs in Li. The centre of this circle is determined by calculating average of longitudes and latitudes of all location in Li, respectively. The radius of this circle is the largest distance between the centre and locations in Li.

d) Finally, the construct circle is the activity regions of user u_i, denoted by R_i.

For each user, we first construct his or her activity region. Figure 1 is an illustration of activity regions of users. Most existing recommendation method use entire check-in records to construct a use-POI matrix, and such matrix describes ratings of POIs by users. However, a large number of elements of this matrix is 0, leading to high sparsity problem.

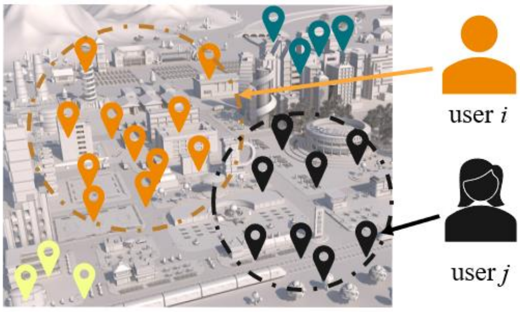

Figure 1: An example of activity regions of users.

Compared with previous method, we do not use the entire check-ins in the dataset to construct a matrix. Instead, we construct a personalized matrix for each user by only considering the check-in records in corresponding activity region, name partial use-POI matrix in this paper. Figure 2 shows an example of construction of partial user-POI matrixes.

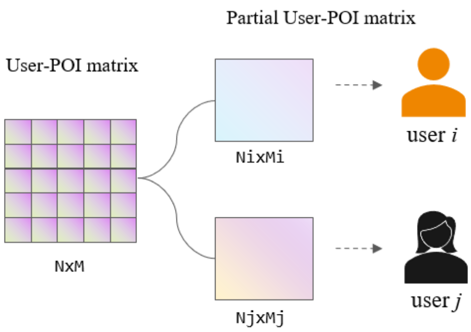

Figure 2: Partial use-POI matrix construction.

Specifically, given a user u_i and corresponding activity region R_i, subsets of both U_N and L_M are constructed, denoted by U_i and L_i, respectively. U_i only contains the users who have visited POIs located in region R_i, and L_i only contains the POIs in in region R_i. Based on U_i and L_i, partial user-POI matrixes of user u_i can be constructed by filling visiting frequencies of POIs in L_i by users in U_i.

3.2 Partial user-POI Matrix Completion

After constructing partial use-POI matrix for each user, Latent Factor Model (LFM) [7] is employed to fill the unknown values of the partial use-POI matrix.

Given a partial use-POI matrix S, we first normalize the values of S to [0, 1], and such normalized values can be regard as users' rating scores for locations. Specially, each element r_{ij} in S is the score of l_j rated by u_i. Then, Latent Factor Model (LMF) is used to update the partial use-POI matrix S as follows. The equation to achieve LMF is shown as equation (1), and figure 3 is an illustration of LMF.

$$e_{ij} = r_{ij} - \widehat{r_{ij}} = r_{ij} - \sum_{k=1}^{K} p_{i,k} q_{k,j} \tag{1}$$

$$SSE = \frac{1}{2} \sum e_{ij}^2 = \frac{1}{2} \sum \left(\left(r_{ij} - \widehat{r_{ij}} \right)^2 \right. \tag{2}$$

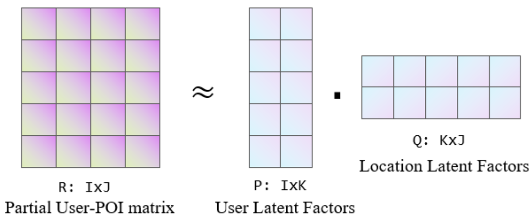

Figure 3: An illustration of latent factor model.

Gradient descent method is applied to achieve matrixes Q and P, which can minimize SSE shown in Equation (2).

3.3 Adaptive Recommendation Generation

In this subsection, we describe our approach for generation of the final POI recommendation list of a target user u_i. Assume that, the partial user-POI matrix of u_i is $S^i_{m \times n}$, and m and n denote the total number of users and POIs in the corresponding activity region of u_i, respectively. Thus, a vector of u_i can be achieved as follows:

$$\overrightarrow{v_i} = [r_{i1}, r_{i2}, \ldots, r_{in}] \tag{3}$$

r_{ij} denotes the rating of POI l_j by user u_i. Given the partial user-POI matrix S_i of the target user u_i, we calculate similarities between and other users in S_i by using Equation (4).

$$sim(u_i, u_t) = \frac{\overrightarrow{v_i} \cdot \overrightarrow{v_t}}{\|\overrightarrow{v_i}\| \times \|\overrightarrow{v_t}\|} \tag{4}$$

Cosine similarity method is applied to measure the similarity between different users as shown in Equation (3) and (4). Thus, a similar user array of can be achieved as follow in Equation (5):

$$SIM(u_i) = [u_1, u_2, \ldots, u_m] \tag{5}$$

$$\overline{sim}(u_i) = \frac{1}{m}\sum_{t=1}^{n} sim(u_i, u_t) \tag{6}$$

Then, we calculate average similarity as described in Equation (6). Next, update *SIM* (u_i) by removing the users whose similarity between u_i is lower than average similarity, denoted by *SIM'* (u_i), such that an adaptive similar user selection can be achieved. Finally, the score of a POI l_j in region R_i by user u_i can be calculated as follow:

$$socre(u_i, l_j) = \frac{\sum_{u_t \in SIM'(u_i)} sim(u_i, u_t) * r_{tj}}{|SIM'(u_i)|} \tag{7}$$

We select N POIs with highest scores calculated by Equation (7) to generate a top-N recommendation list for the target user.

4 Experimental results

In this section we will evaluate the performance of our POI recommendation system. Section 4.1 describe the datasets, evaluation metrics and compared method. Section 4.2 shows the experimental results and analyses.

4.1 Experimental Setting

Datasets: In the experiments, we use two real large-scale datasets [8]: Foursquare-New York and Foursquare-Tokyo. The datasets contain real check-ins in New York city and Tokyo, respectively. The statistics of these two datasets are shown in Table 1. We have removed the locations visited less than 10 times by all the users, and the users with less than 10 check-in records. We randomly select 70% of check-ins of users in these two datasets as training data and the rest 30% of dataset are testing data. We generate top-5, top-10 and top-20 recommendation list, respectively.

Table 1: dataset statistic

Datasets	New York	Tokyo
User number	2293	1080
Location number	6017	3138
Total check-ins	307548	47753

Evaluation Metrics: In order to evaluate the performance of our approach, we employ two metrics: precision and recall. Equation (8) and Equation (9) show the calculation of these two metrics. To facilitate the description of these two metrics, we introduce the following notations.

- Rec (u_i): the recommended POIs provided by recommendation algorithms for user u_i;
- GT (u_i): ground truth of user u_i, which denotes the POIs truly visited by u_i in the testing data.

$$percision = \frac{1}{|U|}\sum_{u_i \in U} \frac{|Rec(u_i) \cap GT(u_i)|}{|Rec(u_i)|} \tag{8}$$

$$recall = \frac{1}{|U|} \sum_{u_i \in U} \frac{|Rec(u_i) \cap GT(u_i)|}{|GT(u_i))|} \quad (9)$$

Compared Approaches: We select four previous works as baseline method and compare the performances of our approaches with these methods. The baseline methods are briefly summarized as follows:

- UCF [9]: traditional user-based collaborative filtering-based POI recommendation method.
- SpertralCF [10]: this paper explored users' latent factors from spectral domain. They employed a bipartite graph to describe the relationship between users and POI. The final recommendation was achieved by completing this bipartite graph.
- ESMP [11]: this method also explores users' activity regions, and the Manshift algorithm is employed to determine the regions. The final recommendation results are achieved also by a CF-based method.
- DWDT [12]: deep walk and tensor decomposition are used in this paper to explore users' preferences. The final recommendation list is generated by considering historical check-ins of a fixed number of similar users.

4.2 Recommendation Performance Evaluation

We have compared our approach with four baseline methods, and by using New York and Tokyo datasets, respectively. The length of the generated recommendation list is set to 5, 10, and 20. Figure 4-Figure 6 show the precisions of recommendation results for different lengths of list, testing on two datasets, and Figure 7-Figure 9 are the results for recalls of the recommendation results.

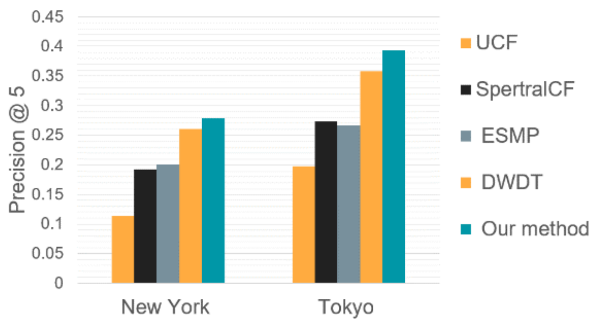

Figure 4: Precisions for top-5 recommendation

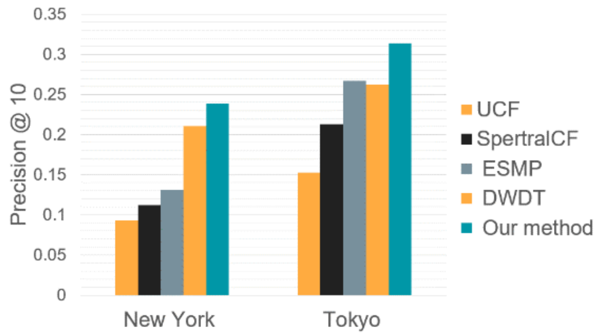

Figure 5: Precisions for top-10 recommendation

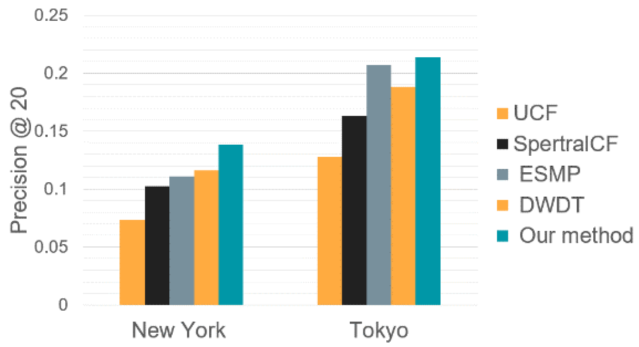

Figure 6: Precisions for top-20 recommendation

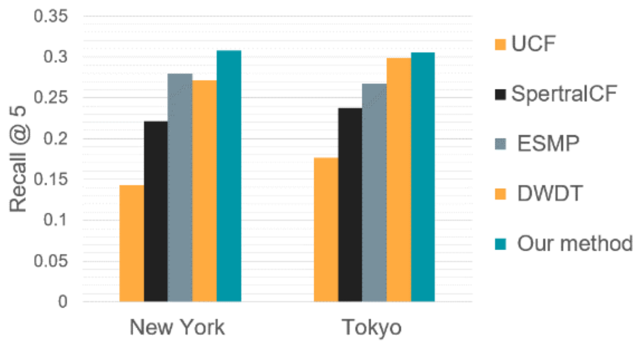

Figure 7: Recalls for top-5 recommendation

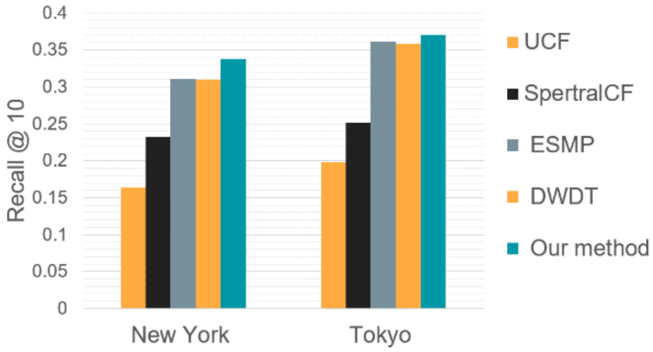

Figure 8: Recalls for top-10 recommendation

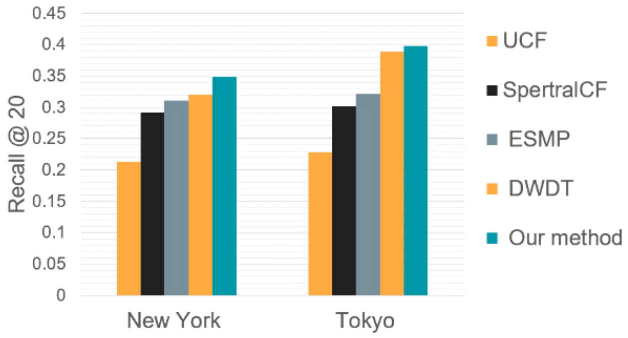

Figure 9: Recalls for top-20 recommendation

Compared with all these four previous methods, our approach could achieve highest performance for both precision and recall. The main reason is that we narrow down the recommendation candidates of users by only considering their corresponding activity regions. Besides, we also adaptively select an appropriate number of users as similar user of the target user, thus recommendation results can be more personalize.

5 Conclusions

In this paper, we investigate the POI recommendation problem and design an adaptive CF-based POI recommendation approach. Most existing approaches are CF-based and they focused on how to measure the similarities between different users. However, after similarity calculations, they simply selected first K users with highest similarity between the target user, and used their historical records to generate recommendations. Such fixed number of similar users is not suitable for the simple reason that, some similar users may be missed and some un-similar users may be included. Thus, we propose a method to select appropriate number of similar users for each different target user. The major advantages of our approach are that data sparsity problem can be alleviated and recommendation results can be more personalized.

References

[1] Ye, M., Yin, P. F., Lee, W. C., & Lee, D. L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. *International Conference on Research & Development in Information Retrieval*. ACM.

[2] Zhang, J., & Chow, C. Y. (2015). GeoSoCa: Exploiting Geographical, Social and Categorical Correlations for Point-of-Interest Recommendations. *the 38th International ACM SIGIR Conference*. ACM.

[3] Cheng, C., Yang H. King, I. (2012). Fused matrix factorization with geographical and social influence in location-based social networks. *In the 26th AAAI Conference on Artificial Intelligence*. 2012: 17—23

[4] Gao, H., Tang, J., Xia, H., & Liu, H. (2013). Exploring temporal effects for location recommendation on location-based social networks. *In Conference on Recommender Systems*. ACM.

[5] Jia-Dong, Z., Chi-Yin, C. (2015). Spatiotemporal sequential influence modeling for location recommendations. *ACM Transactions on Intelligent Systems and Technology*,7(1).

[6] Jiao, X., Xiao, Y., Zheng, W., Wang, H., & Hsu, C. H. (2019). A novel next new point-of-interest recommendation system based on simulated user travel decision-making process. *Future Generation Computer Systems, 100*.

[7] Jenatton, R., Roux, N. L., Bordes, A., Obozinski, G., (2012). A latent factor model for highly multi relational data. *Advances in Neural Information Processing Systems*. 25.

[8] Dingqi Y., Daqing Z., Vincent W. Z., Zhiyong Y., (2015) Modelling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs. *IEEE Trans. on Systems, Man, and Cybernetics: Systems, (TSMC)*, 45(1), 129-142, 2015.

[9] Quan, Y., Gao, C., Ma, Z., Sun, A., & Magnenat-Thalmann, N (2013). Time-aware point-of-interest recommendation. *In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM.

[10] Lei Z., Chun-Ta L., Fei J., Jiawei Z., and Philip S. Y. (2018). Spectral collaborative filtering. *In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM 311–319.

[11] Jiao, X., Xiao, Y., Zheng, W., Xu, L., & Wu, H (2019). Exploring spatial and mobility pattern's effects for collaborative point-of-interest recommendation. *IEEE Access, PP* (99), 1-1.

[12] Zhai, X., Zheng, W., Xiao, Y., & Liu, K. (2022). Point-of-Interest Recommendation System based on DeepWalk and Tensor Decomposition. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 867-872). IEEE.