



ABPR– A New Way of Point-of-Interest Recommendation via Geographical and Category Influence

Jingyuan Gao and Yan Yang^(✉)

Key Laboratory of Database, Heilongjiang University, Harbin, China
yangyan@hlju.edu.cn

Abstract. Point-of-Interest (POI) recommendation has been an important topic on Location-Based Social Networks (LBSN). It could recommend the POI point for users that they have never been. During the latest research, when adding geographical influence, recent research always pick up all the POI points to learn the influence it makes to the users. However, this may reduce the precision of experiment, for it does not take into consideration the reason that influences users in their frequent check-in activity region. To solve this problem, we propose a new POI recommending approach with the activity region, named Activity region Bayesian Personalized Ranking (ABPR), which adds geographical influence into the basket of BPR. This paper outlines the experiments done with Gowalla and Foursquare datasets to demonstrate the effectiveness and advantage of our approach.

Keywords: Location-Based Social Networks (LBSN)
Point-of-Interest (POI) recommendation · Geographical influence

1 Introduction

In recent years, as the rapid developing of social networks, people could share their experiences on Location-Based Social Network services, like Foursquare, Facebook and so on. People could also share their attitude to the POI points they have been on network services. To help people find the place they prefer much more, POI recommending has been proposed and plays an significant role in the stage of LBSN research.

POI Recommending: The goal of POI recommendation is learning the preference of users and recommend the place to users which they are fond of. By now, to divide the types of users, Li [1] proposed the approach of recommend next POI for users by learning the intrinsic and extrinsic interests, which is named as IEMF approach. It argues that users sign outside the activity region only because of the extrinsic reason. But the problems with the IEMF approach are that users may not check in outside the activity region only because of extrinsic reasons, and the creation of the activity region is poorly defined. Our work is different from IEMF in building activity region and the

recommend approach. Another current approach is the PG-rank [2] approach. It combines the geographical influence and the category influence. However, not all the users are affected by the geographical influence, for the POI points inside the activity region produces few influence.

We suggest that POI recommending should add in geographical influence appropriately for it is a significant factor when users are making decisions. For example, user may check-in in a restaurant famous in social networks, or he may give up that thought because of the long distance. On the other hand, we suggest that, we couldn't add geographical influence to all the POI recommendation, for it doesn't play a significant role when it comes to the situation inside their activity region. For example, you may not change your attitude of whether or not check-in in the nearer restaurant because of the distance.

To solve the aforementioned problem, we propose a new Bayesian recommendation approach via location and category influence named ABPR (Activity region Bayesian Personalized Ranking), the purpose of this approach is to analyze the record of POI point of users, then build the activity region. For the POI point inside the region, we are doing the recommending via category approach; for the outside ones, we perform the recommending by using the geographical influence and the category influence, we also updated the parameter with SGD algorithm. Try to imagine that you want to have a meal, there is a restaurant you always check-in, will you change your decision to have a meal in other POI point much closer to where you live?

We think that the related research has the following insufficient parts:

- (1) When adding the geographical influence, most of the research add the geographical influence into all the POI points, but when it comes to the POI points that are close the user, geographical influence is not the most significant reason influencing the users' decision;
- (2) The research involving activity region always builds it with circles, and it may not accurately catch the users' frequently activity region;
- (3) Some users like to explore new POI points, others may prefer the similar one. This has not been well analyzed by recently study.

To solve these problems, we propose the solution of the above– our ABPR approach. The advantage of our approach is:

- (1) We do recommendation via category inside the activity region without the geographical influence;
- (2) We calculate the activity region by using the approach of loss function to promote the precision;
- (3) We add a parameter to measure the users' preference to explore new places.

2 Related Work

POI recommending has received a lot of attention recently, and it could be classified into the following parts:

Geographical Influence Based Recommending

There are tremendous amount of research concerning the geographical influence based recommendation. Some scholars argue users may be affected by the geographical influence to change their mind in check-in behavior in the LBSN. By this mean, [3] first added the geographical influence into the POI recommendation. Then, [4] prominent the influence of geographical influence by using the Gaussian Mixture Model (GMM) and the Matrix Factorization approach (MF), but the reason of check-in is complex and it isn't being analyzed concretely. [5, 6] proved that geographical influence could increase the precision of algorithms, however, both of them didn't take in the geographical influence appropriately. [7] proposed the Potential Dirichlet Distribution to integrate the location-based model. This research indicated that adding the geographical influence may promote the performance of algorithm.

Category Influence Based Recommending

Category based recommending is the earliest and the most widely used approach. [8] proposed the Bayesian Personalized Ranking approach (BPR) which add the thought of Bayesian into POI recommending research in order to analyze the intrinsic information concealed in the check-in data. By this means, [9] proposed the approach of LBPR via listwise Bayesian Personalized Ranking to solve the problem among the successive recommending. [10] also analyzed the last patterns of users to recommend next POI point that user may be most interest in; however, the disadvantage of their research is that they didn't separate the different types of users.

Time-Aware Based Recommending

The newly developed approach aimed to catch the temporal feature of activities, which is well studied recently. Because of the regularity of human daily activity, temporal influenced recommendation helps researchers to catch the preference of POI points much more precisely. [15] firstly add the temporal influence into POI recommendation, [11] divide the time slot into days, months and years, but the time slot is too long to catch the users' activity pattern.

Social Influence Based Recommending

Social influence based recommend is inspired by the intuition that there are some common interests of friends in the LBSN. [12] showed the social relations exert more influence on long distanced travel. [13] proposed the model of PMRE-GTS to show the connect of social relation and temporal influence.

Because of the sparsity and time complexity of the dataset, most of literature only build their models with two influences. And for the long-distance POI points, most of them are not ideal.

3 The ABPR Model

The recommendation task is defined as following:

Given the check-in behaviors of users and the locations of the POI points, we aim at recommending the Top- K locations users might be interested in but never visited before or not often check-in.

3.1 Our Framework

Activity Regions. People’s activity has some type of scope [14], and the recent research mostly added the geographical influence into all the POI points. But the geographical influence effects are much lower than we considered inside the region. Based on such considerations, we propose the concept of activity region. Users usually check-in around one or more regions, such as school, work place and home. We consider that if adding the geographical influence into the recommendation, it will reduce its precision. So, we use the SGD method to create the activity region. We performed category based recommending inside the region, performed the category and geographical based recommending outside the region. For the geographical influence may tremendously influence the check-in decision. Figure 1 shows the idea of activity region.



Fig. 1. The activity region (Color figure online)

The blue circle is the activity region, different color of points shows the different reason of check in.

For the outside points, we add in the geographical influence, here is the lost function of building the activity region:

$$J(\theta_a) = \frac{1}{2m} \sum_{i=1}^m (t_{\theta_a}(x^{(i)}) - y^{(i)})^2 \quad (1)$$

$x^{(i)}$ is the coordinate of check-in location, $y^{(i)}$ is activity region, t_{θ_a} is the matrix of POI points, θ_a is the parameter matrix of t , $(t_{\theta_a}(x^{(i)}) - y^{(i)})^2$ is the distance from outside POI points to the activity. We minimize the sum of the distance:

$$A(\theta_a) = \arg \min \frac{1}{2m} \sum_{i=1}^m (t_{\theta_a}(x^{(i)}) - y^{(i)})^2 \quad (2)$$

To minimize this formula, we use the SGD method to calculate the parameter matrix θ_a and get the activity region, the process of calculating θ_a is:

$$\frac{\partial A(\theta_a)}{\partial \theta_a} = \frac{1}{m} \sum_{i=1}^m (t_{\theta_a}(x^i) - y^i)x^i \quad (3)$$

If we want to get the Minimum value of $A(\theta)$, then $\frac{\partial A(\theta_a)}{\partial \theta_a}$ should be 0. During the iteration, we use α_1 as the step of iteration, the process of iteration is as the following:

$$\theta_a = \theta_a - \alpha_1 \left(\frac{1}{m} \sum_{i=1}^m (t_{\theta_a}(x^i) - y^i)x^i \right) \quad (4)$$

until it convergences.

The Process of Naive Bayesian. posteriori probability could be define as:

$$P(l|l_u) \propto P(l)P(l_u|l) \quad (5)$$

And it equals to:

$$P(l|l_u) = P(l) \prod_{l' \in l_u} P(l'|l) \quad (6)$$

Where $P(l)$ is the posteriori probability of check-in in POI point l , l_u is the set of history check-in points, l' is the point of history check-in.

The geographical Influence. Users may be influenced by the geographical influence and category influence to check-in outside the activity region. By this means, we add the geographical influence into the process of POI recommendation.

We use $U \in R^{k*1}$ matrix as the set of users, $l \in L^{n*1}$ as the set of POI points, θ_l to as the parameter of matrix. Here we use the Euclid distance $D_{l_i l_j}$ between the users and the POI points.

Users may be influenced by the geographical influence, in other words, the distance between the users and the POI points may influence the check-in decision. For example, because of the distance, we may go to check-in in the shop 3 km from where you live, and we may not check-in in the 6 km one. We define $|X(l_i) - X(l_c)|^2$ as the distance from POI points to activity region. If it is much more near to the user, we define it as follow:

$$|X(l_i) - X(l_c)|^2 - |X(l_j) - X(l_c)|^2 \quad (7)$$

We use this formula to calculate l_i or l_j is closer to the activity region. If $D_{l_c l_i} - D_{l_c l_j} < 0$, To make it easier to array, we negate it as:

$$-(D_{l_c l_i} - D_{l_c l_j}) > 0 \quad (8)$$

The Desire to Explore the New POI Point. In our daily lives, some people desire to explore the new POI points, while some people prefers to check-in to the similar ones. Until now, few research involves the desire of individuals to explore. We would like to introduce the desire parameter χ to define the desire to explore the new POI point. If the users may desire to explore the similar ones, we may recommend the POI points based on the category factor.

We define the check-in set as U , ϑ_u as the number of the check-in category, η as the frequency of check-in, we divide ϑ_u by U as follow:

$$\chi = \frac{\vartheta_u}{\eta} \quad (9)$$

Clearly, it is a constant, if it is above the value θ_q , we may consider that the user prefers to check-in in the similar one.

3.2 Parameter Estimation

According to the Bayesian personalized ranking thought, we use the maximum likelihood posteriori probability to build the target formula, and then doing iteration to calculate the parameter and combine the activity region, geographical influence and desire parameter. To optimize the parameter, its derivation process is given by the following formula:

$$\gamma = \operatorname{argmax} \prod_{u \in U} \prod_{l \in L^{n+1}} \prod_{l_i \in L^{n+1}} \prod_{l_j \in L^{n+1}} \chi P(>_{u,l \in a} | \theta) P(\theta) \quad (10)$$

We define γ as the probability of check-in, The parameter $\theta = \{X(L), X(U), X(a), X(q)\}$ is the parameter that needs to be optimized, we change it by using the sigmoid formula $\sigma = \frac{1}{1+e^{-x}}$, we define the probability as:

$$P(>_{u,l \in a} | \theta) = \sigma(x_{u,i,j}) \quad (11)$$

So the optimize formula could be changed as:

$$\begin{aligned} \gamma &= \operatorname{argmax} \log \prod_{u \in U} \prod_{l \in L^{n+1}} \prod_{a \in R^+} \prod_{q \in R^+} \chi P(>_{u,l \in a} | \theta) P(\theta) \\ &= \operatorname{argmin} \sum_{u \in U} \sum_{l \in L^{n+1}} \sum_{l_i \in L^{n+1}} \sum_{l_j \in L^{n+1}} -\log \sigma(x_{u,i,j}) + \log P(\theta) + \theta_q \chi \end{aligned} \quad (12)$$

The above formula is inside the activity region, when it comes to the outside the region, it can be defined as:

$$= \operatorname{argmin} \sum_{u \in U} \sum_{l \in L^{n \times L}} \sum_{l_i \in L^{n \times 1}} \sum_{l_j \in L^{n \times 1}} [-\log \sigma(x_{u,i,j}) + \lambda_\theta \|\theta^2\| + \theta_q \chi] + A(\theta_a) \quad (13)$$

For the outside ones, we define the geographical influence as the most significant factor, based on this idea, we change the above formula as:

$$\gamma = \operatorname{argmax} \prod_{u \in U} \prod_{l \in L^{n \times 1}} \prod_{l_i \in L^{n \times 1}} \prod_{l_j \in L^{n \times 1}} P(>_{u,l \in a} | \theta) P(\theta) \quad (14)$$

$$\text{Well the } P(>_{u,l \in a} | \theta) = P(-(D_{l_c l_i} - D_{l_c l_j}) > 0 | \theta)$$

$$= -\sigma(D_{l_c l_i} - D_{l_c l_j}) \quad (15)$$

So it could be changed as:

$$\begin{aligned} \gamma &= \operatorname{argmax} \log \prod_{u \in U} \prod_{l \in L^{n \times 1}} \prod_{a \in R^+} \prod_{q \in R^+} P(>_{u,l \in a} | \theta) P(\theta) \\ &= \operatorname{argmin} \sum_{u \in U} \sum_{l \in L^{n \times L}} \sum_{a \in R^+} \sum_{q \in R^+} -\log \sigma(D_{l_c l_i} - D_{l_c l_j}) + \lambda_\theta \|\theta^2\| \end{aligned} \quad (16)$$

3.3 Algorithmic Description

We firstly extract all the train set, then optimize the parameter, we could get the check-in dataset ψ , it contains $\{u, l^c, \vartheta_u, \eta\}$, the four parameter defines users; the users' location; the location of POI points and its category; the frequency of check-in. The approach of update the parameter is:

$$\theta \leftarrow \theta - \alpha \frac{\partial \gamma}{\partial \theta} \quad (17)$$

The α defines the learning step of SGD. To express the idea of the algorithm more Intuitively, we list the process of the algorithm as follow:

The algorithm of ABPR

Input:check-in dataset,learning step
 α , regularization parameter λ_θ

Bulid the activity region;

Repeat

For every $\{u, l^c, \mathcal{G}_u, \eta\}$

 If $A(\theta_a)^t=1$ (inside the activity region)

 Update X(L),X(U),X(a),X(q);

 Else $A(\theta_a)^t=0$

 Update X(L),X(U),X(a);

Until convergence

Return the parameter $\{X(L),X(U),X(a),X(q)\}$;

4 Experiments

4.1 Dataset

We use the Gowalla and Foursquare datasets to evaluate the model performance. each check-in record contains the user’s ID and the POI point location, each location has the latitude and the longitude information. The data statistics are shown in Table 1.

Table 1. The data statistics

Dataset	No. Check in	No. Users	No. POI
Foursquare	18492	1228	3360
Gowalla	34120	1838	2952

4.2 Evaluate Parameter

We quantitatively evaluate the model performance by using top- K recommending performance, i.e, Precision@ K and Recall@ K . They are defined as following:

$$\text{Precision@K} = \frac{1}{n} \sum_{i=1}^n \frac{S_i(K) \cap \tau_i}{K} \quad (18)$$

$$\text{Recall@K} = \frac{1}{n} \sum_{i=1}^n \frac{S_i(K) \cap \tau_i}{|\tau_i|} \quad (19)$$

Where $S_i(K)$ is a set of top- K unvisited locations recommend to user i excluding those locations in the training, τ_i is the set of locations that are visited by users i .

4.3 Performance Comparison

We compare the proposed model with the following methods:

BPR [8]: optimizes the ordering relationship of users' preferences for the observed and unobserved location;

USG [3]: incorporates the geographical influence, social network and user interest into collaborative filtering in an additive manner.

IEMF [1]: learning users' intrinsic and extrinsic interests for POI recommendation.

We experiment these methods and our ABPR with the real world datasets, and here is the comparison result (Fig. 2):

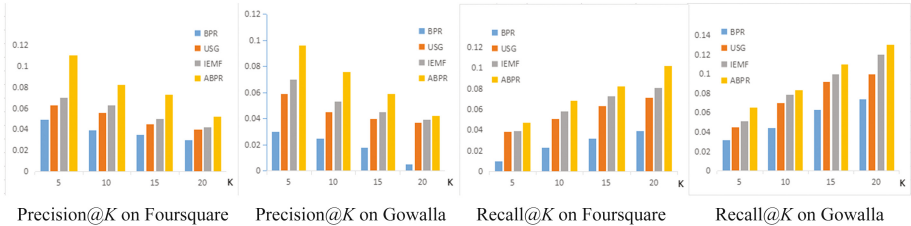


Fig. 2. Performance comparison

We evaluated the model performance by using precision@ K and recall@ K , our model gains about 15.4% and 17.3% improvement in precision@5 and recall@20 in Gowalla data; 12.6% and 13.8% in Foursquare data. Through analyzing the experimental result, we found the following observations.

First of all, our method outperforms all other methods. This superior result is for modeling the users' preference. By this means, we've done it better.

Secondly, the methods with geographical influence (USG, IEMF, ABPR) performs much better than the one without geographical influence (BPR). This further illustrates the benefits of doing POI recommending with geographical influence.

Lastly, we'd like to take about the influence of the number of iterations of building activity region, and the outperform of the method of building the activity region. The influence of it will be illustrated in Fig. 3. The red line of the figure shows our precision of our activity region, the green one shows the method of STELLAR [16], it adds the geographical influence to separate the POI points only with distance.

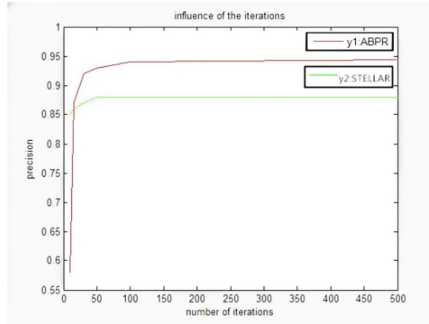
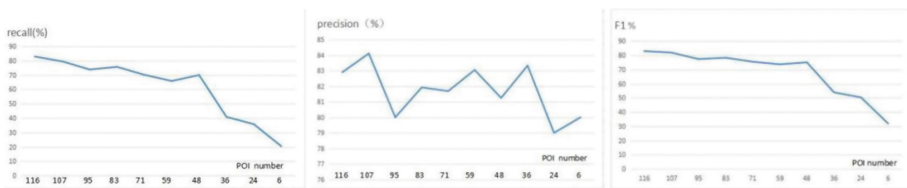


Fig. 3. Influence of the number of iterations (Color figure online)

Clearly, our method of building the activity region of ABPR performs much more better than the STELLAR [16] of adding geographical influence. STELLAR abandons the POI points over 10 km far. To illustrate that, we compared with the two methods to show the advantage of our method.

We can also conclude from the figure that when the iteration comes to more than 100, it convergences. So we suggest that the iteration may be getting its best situation when more than100 iteration is performed.

We also analyzed how does POI points effect the performance of activity region. The Fig. 4 shows the influence:



Precision of creating activity region Recall of creating activity region F1 score of creating activity region

Fig. 4. The connection between activity and the POI number

We found that our method does well when there are more than 30 POI points. The precision is not affected as much when the number of POI points changes, it remains at 80%. However, it affects much on the percentage of recall. The F1 score shows when POI point drops down to lower than 30, the score disparity of precision and recall gets higher. If the POI points are lower than 30, it is undesirable in building the activity region. Here are four examples of the situation of it (Fig. 5):

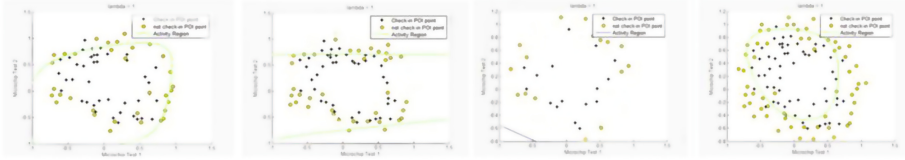


Fig. 5. Four situations of the POI points lower than 40 and more than 40

It seems that our approach is undesirable when the POI points is lower than 30 for the algorithm try to contain all the POI points that are being checked-in, but it does well in other situations, like the forth figure which has 137 POI points.

5 Conclusion

To further the study of personalized recommending, in this paper, we propose a new method to catch the preference of users based on the geographical influence and the category influence. We use the approach of SGD to estimate the parameter, and performance evaluation in the real world LBSN datasets shows our approach of POI recommending outperforms other methods. More specifically, our proposed method is more capable in distinguishing different types of users. For future works, we could consider the social relations and temporal influences on personalized recommending.

References

1. Li, H.Y., Ge, Y., Lian, D.F., Liu, H.: Learning user's intrinsic and extrinsic interests for point-of-interest recommendation: a unified approach. In: University of North Carolina at Charlotte, America (2017)
2. Feng, S., Li, X., Zeng, Y., Cong, G., Chee, Y.M., Yuan, Q.: Personalized ranking metric embedding for next new POI recommendation. In: IJCAI, pp. 2069–2075 (2015)
3. Ye, M., Yin, P., Lee, W.-C., Lee, D.-L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: SIGIR, pp. 325–334 (2011)
4. Cheng, C., Yang, H., Lyu, M.R., King, I.: Where you like to go next: successive point-of-interest recommendation. In: IJCAI 2013 (2013)
5. He, J., Li, X., Liao, L.: Category-aware next point-of-interest recommendation via listwise Bayesian personalized ranking. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017). Beijing Institute of Technology (2017)
6. He, J., Li, X., Liao, L.: Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In: AAAI 2016. Beijing Institute of Technology (2016)
7. Kurashima, T., Iwata, T., Hoshida, T., Takaya, N., Fujimura, K.: Geo topic model: joint modeling of user's activity area and interests for location recommendation. In: WSDM, pp. 375–384 (2013)
8. Rendle, S., Christoph, F., Zeno, G.: BPR: Bayesian Personalized Ranking from Implicit Feedback. University of Hildesheim, Germany (2009)

9. Cho, E., Mysers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social network. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, USA, pp. 1082–1090 (2011)
10. Han, Z., Liu, Y.: PMRE-GTS: a new type of continuous POI recommending model via temporal influence and social influence. In: CCF of 2017 (2017)
11. Liu, Y., Wei, W., Sun, A., Miao, C.: Exploiting geographical neighborhood characteristics for location recommendation. In: Proceedings of CIKM, pp. 739–748 (2014)
12. Yao, L., Sheng, Q.Z., Qin, Y., Wang, X., Shemshadi, A., He, Q.: Context-aware point-of-interest recommendation using tensor factorization with social regularization. In: SIGIR, pp. 1007–1010 (2015)
13. Li, X., Cong, G., Li, X., Pham, T.N., Krishnaswamy, S.: Rank-GeoFM: a ranking based geographical factorization method for point of interest recommendation. In: SIGIR, pp. 433–442 (2015)
14. Hu, B., Ester, M.: Social topic modeling for point-of-interest recommendation in location-based social networks. In: ICDM, pp. 845–850 (2014)
15. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Time-aware point-of-interest recommendation. In: SIGIR 2013, 28 July–1 August 2013
16. Zhao, S., Zhao, T., Yang, H., Micheal, R., King, I.: STELLAR: spatial-temporal latent ranking for successive point-of interest recommendation. In: AAAI 2017, pp. 315–321 (2017)
17. Jinghua, Z., Xuming, Y., Yake, W., et al.: Structural holes theory-based influence maximization in social network. In: The 12th International Conference on Wireless Algorithms, Systems, and Applications, WASA 2017, Guilin, China, pp. 860–864 (2017)