

Research on Point of Interest Recommendation Algorithm Based on Spatial Clustering

Liguo Zheng

*Dept. of Computer and Information Engineering School, Harbin Normal University,
China
zhenglg@hrbnu.edu.cn*

Abstract

In order to solve the problem of recommendation of points of interest, this paper proposes an algorithm of recommendation of points of interest based on user check-in space clustering. According to the administrative region information of interest points in LBSN and the distribution characteristics of user check-in, a new spatial clustering algorithm is designed in this paper. First, according to the distribution of user check-ins, the whole data set was clustered in cities, and the user rating information was normalized. Then the recommendation scores of candidate recommendation points were calculated according to the user preference model, social relation model and geographical correlation model. The final recommendation list is obtained by calculating the recommendation probability of the points of interest. Experiments on the Yelp data set show that the proposed algorithm has higher precision and recall rate than the traditional algorithm.

Keywords: *LBSN, POI, Recommendation algorithm, Space clustering, YELP*

1. Introduction

In recent years, with the rapid rise of smart phones and GPS positioning technology, various social platforms can obtain users' geographical location information through mobile apps. It makes the original social network also have the geographical correlation attribute, and then produces the social network based on location [1]. The LBSN platform uses GPS positioning technology to obtain the current location information of users and link the social network with the real world [2]. While realizing information sharing in social network, it also realizes the spread of location in social network.

As a new platform for connecting social networks to the physical world, location-based social networks are having a giant impact on our daily lives. Well-known social platforms include Foursquare, Brightkite, Gowalla, Yelp, and social software like WeChat, Renren, and Dianping. Through these LBSN platforms, users can share their current mood, what they have seen, recommend a good restaurant to friends, or share a meaningful article with their circle of friends. Location-based social networks not only affect people's social life, but also bring new opportunities and great business value to related interconnected enterprises [3].

Location-based social networks have also opened up a series of new research topics in the academic field, such as user behavior analysis, motion pattern research [4] and various real-life applications. The work of providing users with personalized information filtering and

Article history:

Received (February 18, 2020), Review Result (March 22, 2020), Accepted (April 29, 2020)

recommendation based on multi-source heterogeneous information such as user preferences, geographical location information and social relations has won wide attention in the academic community

Due to the surge in the number of users of mobile apps and merchants' settlement on social platforms, the geographical attribute information of users' check-in and description information about merchants obtained by LBSN platform also showed an exponential growth. How to find interested merchants for users from a large number of information resources is the current problem faced by the development of LBSN platform. Therefore, according to the user's interests and preferences and social factors, screening valuable information for the user and recommending it to the user becomes a problem to be solved by relevant enterprises and researchers. Location Recommendation, an emerging application, is to recommend a suitable Point-Of-Interest (POI) for users. Therefore, it has become an emerging topic studied by enterprises and academia.

The recommendation function of location-based social network can not only help users find places they may be interested in, but also bring huge business value to relevant Internet enterprises and entity businesses. Therefore, it is of great research value to improve and improve the location recommendation algorithm.

2. Review

Location recommendation algorithm is based on the user's check-in history, score information, comment text, check-in time, social relations and other heterogeneous information to build the user's preferences, social relations and geographical characteristics. And according to similar users or friends of the check-in situation for users to recommend some places he has not visited. In the location recommendation algorithm, three feature models are generally built according to the check-in data of users: the user preference feature model, the social relationship feature model and the geographical factor feature model. The following is an introduction to the construction methods of three feature models and the research on location recommendation based on the integration of multiple models.

2.1. Build user preference characteristics

When constructing user preference characteristics, the following algorithms can be used: content-based recommendation algorithm [5][6], recommendation algorithm based on collaborative filtering and hybrid recommendation algorithm [7]. Content-based recommendation is to learn the user's interest preference according to the user's evaluation of the project or the project's own attributes, and to recommend the points of interest by examining the matching degree between the user's information and the project. In literature [6], Zhang et al. believed that the check-in frequency of users could not truly reflect their preferences, but was implied in the text comment information package of users on points of interest. They analyze and study user comments and make location recommendations by analyzing the matching degree between user preferences and points of interest. In literature [8], Yang et al. proposed the User Preferences on Spatiotemporal Topics model (UPOST), which learned spatiotemporal theme from the historical semantic location of users and inferred the users' preference for different types of locations in different periods. In literature [9], multi-attribute network is used to effectively reflect some attributes and measure the direct and indirect similarity between items, and the correlation between items is considered according to the clustering technique. In the above content-based recommendation models,

relevant models in machine learning are used for analysis, which can well solve the problems of information overload and data sparsity.

2.2. Build social relationship characteristics

In location recommendation, social relations are mainly based on the similarity of interests and preferences between users and friends and the influence of friends on users' strong organizational relationships. Make it possible for the user to visit businesses or places he has not been to. Wen et al., in literature [10], captured information such as interactions between communities, physical activities and time effects. Thus, the social influence among users can be inferred and a new social influence recommendation framework can be constructed. Literature [11] analyzed the closeness between friends to build the strength of social association between users, and at the same time considered the spatial and temporal association between users when making location recommendations.

2.3. Build geography characteristics

In location recommendation, geographical factors refer to the physical distance that affects the user's check-in behavior. Generally, due to the influence of geographical factors, users tend to conduct activities in a specific area, which is called the "active area". The geographical influence on users' check-in behavior is personalized. According to this theory, a personalized geographical location recommendation framework named iGeoRec is proposed. In literature [12], Lin et al. analyzed the distance between the user's current location and the recommended location, constructed geographical factors by using naive Bayesian classification model, and made location recommendation by combining the user's personal preference and the social influence model built by collaborative filtering algorithm. In literature [13], geographic information is converted into matrix decomposition, and information such as check-in times and comment text are integrated into the matrix decomposition model for location recommendation. Zheng et al., in literature [14], proposed a recommendation service for location-aware points of interest located in urban shopping centers. This model firstly mined users' historical shopping tracks to obtain users' location preferences. At the same time, the user's personal interest is taken into account to generate the final recommended POIs. Experiments show that this method has good recommendation performance in several existing benchmark methods.

3. Point of interest recommendation algorithm based on user check-in space clustering

In this paper, a spatial clustering algorithm is designed based on the administrative region information of interest points in LBSN and the distribution characteristics of user check-ins. The rating information of users is normalized, and then the distribution of each user's check-in place in the cluster subset is determined, and the points of interest are recommended based on the characteristics of user rating and social relations. Experiments on real Yelp data sets verify the effectiveness of the proposed algorithm.

3.1. Build characteristic model

First, we normalize the rating information of users. Then, the multi-source heterogeneous information, such as the number of check-ins and scores of similar users at the points of interest, is integrated, and the user-based collaborative filtering model and the social-relation-

based collaborative filtering model are constructed. Finally, according to the popularity of the points of interest and the naive Bayesian Algorithm, the geographical correlation model is constructed.

Make $U = \{u_1, u_2, u_3 \dots, u_N\}$ and $L = \{l_1, l_2, l_3 \dots, l_M\}$ respectively represent the user set and the collection of points of interest, where N and M respectively represent the number of users and the number of points of interest in the data set. When user u_i accesses POI_{l_j} and gives a score, $r_{i,j}$ is used to represent the score information, and $y_{i,j}$ is used to represent the number of check-in times of u_i in l_j . In addition, the popularity of POI_{l_j} is represented by p_{l_j} , which represents the total number of check-ins of all users in l_j .

The rating information of users on the points of interest often indicates the degree of users' preference for the site, but the same rating may represent different degrees of preference depending on the rating scale of each user on the points of interest.

In this paper, it is sorted according to the chronological order of users accessing the same POI, as shown in [Figure 1], and then the score information of users is weighted. The calculation formula is defined as follows:

$$r'_n = \begin{cases} r_1, & n = 1 \\ \frac{(r'_{n-1} + r_n)}{2}, & n > 1 \end{cases} \quad (1)$$

Where, r_n is the NTH score of the current user on the same POI, and r'_n is the new score obtained by weighted treatment. When the user has k scores in the same POI, the final score is r'_k after weighted by [Formula 1].

No	User ID	Time	POI ID	Stars
1	18500245	2014-03-18	15224152	4
2	18500245	2014-04-15	15224152	5
3	18500245	2014-04-29	15224152	4
4	18500245	2014-06-02	15224152	4
5	18500245	2014-08-10	15224152	3

Figure 1. Current user rating information on the same POI

Since the scale of each user's rating is different, the weighted rating of the same user is then normalized by the method of decoupling. The definition is shown in formula 2

$$Pr_{i,j}(r'_{i,j} \text{ is preferred}) = Pr(\text{Rating} \leq r'_{i,j}) - Pr(\text{Rating} = r'_{i,j})/2 \quad (2)$$

Where, $r'_{i,j}$ is the final score information obtained by user u_i after weighting POI l_j , and $Pr_{i,j}$ is the score obtained after normalization, and $Pr_{i,j} \in (0,1)$

When a user visits a POI many times, it means that the location may be close to the user's residential location or that the user prefers the POI. Therefore, the simple binary representation of the user's check-in frequency data does not accurately capture the current user's preference for POI. When calculating user similarity, this chapter USES the number of check-in times of users at points of interest to calculate the similarity between users. The formula is defined as follows:

$$w_{i,k} = \frac{\sum_{l_j \in L} y_{i,j} y_{k,j}}{\sqrt{\sum_{l_j \in L} y_{i,j}^2} \sqrt{\sum_{l_j \in L} y_{k,j}^2}} \quad (3)$$

In the formula, $y_{i,j}$ and $y_{k,j}$ respectively represent the check-in frequency of user u_i and u_k in POI l_j , where l_j belongs to a place on the set of points of interest L, and $w_{i,k}$ represents the similarity between two users.

After obtaining the similarity between users, this paper selects the first N' similar users with the highest degree of similarity with the current user u_i to form the user neighbor set U' , and $U' \subset U$. Since the number of check-ins and score information of users on the points of interest can also reflect users' interest preference on the POI to a certain extent. Therefore, when calculating the possibility for user u_i to sign in for POI l_j that has not been accessed by user's collaborative filtering model, this paper integrates the check-in times and scoring information of similar users. Its definition is shown in formula 4,

$$P_{i,j}^u = \frac{\sum_{u_k \in U'} w_{i,k} \cdot y_{k,j} \cdot Pr_{k,j}}{\sum_{u_k \in U'} w_{i,j}} \quad (4)$$

In location recommendation, social relationship is mainly based on the similarity of interests and hobbies between friends, as well as the strong organizational relationship between users and friends. When calculating the social similarity between users, this paper considers the social relationship between two users and the repeated possession ratio of check-in data in the same place to calculate the similarity between users and their friends. The specific calculation formula is as follows,

$$sw_{i,k} = \eta \frac{|F_i \cap F_k|}{|F_i \cup F_k|} + (1 + \eta) \frac{|L_i \cap L_k|}{|L_i \cup L_k|} \quad (5)$$

Where, F_i represents the friend set of users u_i , l_j represents the set of check-in interest points of user u_i , η is a parameter, $\eta \in [0,1]$

According to the similarity calculated from the social relationship, the similar users with the highest score of N' before are selected to form the user neighbor set U' . According to the collaborative filtering algorithm of social relations, when calculating the possibility of user u_i to check in the unvisited POI l_j it also integrates the information such as the number of checkin times and score of social similar users at the points of interest. The formula is as follows:

$$P_{i,j}^s = \frac{\sum_{u_k \in U'} sw_{i,k} \cdot y_{k,j} \cdot Pr_{k,j}}{\sum_{u_k \in U'} sw_{i,k}} \quad (6)$$

3.2. The clustering algorithm fuses user check-in distribution

In the Yelp data set, the information of the point of interest includes the longitude and latitude information of the business, as well as the information of the corresponding city. Through the collation and analysis of the experimental data set, it can be seen from the data that the administrative area where the points of interest are located and the number of checkups of users are mostly concentrated. The check-in records of users in the data set are mostly distributed in a few cities. This chapter reevaluates the city center where check-ins are relatively concentrated based on the number of check-ins, and takes this center as the clustering center of location clustering.

In the difference analysis of the training data set and the test data set, we assume that a user u_i has checked in at n points of interest, and these locations are distributed in m cities, then $L_i = \{l_1, l_2, l_3, \dots, l_n\}$ and $C_i = \{c_1, c_2, c_3, \dots, c_m\}$ are respectively the collection of user u_i check in points of interest and the city information distributed. In this chapter, c_a^i is used to represent the city c_a visited by user u_i . When the city of the sign-in POI l_j of user u_i is c_a ,

$c_a^{i,j}$ is used to represent it. The average distance between the set of points of interest l_j visited by user u_i is defined as d_i , and the value of its d_i is expressed by formula 7:

$$d_i = \frac{2 \sum_{j=1}^n \sum_{k=j+1}^n d(l_j, l_k)}{n \cdot (n-1)} \quad (7)$$

Where $d(l_j, l_k)$ represents the distance between POI l_j and l_k .

When the clustering urban c_a^i recommends the points of interest for users, this paper selects only the POIs owned by the city and not visited by the current users as the candidate recommendation points of interest. And for candidate recommendation points of interest l_j , if the average distance between $d_{i,j}$ and the set of historical check-in points of interest of user u_i is less than $\lambda \times d_i$, the recommendation probability will be calculated. Otherwise, filter out these candidate recommendations. Then, according to the score of recommendation probability, the previous $K_{c_a^i}$ interest points with the highest score were selected as the recommended positions of this cluster city. The calculation formula of the recommended probability is defined as follows:

$$S_{i,j} = (1 - \alpha - m')S_{i,j}^u + m'S_{i,j}^s + \alpha S_{i,j}^g \quad (8)$$

The weight parameter m is used as the adaptive adjustment parameter of the social influencing factors. When $m \leq 10$, there is $m = \frac{m}{10}$, If $m > 10$, then $m = 1$, and we have $0 < \alpha + m \leq 1$. That is, when the number of check-in cities of users is more than 10, according to the adaptive adjustment parameters, only the collaborative filtering algorithm based on social relations is used to recommend points of interest in the location recommendation. parameter α is to set the proportion of geographical correlation model in the recommendation.

When making recommendations based on user similarity, social relationship and geographic information. The values of $P_{i,j}^u$, $P_{i,j}^s$ and $P_{i,j}^g$ are recommended by different models, so the standards of these recommended results are different. Therefore, the framework integration cannot be simply weighted processing, the standardized processing formula used in this paper is as follows,

$$S_{i,j}^u = \frac{P_{i,j}^u}{Z_i^u}, Z_i^u = \max_{\max l_j \in (L-L_i)} \{P_{i,j}^u\} \quad (9)$$

$$S_{i,j}^s = \frac{P_{i,j}^s}{Z_i^s}, Z_i^s = \max_{\max l_j \in (L-L_i)} \{P_{i,j}^s\} \quad (10)$$

$$S_{i,j}^g = \frac{P_{i,j}^g}{Z_i^g}, Z_i^g = \max_{\max l_j \in (L-L_i)} \{P_{i,j}^g\} \quad (11)$$

4. Experimental results

4.1. Experimental data

The method presented in this paper was validated using the Yelp data set. Yelp is a famous business review website in the United States, where users can check in, rate, write reviews and make friends among users, and each POI has a category logo. In the Yelp data set, the time range of user check-in data is from 2004-02-12 to 2015-06-24, with a total of 4751,523 check-in data. Since the user's check-in time span is relatively large, and most of the check-in records are concentrated after 2011, this paper preprocesses the data set first. This chapter selects the check-in records from 2012 to 2015, and only keeps the data with more than 10

check-in records of users and points of interest as the standard data set. In the standard data set, there were 35,241 independent users, 21,452 points of interest, 452,125 social relationships, and 785,452 check-in records.

4.2. Evaluation standard

In this paper, two measures in the fields of information retrieval and statistics, Precision and Recall, were used as the evaluation indexes of this experimental method. The average time recommended by all users was used to compare the time performance of different models. The three evaluation indicators are defined as follows,

$$Precision = \frac{1}{N} \sum_{i=1}^N \frac{S_i(k) \cap T_i}{k} \quad (12)$$

$$Recall = \frac{1}{N} \sum_{i=1}^N \frac{S_i(k) \cap T_i}{|T_i|} \quad (13)$$

$$AverTime = \frac{TotalTime}{N} \quad (14)$$

4.3. Analysis of experimental results

As shown in [Figure 2] and [Figure 3], the model proposed in this paper is better than USG and USPB models in the two performance indexes of Precision and Recall. When 5 points of interest are recommended to users, the accuracy performance of the model proposed in this paper is 8.57% higher than that of the USG model, and 21.63% higher than that of the USPB model. The model proposed in this paper is 7.15% higher than USG model and 26.78% higher than USPB model. When recommending other points of interest of different Numbers, the accuracy and recall rate of the model proposed in this paper are also improved compared with USG model and USPB model. This demonstrates the effectiveness of adaptive location recommendations based on whether the user's check-in behavior has an "active area". Then, the PROPOSED model and the USG model are divided into PROPOSED, SCUSPG, GCUSPG and UUSG, SUSG and GUSG algorithms according to the different characteristics of the recommendation.

When five points of interest are recommended to users, the model splitting method proposed in this paper also improves the accuracy and recall rate compared with the USG model splitting method. This shows the effectiveness of considering location clustering, check-in frequency and user rating information. Then, this chapter combines the UUSG model with spatial clustering, check-in frequency, score normalization and other characteristics, and when recommending 5 points of interest, its recommendation accuracy increases by 5.89%, 0.15% and 5.62%, and the recall rate increases by 6.01%, 0.18% and 1.8%, respectively.

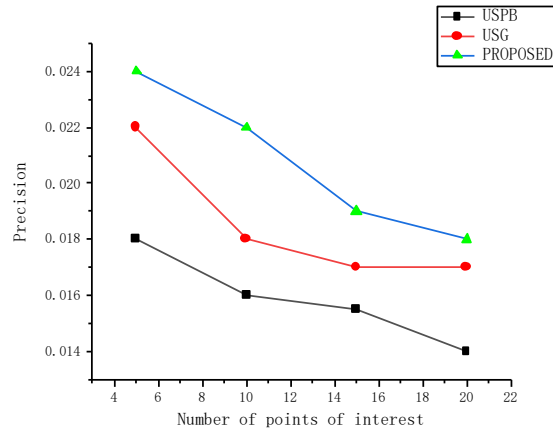


Figure 2. Precision of recommended points of interest

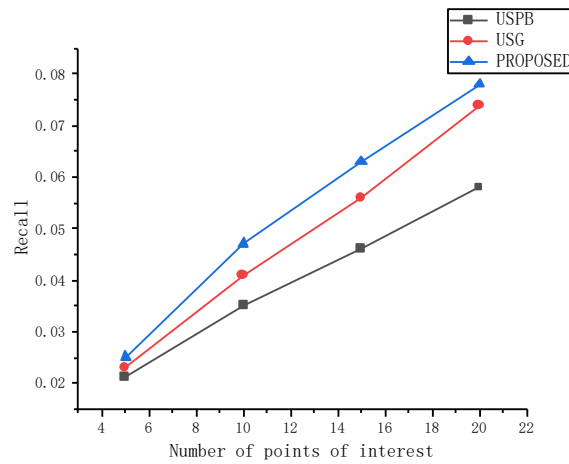


Figure 3. Recall of recommended points of interest

5. Conclusion

In order to solve the problem of recommendation of points of interest, this paper proposes an algorithm of recommendation of points of interest based on user check-in space clustering. According to the administrative region information of interest points in LBSN, the city is redivided by clustering. Then, three basic recommendation algorithms are used to build the model. At the same time, this chapter also combines information such as user rating and check-in frequency of interest points. Experimental results show that the improved algorithm is better than the traditional recommendation algorithm in accuracy and recall rate, and to a certain extent, it can solve the problem of scalability of collaborative filtering.

Reference

- [1] Julong Pan, Zhengwei Zuo, Zhanyi Xu, and Qun Jin, "Privacy protection for LBS in mobile environments: progresses, issues and challenges," International Journal of Security and Its Applications, vol.9, no.1, January, pp.249-258, (2015) DOI:10.14257/ij sia.2015.9.1.24

- [2] Jinying Jia and Fengli Zhang, “Non-deterministic K-anonymity Algorithm Based Untrusted Third Party for Location Privacy Protection in LBS,” *International Journal of Security and Its Applications*, vol.9, no.9, pp.387-400, (2015) DOI:10.14257/ijisia.2015.9.9.33
- [3] Jeong-Sig Kim, Eung-Sung Kim, and Jin-Hong Kim, “Bayesian Model-based Personalized Recommendation Service,” *International Journal of u - and e - Service, Science and Technology*, vol.10, no.7, July, pp.55-64, (2017) DOI:10.14257/ijunnesst.2017.10.7.06
- [4] Woo-Soo Jeong, Seung-Woan Chai, and Kyoungsik Min, “An analysis of the economic effects for the lbs-related industry in Korea,” *International Journal of u - and e - Service, Science and Technology*, vol.10, no.9, pp.145-150, (2017) DOI:10.14257/ijunnesst.2017.10.9.15
- [5] Zhang C and Wang K., “POI recommendation through cross-region collaborative filtering,” *Knowledge & Information Systems*, vol.46, no.2, pp.369-387, (2016)
- [6] Zhang J D, Chow C Y, and Zheng Y., “ORec: An opinion-based point-of-interest recommendation framework,” In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. Melbourne: ACM Press, pp.1641-1650, (2015)
- [7] Burke R., “Hybrid recommender systems: Survey and experiments,” *User Modeling and User-Adapted Interaction*, vol.12, no.4, pp.331-370, (2002)
- [8] Yang S, Huang G, and Xiang Y, et al., “Modeling user preferences on spatiotemporal topics for point-of-interest recommendation,” In *Proceeding of 2017 IEEE International Conference on Services Computing*. Honolulu: IEEE Press, pp.204-211, (2017)
- [9] Son J and Kim S B, “Content-based filtering for recommendation systems using multiattribute networks,” *Expert Systems with Applications*, vol.89, pp.404-412, (2017)
- [10] Wen Y T, Lei P R, and Peng W C, et al., “Exploring social influence on location-based social networks,” *IEEE International Conference on Data Mining*. Shenzhen: IEEE Press, pp.1043-1048., (2014)
- [11] Fang M Y and Dai B R., “Power of bosom friends, POI recommendation by learning preference of close friends and similar users,” *Big Data Analytics and Knowledge Discovery*. Porto: Springer Press, pp.179-192, (2016)
- [12] Lin K, Wang J, and Zhang Z, et al., “Adaptive location recommendation algorithm based on location-based social networks,” *International Conference on Computer Science & Education*. Cambridge: IEEE Press, pp.137-142, (2015)
- [13] Xu X, Zhao P, and Liu G, et al., “A Hybrid method for POI recommendation: combining check-in count, geographical information and reviews,” *Asia-Pacific Web Conference*. Suzhou: Springer Press, pp.162-173, (2016)
- [14] Zheng Z, Chen Y, and Chen S, et al., “Location-Aware POI recommendation for indoor space by exploiting wifi logs,” *Mobile Information Systems*, 2017(9601404), pp.1-16, (2017)

This page is empty by intention.