A Point of Interest Recommendation Approach by Fusing Geographical and Reputation Influence on Location Based Social Networks

Jun Zeng, Feng Li(✉), Junhao Wen, and Wei Zhou

Graduate School of Software Engineering,
Chongqing University, Chongqing, China
{zengjun, lifeng, jhwen, zhouwei}@cqu.edu.cn

Abstract. With the rapid development of location-based social networks (LBSNs), more and more people form the habit of sharing locations with their friends. Point of interest (POI) recommendation is aiming to recommend new places for users when they explore their surroundings. How to make proper recommendation has been a key point on the basis of existing information. In this paper, we propose a novel POI recommendation approach by fusing user preference, geographical influence and social reputation. TFIDF is used to represent user preference. Then, we further improve recommendation model by incorporating geographical distance and popularity. In the dataset, we find friends in LBSNs share low common visited POIs. Instead of directly getting recommendation from friends, users attain recommendation from others according to their reputation in the LBSNs. Finally, experimental results on real-world dataset demonstrate that the proposed method performs much better than other recommendation methods.

Keywords: POI recommendation · Collaborative filtering
Location-based social networks · Geographical influence

1 Introduction

As an extension of traditional social network, location-based social networks (LBSNs) has been booming, e.g., Foursquare and Brightkite [1–8]. Check-in has gradually become a new life style. More and more people form the habit of sharing location with their friends. Then, users generate lots of check-in records. In general, check-in data includes user information, time and location information and reviews [9, 10]. Point of interest (POI) recommendation is aiming to recommend new places which users have not visited before when they explore their surroundings.

POI recommendation recently has been a hot topic. Collaborative filtering (CF) is widely applied to POI recommendation because of simplicity and extendibility. There are much work about memory-based CF [1, 2, 9, 11] and model-based CF [3, 4, 6–8, 10, 13, 14]. For example, Ye et al. [11] incorporated social relationship into recommendation. Friend-based CF (FCF) [1] directly recommends POIs for users from those locations their friends visited before. Compared with traditional user-based CF (user-CF), this method has lower computational overhead. However, the method is limited

I. Romdhani et al. (Eds.): CollaborateCom 2017, LNICST 252, pp. 232–242, 2018.
https://doi.org/10.1007/978-3-030-00916-8_22

because the tastes of one users' friends maybe vary greatly. Some matrix factorization (MF) based methods can easily fuse factors, such as geographical distance and social relation, but a common problem is how to tune the appropriate parameters.

In this paper, we propose a novel method to improve POI recommendation. The main contributions are as follows:

- TF-IDF way is adopted to describe user preference. As a supplement of traditional similarity, TF-IDF demonstrate user preference in a more personalized way.
- Different from existing work, K-medoids method is utilized to cluster user's visited locations to find check-in center for each user.
- Instead of directly generating recommendation from friends, every user generate different contribution degree according to their social reputation in social network.
- We incorporate improved user preference, geographical influence and social reputation into user-CF. Experimental result on the real world dataset show our model outperforms other evaluated methods.

The rest of the paper is organized as follows. Section 2 briefly reviews the related work about the POI recommendation. Section 3 elaborates the proposed method which fuses TF-IDF based user preference, geographical influence and social reputation. Section 4 conducts experiments on real dataset to evaluate the performance of the proposed method. Section 5 concludes this paper.

2 Related Work

The main task of POI recommendation is to recommend new POIs that users may be interested in for users by analyzing their historical behaviors. This section reviews existing POI recommendation techniques on how they employ some factors:

- Reviews & Category. Reviews and category information provide a better understanding about the POIs. However, there are only a few studies having utilized category and reviews information for POI recommendation. Gao et al. [3] studied content information including POI properties, user interest and sentiment indications from users' tips. They use the sentiment scores of tips to infer user interest. Bao et al. [2] modeled user preference with a weighted category hierarchy (WCH) like a tree, then selected local experts for each location category using HITS. Scores for candidate locations are predicted from the opinions of the selected local experts. Hu et al. [8] discovered that users' rating for a specific location is determined by the intrinsic characteristics of the location (review, category and popularity) and the extrinsic characteristics (geographical neighbors).
- Geographical information. Geographical information has a significant impact on human decision making. Yuan et al. [9] assumed human tend to visit nearby POIs to their previous locations and modeled the willingness of visiting a location with a power law distribution. Zhang et al. [18] utilized kernel estimation with an adaptive bandwidth to model the geographical correlation between POIs and estimate the relevance score for users on unvisited locations. Cheng et al. [12] modeled the probability of a user's check-in on a location as Multi-center Gaussian Model.

- Social Influence. Some existing work also considered social influence. Ye et al. [11] argued that friends tend to have similar behavior and friends might provide good recommendation for a given user due to their potential correlated check-in behavior. Cheng et al. [12] fused MF with social influence for POI recommendation. However, experimental results on [12] demonstrated social influence is not so important because of low common visited POIs between friends. In this paper, we adopt a new method to exploit the social reputation, not simple friend link.

Our study differentiates itself from these existing work in some aspects. First, TF-IDF technology is adopted to represent personalized user preference. Second, we use K-medoids method to cluster user's visited locations to find check-in centers for each user, not common center [12]. Then, geographical influences including distance and location popularity are not incorporated by naïve Bayes [9]. Finally, we use social reputation to measure users' contribution degree, not simple social link in [11, 16, 18].

3 Fused Method with User Preference, Geographical Information and Social Reputation

This section first defines the problem, and then presents a unified framework to perform POI recommendation.

3.1 Problem Definition

Let $U = \{u_1, u_2, u_3, \ldots, u_{|U|}\}$ and $L = \{l_1, l_2, l_3, \ldots, l_{|L|}\}$ be the user set and location set, where $|U|$ and $|L|$ denote the number of users and locations. Given users' check-in data (U and L) and the corresponding social relationship S, the task of POI recommendation is defined as in (1). We calculate score for locations that user never visited before, and return a ranked list of candidate POIs to users.

$$U \times L \times S \rightarrow Score \tag{1}$$

3.2 TF-IDF Based Similarity Between Users

User-CF finds similar users based on a similarity measure. Then scores on items are calculated by a weighted combination of historical ratings from similar users. We can easily get a $|U| \times |L|$ check-in matrix C from check-in records. $C_{u,l}$ means check-in frequency that user $u \in U$ has checked in $l \in L$. If $C_{u,l} = 0$, it means user u has never visited location l. The recommendation score that user u will check in a location l is denoted as $Score_{u,l}$ as in (2), where $Sim_{u,v}$ is the similarity between user u and user v.

There are many ways to measure similarity weight between users, such as cosine similarity and Pearson's correlation coefficient. We adopt the widely used cosine similarity $Sim_{u,v}$ between users u and user v as in (3).

$$Score_{u,l} = \frac{\sum_{v} Sim_{u,v} C_{v,l}}{\sum_{v} Sim_{u,v}} \tag{2}$$

$$Sim_{u,v} = \frac{\sum_{l \in L} C_{u,l} C_{v,l}}{\sqrt{\sum_{l \in L} C_{u,l}^2} \sqrt{\sum_{l \in L} C_{v,l}^2}} \tag{3}$$

Some works used visited/unvisited way [9, 11] or check-in frequency [2, 3, 7] to represent user preference. The former only shows whether users visited a location or not. The latter roughly indicates user preference. These methods can't really demonstrate user preference precisely. In this case, we use TF-IDF to describe user preference in a fine-grained way.

TF-IDF [19] is widely used in information retrieval, document classification and other related fields. It measures the importance of a word to a document, just like the importance of visited locations to users. On the one hand, the importance of the word is proportional to the frequency of the word (TF) in the document. Meanwhile, the importance of the word is inversely proportional to the frequency in the corpus (IDF). Similarly, the more frequently user u checks in location l, the more important the location l is to user u as in (4). If less people visit location l, but user u visits location l individually, it means user u prefers location l relatively as in (5). We assume TF-IDF can represent personalized preference as in (6), not in a generalized way.

$$TF_{u,l} = \frac{total_{u,l}}{total_u} \tag{4}$$

$$IDF_l = log\left(\frac{|U|}{total_l} + 1\right) \tag{5}$$

$$TFIDF_{u,l} = TF_{u,l} IDF_l \tag{6}$$

where $total_{u,l}$ is check-in frequency for user u in location l, $total_u$ is the total number of check-ins for user u. $total_l$ is the total check-in frequency in location l, $TFIDF_{u,l}$ is the importance of POI l to user u.

Cosine similarity between users is improved by TF-IDF as in (7). An interesting phenomenon is that $Sim_{u,v}$ is not equal to $Sim_{v,u}$, which is greatly different from traditional similarity. As mentioned above, we use TF-IDF to represent the importance of locations for users. When measuring similarity of pairwise users, different locations generate different contribution according to their importance for users, not always the same. If location l is more important for user u, location l should be endowed with a larger weight when measuring similarity between others and u. In fact, this is a more personalized way, because we consider user real preference.

$$Sim_{u,v} = \frac{\sum_{l \in L} C_{u,l} C_{v,l} TFIDF_{u,l}}{\sqrt{\sum_{l \in L} C_{u,l}^2} \sqrt{\sum_{l \in L} C_{v,l}^2}} \tag{7}$$

3.3 Social Reputation Influence for Recommendation

In Brightkite dataset [14], we find the average ration of common visited locations between friends is only 0.45%. It implies that less than 1% locations are commonly visited by friends and friends' preference about POIs may vary greatly. Therefore, we think it is unreasonable that some researchers directly generate recommendation for users from limited friends [1], or require target users' user-specific latent vector are closed to their friends [16, 17].

Empirical result demonstrates users tend to seek advice from people with high reputation. We assume users' reputation has an effect on recommendation result. Users with high reputation generate more contribution than those who have less reputation.

There are some link analysis approaches to measure weight of nodes in networks, such as PageRank, Hyperlink - Induced Topic Search (HITS) and TrustRank. Google Search uses PageRank to assign a numerical weighting to web pages to present their relationship and importance according to hyperlinks between them. We adopt widely used PageRank to calculate the weight level of user u, namely user reputation score $PageRank_u$. We normalize each reputation as in (8), where Rep_u is the social reputation of user u.

$$Rep_u = \frac{PageRank_u}{max(PageRank_{u \in U})} \tag{8}$$

3.4 Geographical Influence for Recommendation

When recommending locations to users in mobile environment, geographical distance is a very important factor. Meanwhile, popularity of locations should also be considered. In this section, the concerned geographical influence includes distance and popularity. For simplicity, the effect of geographical distance dis is denoted as $g(dis)$. Some researchers considered the relationship between check-in probability and geographical distance follow the power-law distribution [1, 9, 11] as in (9) where a and b are

$$g(dis) = a \times dis^b \tag{9}$$

parameters of a power-law distribution. Others think the relationship between probability of check-in and geographical distance follow inverse proportion [12] as in (10) or exponential function [15] as in (11). The common basic idea is almost consistent as follow: human tends to visit nearby POIs, and check-in probability of visiting a POI decreases as distance increases. Thus, we try to find the relatively appropriate method.

$$g(dis) = \frac{1}{dis} \tag{10}$$

$$g(dis) = e^{-dis} \tag{11}$$

Users always prefer visit nearby POIs close to some check-in centers like home or office [8, 11, 14]. In this case, an important work is to find check-in centers for each user. In order to avoid check-in outliers, we choose k-medoids method that is different from [12] to cluster locations user u visited and find the check-in centers for user u, like office and home. Nearby POIs closed to the center are good choices when recommending POIs for users.

Generally, the popularity of POIs also affects user decision making. A popular POI could provide better user experience in some aspects, so we adopt method in [6] to measure popularity of POI as in (12) where $p(l)$ is the popularity of location l, $totalCk_l$ is the total check-in frequency in location l, $totalPeo_l$ is the number of people who checked in location l.

$$p(l) = \frac{1}{2}\left\{\frac{totalCk_l}{max(totalCk_{l \in L}) - 1} + \frac{totalPeo_l}{max(totalPeo_{l \in L}) - 1}\right\} \tag{12}$$

3.5 Unified Framework for POI Recommendation

Each factor mentioned above, such as TF-IDF based user preference, geographical information, social reputation, can be utilized to improve POI recommendation. Naturally, we proposed a unified framework named TSG to integrate these factors. Let $Score_{u,l}$ denotes the check-in score of user u at location l as in (13), where $dis_{u,l}$ is the nearest geographical distance between POI l and u's check-in center, such as user u's home or office.

It is worth mentioning that the product rule has been widely used to fuse different factors for POI recommendation in the previous work [6, 11, 12, 18] and has shown high robustness. In this framework, we calculate check-in score of unvisited locations and return a top-N POIs list for user u.

$$Score_{u,l} = \frac{\sum_{v} Sim_{u,v} C_{v,l} g(dis_{u,l}) p(l) Rep_v}{\sum_{v} Sim_{u,v}} \tag{13}$$

4 Experimental Evaluation

In this section, we design and conduct several experiments to compare the recommendation qualities of the proposed method with other CF methods.

4.1 Dataset Analysis and Metric

We set a bounding box and extract 1422625 Brightkite check-ins from the dataset in [14]. The dataset is crawled from online LBSNs—Brightkite from Apr. 2008 to Oct. 2010. Check-in dataset includes user ID, location ID, longitude and latitude of POI, check-in timestamp and social relationship. To reduce noise data, we remove users who have fewer than 5 check-ins and locations which is less than 5 check-ins. The check-in density of the dataset is 1.36×10^{-3}. We randomly choose 70% as the training set and the remaining 30% as the testing set.

To evaluate the performance of the proposed method, we use precision Pre@N and recall Rec@N as the evaluation metric. Pre@N means the ratio of hit POIs to the recommendation list. Rec@N means the ratio of hit POIs to the ground truth. N is the length of the recommendation list. In our experiment, we set N = 5, 10, 20. To clearly compare with other methods, we focus on the relative improvements we achieved, instead of the absolute values.

4.2 Experimental Results

Performance Comparison Between TFIDF-Based Similarity and Cosine Similarity. In Sect. 3.2, we think the importance of POIs to users has an effect on similarity between users. Thus we use TF-IDF to enhance traditional user preference. In order to verify whether TF-IDF efficiently improves the performance of user-CF, we compare the recommendation results from traditional cosine similarity in (3) and TFIDF-based similarity in (7).

The precision and recall for them are plotted in Fig. 1 on top 5/10/20. In these figures, TF-IDF based way always exhibits the better performance under all values of N. It means TF-IDF based similarity can improve recommendation. And it describes user preference in a more accurate and personalized way.

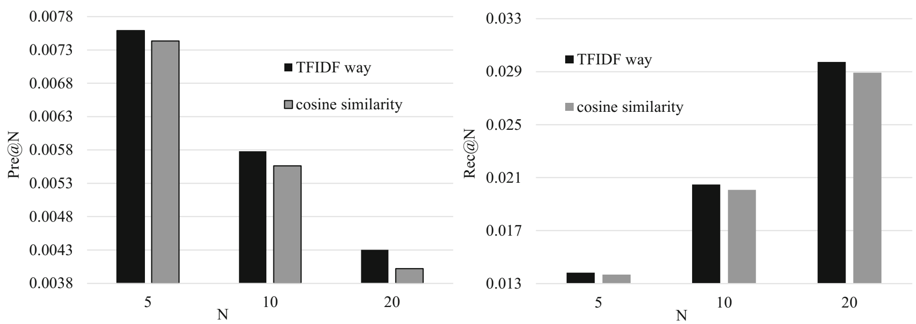

Fig. 1. Performance about different similarity

Performance on Different Methods Modeling Geographical Distance Influence. In Sect. 3.4, there are three methods to model geographical distance in POI recommendation as follow: power law distribution, exponential function and inverse proportion. We respectively incorporate these three methods into user-CF as in (14).

$$Score_{u,l} = \frac{\sum_{v} Sim_{u,v} C_{v,l} g(dis)}{\sum_{v} Sim_{u,v}} \tag{14}$$

Figure 2 shows the compared performance among different models on top 5/10/20 recommendation. We observe the inverse proportion is always perform better than other methods in terms of both Pre@N and Rec@N. It is reasonable since the probability of user checking in location to distance presents relatively slow descending tendency.

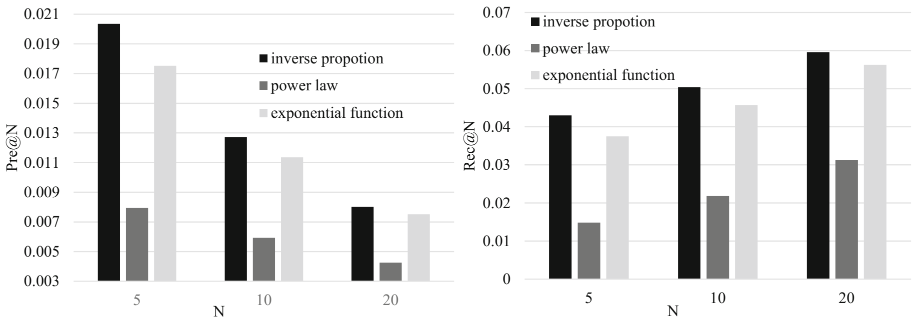

Fig. 2. Performance for different models of geographical distance

Comparison with Other POI Recommendation Approaches. Three factors, TF-IDF based user preference, geographical influence and social reputation, are incorporated into our unified models, denoted by TSG. We compare the proposed method with the following CF methods:

User-CF: the classical method is widely used in all kinds of applications.
FCF: In this way, users only get recommendation from their friends in LBSNs [1].
GM-FCF: this is the development of FCF. The method uses power law distribution to model geographical distance between friends [1].
USG: Three factors, namely user preference, social influence and geographical influence, are incorporated into unified CF [11].

Figure 3 reports the comparison results of top 5/10/20 recommendation with the baseline methods. TSG outperforms other methods significantly in all metrics. For example,

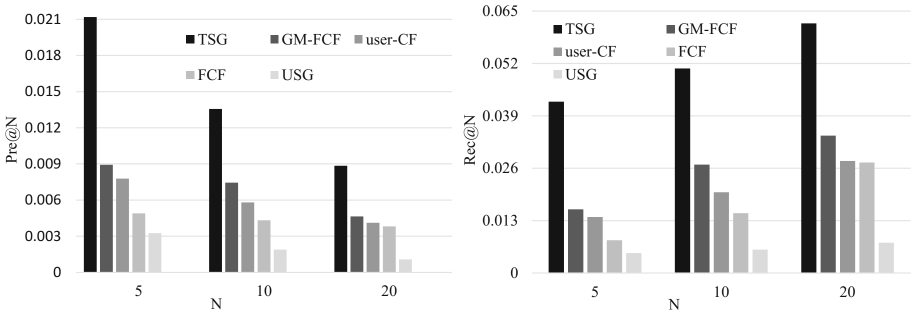

Fig. 3. Performance comparison

TSG attains 0.023 Pre@5, while user-CF achieves 0.0078. This implies that TFIDF-based user preference, geographical influence and social reputation can improve POI recommendation. TSG achieves much significantly better than FCF and GM-FCF. In other word, getting recommendation from users with high reputation performs better than directly from friends. Because friends in LBSNs may be strangers, and their preference may vary greatly. Users can't ensure whether their relationships are reliable. Seeking advice from those who have high reputation is confirmed in real life.

5 Conclusion and Future Work

The paper proposes a unified framework which fuses TF-IDF based user preference, geographical influence and social reputation in POI recommendation. TF-IDF technology is utilized to describe personalized user preference. Furthermore, we model geographical influence including geographical distance and popularity, then generate recommendation for users according to other's reputation in LBSNs. Experiment results on real dataset show that the proposed method performs better than other methods.

There are several directions to investigate in the future. First, information in LBSNs has not been fully utilized, such as user generated content (UGC), detailed POI information and social activities. Then, user influence should be determined by their activities on the social network rather than just friends. These information will be analyzed in our future work. We try to find and ensure the efficiency of temporal influence in POI recommendation.

Acknowledgement. This research is supported by the National Natural Science Foundation of China (Grant No. 61502062, Grant No. 61672117 and Grant No. 61602070), the China Postdoctoral Science Foundation under Grant 2014M560704, the Scientific Research Foundation for the Returned Overseas Chinese Scholars (State Education Ministry), and the Fundamental Research Funds for the Central Universities Project No. 2015CDJXY.

References

1. Ye, M., Yin, P., Lee, W.-C.: Location recommendation for location-based social networks. In: 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 458–461. ACM Press, New York (2010)
2. Bao, J., Zheng, Y., Mokbel, M.F.: Location-based and preference-aware recommendation using sparse geo-social networking data. In: 20th International Conference on Advances in Geographic Information Systems, pp. 199–208. ACM Press, New York (2012)
3. Gao, H., Tang, J., Hu, X., Liu, H.: Content-aware point of interest recommendation on location-based social networks. In: 29th AAAI Conference on Artificial Intelligence, pp. 1721–1727. AAAI Press, Menlo Park (2015)
4. Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., Rui, Y.: GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 831–840. ACM Press, New York (2014)
5. Liu, Q., Ma, H., Chen, E., Xiong, H.: A survey of context-aware mobile recommendations. Int. J. Inf. Technol. Decis. Mak. **12**, 139–172 (2013)
6. Liu, B., Fu, Y., Yao, Z., Xiong, H.: Learning geographical preferences for point-of-interest recommendation. In: 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1043–1051. ACM Press, New York (2013)
7. Gao, H., Tang, J., Hu, X., Liu, H.: Exploring temporal effects for location recommendation on location-based social networks. In: 7th ACM Conference on Recommender Systems, pp. 93–100. ACM Press, New York (2013)
8. Hu, L., Sun, A., Liu, Y.: Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In: 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 345–354. ACM Press, New York (2014)
9. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Time-aware point-of-interest recommendation. In: 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 363–372. ACM Press, New York (2013)
10. Yuan, Q., Cong, G., Sun, A.: Graph-based point-of-interest recommendation with geographical and temporal influences. In: 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 659–668. ACM Press, New York (2014)
11. Ye, M., Yin, P., Lee, W.-C., Lee, D.-L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: 34th International ACM SIGIR Conference on Research and Development in Information, pp. 325–334. ACM Press, New York (2011)
12. Cheng, C., Yang, H., King, I., Lyu, M.R.: Fused matrix factorization with geographical and social influence in location-based social networks. In: 26th Conference on Artificial Intelligence, pp. 17–23. AAAI Press, Menlo Park (2012)
13. Tang, J., Hu, X., Gao, H., Liu, H.: Exploiting local and global social context for recommendation. In: 23rd International Joint Conference on Artificial Intelligence, pp. 2712–2718. AAAI Press, Menlo Park (2013)
14. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1082–1090. ACM Press, New York (2011)
15. Liu, S.D., Meng, X.W.: Approach to network services recommendation based on mobile users' location. J. Softw. **25**, 2556–2574 (2014). (in Chinese)
16. Feng, Y., Li, H., Chen, Z.: Improving recommendation accuracy and diversity via multiple social factors and social circles. Int. J. Web Serv. Res. **11**, 32–46 (2014)

17. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: 4th ACM International Conference on Web Search and Data Mining, pp. 287–296. ACM Press, New York (2011)
18. Zhang, J.-D., Chow, C.-Y.: GeoSoCa: exploiting geographical, social and categorical correlations for point-of-interest recommendations. In: 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 443–452. ACM Press, New York (2015)
19. Spertus, E., Sahami, M., Buyukkokten, O.: Evaluating similarity measures: a large-scale study in the orkut social network. In: 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 678–684. ACM Press, New York (2005)