Somenath Chakraborty

Computing Sciences and Computer Engineering, The University of Southern Mississippi, Hattiesburg, USA Somenath.Chakraborty@usm.edu

DOI: http://dx.doi.org/10.56828/jser.2022.1.2.3 Article history: Received (January 30, 2022); Review Result (March 4, 2022); Accepted (May 15, 2022)

Abstract: This research suggests a point-of-interest recommendation model that incorporates social interactions and some geographical parameters in order to address the issue of data sparsity in Point-of-Interest (POI) recommendation. The model first suggested a user similarity measurement method based on the shared sign-in and distance between users in the social relationship, and based on the collaborative filtering of users, it was possible to determine the degree to which the social relationship influenced the user's sign-in behavior. Second, divide the local activity area for each user, fully evaluate the check-in correlation between the nearby POIs that haven't been visited, and determine the extent to which the user's check-in behavior is influenced by the surrounding geography. Finally, a POI recommendation model that is more in line with the user's preference is constructed, which effectively solves the problem of data sparsity. It is based on the weighted matrix decomposition, the user's preference model and combined with the social relationship and local geographic factors considered above. Experiments on the Gowalla dataset show that the proposed SLGMF algorithm has good recommendation performance.

Keywords: Social relationships, Local geographic factors, Point-of-Interest recommendation, Recommendation model

1. Introduction

At present, with the development of technologies such as smartphones and mobile Internet, Location-based Social Networks (LBSN) have been widely used, such as Gowalla and Foursquare. On the LBSN platform, users can sign in and comment on the currently visited POI (such as restaurants, bookstores, tourist attractions, etc.), and can share their sign-in information with social friends. In addition to user sign-in information, LBSN also contains massive data such as social relationships, points of interest, user comments, etc. Fully mining these data can better analyze user sign-in behavior, grasp user group interest characteristics and access rules, and help improve personalization POI recommends the quality of service. On the LBSN platform, POIs are large in scale and widely distributed, and there are often very few POIs that each user has visited and checked in. This makes user check-in data very sparse and poses challenges for POI recommendations.

Aiming at the sparseness of user sign-in data, this paper proposes a POI recommendation algorithm that integrates effective data such as social relationships and geographic location information to improve recommendation quality. The main contributions of this article are as follows:

(1) When constructing the social relationship influence model, comprehensively consider the common sign-in and distance relationship between users, and propose a user similarity measurement method to reduce similarity calculation errors.

(2) When constructing the geographical factor influence model, by dividing a local activity area for each user, analyzing the sign-in correlation among the unvisited POIs in the area where the user belongs, and more fully analyzing the geographical preference of each user's sign-in.

(3) When recommending points of interest, the social relations and local geographic factors considered above are integrated into the weighted matrix factorization model to build a POI recommendation model that is more in line with user preferences, which alleviates the problem of data sparsity to a certain extent. Experiments show that the algorithm proposed in this paper has a higher accuracy and recall rate.

2. Basic Theory

A crucial personalized service offered by LBSN is the POI recommendation feature, which tries to provide users with a list of POIs they may be interested in but have not yet visited. At present, a lot of research work has been carried out on POI recommendations for location social networks, and some specific scenarios have also begun to emerge in large numbers, such as restaurant recommendations [1], travel route recommendations, etc. [2].

At present, most POI recommendations are based on user sign-in data combined with rich contextual information to alleviate data sparseness, and to mine users' preference for POIs that have not yet been accessed. For example, Lapidoth [3] models the distance distribution between POIs using a multi-center Gaussian distribution and adds social links to generate POI suggestions. In order to prevent inaccuracies brought on by the uniform distribution of all users and to construct a model for the influence of geographic factors, Helu et al [4] employed kernel density estimation to mimic the distance distribution between any two POIs. In order to combine sequence influence, social influence, and geographic factor influence into a single recommendation framework, Mathonat et al [5] mine sequence patterns from the position sequence. Yali et al., [6] proposed a personalized point-of-interest recommendation model based on collaborative filtering based on friends, comprehensively considering social relationships and geographic location characteristics.

In addition, there are also many studies using other contextual information for POI recommendation, such as Gulam et al., [7] for semantic analysis and emotional calculation of user reviews, and predict user preferences based on users' emotional tendencies. Strickland et al. [8] integrate POI popularity characteristics into user-based collaborative filtering methods and combine social relationships and geographic factors to make POI recommendations.

Although the fusion of rich contextual information can effectively alleviate the problem of data sparsity, the existing POI recommendation algorithms that integrate contextual information still have the following problems: When studying the influence of social relationships, the selection of factors affecting user similarity is often relatively simple. It is difficult for users with less sign-in information to construct a social influence model; when analyzing the influence of geographic factors, the existing methods for calculating the

geographic correlation between global POIs cannot fully analyze the geographic preferences of each user. Therefore, this paper proposes a POI recommendation algorithm that integrates social relationships and local geographic factors, which can effectively improve the accuracy and recall of recommendations.

3. Point of Interest Recommendation Model

This chapter analyzes the influence of social relationships and geographic factors on the user's check-in behavior and builds a POI recommendation model that is more in line with user preferences.

3.1. Related symbols and definitions

Table 1 shows the lists of relevant symbols and their meanings for the convenience of follow-up discussion. At the same time, the sign-in matrix and sign-in hot spots are defined.

Symbol	Related description
U, I	Collection of all users, collection of all POIs
u, i	User $u \in U$, POI $i \in I$
R	User sign-in matrix
С	The 0, 1 matrix corresponding to the sign-in matrix R
S	User's social preference matrix for POI
G	User's geographic preference matrix for POI
Е	User's final preference matrix for POI
Р	User hidden feature matrix
Q	POI hidden feature matrix
K	Hide feature matrix dimensions
W	The weight matrix corresponding to the sign-in matrix R

Table 1: Description of related symbols

Definition 1 (check-in matrix) According to the user's historical check-in records, a checkin matrix $R_{|U| \times |I|}$ is constructed. Each element r_{ui} in the matrix represents the number of times user *u* has checked in toPOI*i*.

Definition 2 (check-in hotspot) the POI with the highest number of check-ins by user u.

3.2. Construction of social relationship influence model

In LBSN, users' check-in behavior is influenced to some part by their social contacts, which pushes users to visit POIs they haven't visited before. For example, after going to a restaurant or bookstore with a good experience in daily life, you are likely to invite or recommend your friends to go. These may cause users to sign in at the same POI, and there are certain behaviors among social users. Therefore, considering the influence of the user's social relationship on the user's check-in behavior is beneficial to improving the quality of POI recommendations.

For the POIs that the user has not visited, this paper uses the user-based collaborative filtering method to calculate the degree of influence of social relationship information on the user's check-in behavior. The calculation formula is as follows:

$$s_{ui} = \frac{\sum_{v \in F_u} sim_{uv} \times c_{vi}}{\sum_{v \in F_u} sim_{uv}} \#$$
(1)

Where the definition of $c_{\nu i}$ is as follows:

$$c_{vi} = \begin{cases} 1, r_{vi} > 0\\ 0, r_{vi} = 0 \end{cases} \#$$
(2)

Where s_{ui} represents user u's preference for POI*i*. F_u represents the collection of all friends of user u on the social network, user v is the social friend of user u, c_{vi} represents the check-in status of friend v to POIi, and sim_uv represents the similarity between users u and v.

For the user similarity \sin_{uv} in formula 1, different studies use different methods [3][4][5][6]. In most studies, only the similarity of users based on common sign-in is considered, which cannot effectively calculate their social preferences for users who have less sign-in information. Therefore, this paper considers the previous similarity research based on common sign-in and at the same time measures the difference of interest between users through the distance relationship between users, and reduces the calculation error of user similarity.

On the one hand, it is generally believed that the more common check-ins between users, the more similar their preferences are. Among the similarity measurement methods based on behavioral relevance, the Jaccard method is more effective in calculation efficiency and recommendation accuracy. The formula for calculating user check-in similarity based on the Jaccard method is as follows:

$$sim_c = \frac{|R_u \cap R_v|}{|R_u \cup R_v|} \#$$
(3)

Where R_u and R_v represent the set of POIs signed in by users u and v, respectively.

On the other hand, considering that users who are closer to each other have more shared interest points, more common sign-ins will be generated. Therefore, for closer users, the similarity of their check-in behavior is higher. This paper calculates the similarity of user distance by improving the Sigmod function.

The Sigmod function is a common sigmoid function, and the formula is defined as follows:

$$S(x) = \frac{1}{1 + e^{-x}} \#$$
(4)

The function is continuous, smooth, strictly monotonic, symmetric about the center of the point (0,0.5), and grows nonlinearly on the interval $[-\infty, +\infty]$. Based on the above assumptions, the distance similarity should decrease with the distance between users, so consider improving the Sigmod function to measure the similarity based on the user distance. The improved user distance similarity calculation formula is as follows:

$$sim_d = \frac{2}{1 + e^{dis(u,v)}} \#$$
 (5)

Where dis(u, v) represents the distance between user u and v, which is calculated by the Haversine method [16] according to the longitude and latitude information of the user's sign-in hotspot. Calculated as follows:

$$hav\left(\frac{dis(u,v)}{R}\right) = hav(\varphi_2 - \varphi_1) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot hav(\sigma_2 - \sigma_1) \#$$
(6)

$$hlav(\theta) = \frac{1 - \cos\theta}{2} \# \tag{7}$$

Where R is the radius of the earth, φ_1 and φ_2 represent the latitudes of users u and v signing in to the hotspots respectively, and σ_1 and σ_2 represent the longitudes of signing hotspots with u and v respectively.

Based on the above design, this paper linearly weights the user sign-in similarity and distance similarity to obtain the final similarity of users' u and v. The calculation formula is as follows:

$$sim_{uv} = (1 - \alpha)sim_c + \alpha sim_d \#$$
(8)

Where α is a parameter that adjusts the weight of user sign-in similarity and distance similarity.

Substituting formula (8) into formula (1) can obtain the user's social preference matrix S for points of interest under the influence of social relationships.

3.3. Local geographic factors affect the model construction

In location social networks, geographic location information is its unique contextual information, and it is an important factor to improve the quality of POI recommendations. At present, the model that considers the influence of geographical factors mainly analyzes the distribution of the entire user's sign-in POIs, and calculates the user's sign-in probability to the POIs through the distance relationship between the POIs, such as the commonly used Gaussian distribution [3] and kernel density estimation. To further study the influence of geographic factors, it is found that user sign-in shows a spatial clustering phenomenon, indicating that users like to visit neighboring POIs [9]. If users travel to a certain city, they will often visit all POIs concentrated in adjacent areas. The check-in situation between neighboring POIs has certain relevance [10]. Therefore, this paper considers dividing a local activity area for each user, fully analyzing the sign-in correlation between the unvisited POIs in the area, and constructing a local geographic factor influence model that is more in line with user preferences.

Step 1: Divide the user activity area

First, find the sign-in hotspot of user u, and then divide a local active area with a radius of β from the sign-in hotspot for user u.

Step 2: Look for POIs in the area that users have not yet visited

In the user activity area divided in step 1, look for the set of POIs that user u has not visited, and analyze the sign-in situation of neighbor POIs whose distance from POI i in the set is less than γ .

Step 3: Calculate the degree of geographic preference

Studies have found that users tend to have centralized access to neighboring POIs, and related research has been launched on the impact of neighboring POIs' checks. Where Hossein et al. [11] believed that the number of users sign-ins to neighboring POIs would have a certain impact on users u sign in to POI and constructed a geographical preference model

that is more in line with user preferences. Therefore, according to the sign-in situation of neighbor POIs, this paper adopts the geographical preference definition in Bokde et al. [12] to calculate the sign-in probability of user u forPOI*i*, the formula is as follows:

$$g_{ui} = 1 - \frac{N_i^u}{|L_u|} \#$$
(9)

Where N_i^u represents the number of check-ins of all neighbor POIs of POIi by user u, and L_u represents the set of all POIs that user u has visited.

3.4. A point-of-interest recommendation model integrating social relationships and local geographic factors

Point of Interest Recommendation In LBSN, contextual factors like social connections and location will have an impact on a user's check-in behavior, but the most crucial factor is the user's preferences as shown by their prior check-in data. Matrix factorization technique is frequently utilized in POI suggestions to mine user preferences. The User-POI check-in matrix will, however, only display the POIs that the user has visited, leading to a significant number of missing items for those POIs. Therefore, the user's preference model cannot be created using the general matrix factorization method. Hu et al., [13] first proposed a Weighted Matrix Factorization (WMF) model on large-scale implicit data. This model assigns a larger weight to the POIs that the user has visited, and a smaller one to the POIs that have not yet been visited.

In the POI recommendation, the WMF method decomposes the User-POI check-in matrix into two low-dimensional hidden feature matrices $P_{m \times k}$ and $Q_{n \times k}$. The users' preference matrix for POI is the two hidden feature matrices. This article builds a user preference model based on the WMF model, and its objective function is defined as follows:

In POI recommendation, the WMF method decomposes the User-POI check-in matrix into two low-dimensional hidden feature matrices $P_{m \times k}$ and $Q_{n \times k}$, and the user's preference matrix for POI is the two hidden feature matrices. This paper builds the user's preference model based on the WMF model, and its objective function is defined as follows:

$$f = \sum_{u=1}^{|U|} \sum_{i=1}^{|I|} w_{ui} (c_{ui} - p_u^T q_i)^2 + \lambda (\|p_u\|^2 + \|q_i\|^2) \#$$
(10)

$$c_{ui} = \begin{cases} 1, r_{ui} > 0\\ 0, r_{ui} = 0 \end{cases} \#$$
(11)

Where p_u and q_i respectively represent the hidden feature vector corresponding to user u and POIi, and λ represents the regularization coefficient. w_{ui} represents the weight corresponding to the number of check-ins r_{ui} for each user. This paper adopts the weight setting method in the Zhao et al., [14], and the weight coefficient of the weighting matrix w is set as follows:

$$w_{ui} = 1 + lb\left(1 + \frac{r_{ui}}{\varepsilon}\right) \#$$
(12)

Where ε is used to control the rate at which the weight coefficient increases with the number of times the user visits the POI.

The WMF model extracts the user's preference for POI based on the user's check-in frequency information but does not consider the influence factors such as social relationships and geographic location. This paper improves on the WMF model, integrates the social relations and local geographic information considered above into the WMF model, and builds a POI recommendation model SLGMF that is more in line with user preferences. The objective function of this model is shown in equation (13):

$$f = \sum_{u=1}^{|U|} \sum_{i=1}^{|I|} w_{ui} (c_{ui} - e_{ui})^2 + \lambda (\|p_u\|^2 + \|q_i\|^2) \#$$
(13)

$$e_{ui} = p_u^T q_i + s_{ui} + g_{ui} \#$$
(14)

Where e_{ui} represents user u's final preference for POI*i*, $||p_u||^2$ and $||q_i||^2$ represent the regularization items of user u and POI*i* respectively, and the purpose of adding is to prevent overfitting.

In this paper, the objective function is optimized by the alternating least squares method. The main idea is to first fix the matrix P in turn, optimize the matrix Q, and then fix the matrix Q, optimize the matrix P, alternate repeatedly, continuously modify the p and q components, and obtain the optimal matrix P And matrix Q.

4. Experimental Results

4.1. Data set

The experiment selected the data set Gowalla [15]. Gowalla collected all check-in data, user social relationships, and POI information from April to December 2012. To make the experimental data more effective, the low-value data with less than 20 users and POI check-ins in the data set are removed. The processed standard data set has a total of 5318 users, 29803 POIs, 590683 sign-in records, and 42007 social relationships.

4.2. Evaluation index

The experiment selects the commonly used accuracy and recalls as evaluation indicators to measure the performance of the recommendation algorithm. The definitions are as follows:

$$Precision = \frac{1}{/U} \sum_{u=1}^{/U/} \frac{/R_u \cap T_u}{/R_u} \#$$
(15)

$$Recall = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{|R_u \cap T_u|}{|T_u|} \#$$
(16)

Where |U| represents the number of users, Ru represents the list of POIs recommended by the user u, and Tu represents the list of POIs that the user u has visited.

4.3. Comparison method and parameter setting

To verify the performance of the proposed SLGMF algorithm, the experiment selects two POI recommendation algorithms that integrate social relationships and geographic factors: MGMPFM [3] and LORE [5]. In addition, to examine the impact of different factors on POI recommendation, this paper splits the SLGMF algorithm into a recommendation algorithm S-MF that only integrates social relationships, and a recommendation algorithm LG-MF that only integrates local geographic factors, and compares it by itself.

MGMPFM: Integrate social and geographic data based on probability matrix decomposition for POI recommendation. Multi-center Gaussian model the impact of geographic elements.

LORE: In order to recommend POIs, sequence patterns from the user's sign-in positions are mined, together with social and geographic criteria.

S-MF: The recommendation algorithm described in this paper's model solely incorporates data on social relationships.

LG-MF: This paper proposes a recommendation algorithm model that only incorporates geographic location data.

SLGMF: The model put out in this work is a recommendation system built on a weighted matrix decomposition and thorough analysis of local geographic and sociological parameters.

In terms of parameter settings, the selected comparison model should be as consistent as possible with the original literature to maximize the performance of each algorithm. For the SLGMF model proposed in this paper, the hidden feature matrix dimension K = 25, the user similarity influence factor $\alpha = 0.6$, the geographical factor influence factor $\beta = 30$, and $\gamma = 12$ are respectively set to make the recommendation effect optimal. In addition, the average radius *R* of the earth is 6371.

4.4. Experimental results and analysis

In this paper, Top-N (N=5, 10, 15, 20, 25, 30) recommendations are performed on each POI recommendation algorithm on the same data set. Figures 1 and Figure 2 show the accuracy and recall rates of Top-N recommendations. It can be seen that the recall rate grows as N increases but the accuracy rate falls. This is because the more POI returned, the more likely it is to find more POIs that users are willing to visit.



Figure 2: Recommendation recall

Figures 1 and 2 show that, regardless of the number of N, the SLGMF algorithm has enhanced accuracy and recall, demonstrating the efficacy of the SLGMF algorithm described in this paper. These two POI recommendation algorithms, MGMPFM and LORE, incorporate social interactions and geographic features.

This study further divides the SLGMF model into two POI recommendation models: one that only incorporates social information and the other that only incorporates geographic location information. Experiments are used to study the recommendation effect of fusing single context information. The results are shown in Figure 1 and Figure 2. When performing Top-5 recommendations for users, the accuracy of the SLGMF algorithm is improved by

14.6% relative to the S-MF algorithm, and 11.2% relative to the LG-MF algorithm. In the recall rate, the SLGMF algorithm has increased by 24.3% compared to the S-MF algorithm and has increased by 22.8% compared with the LG-MF algorithm. This shows that the integration of multiple contextual information based on weighted matrix decomposition in this paper can indeed effectively alleviate the problem of data sparsity and improve recommendation performance, and the more effective contextual information is incorporated, the better the recommendation effect. In addition, comparing the S-MF and LG-MF algorithms, it can be found that the recommendation performance of LG-MF that only integrates geographic factors is better than that of S-MF that only integrates social factors.

4.5. Influence of model parameters

When constructing the user's preference model, the user's sign-in matrix is decomposed into two hidden feature matrices P and Q with dimension K through WMF. The value of parameter K determines the user's preference for POI. This article selects different K values and recommends Top-5 to users. The recommendation results are shown in [Figure 3]. It can be seen that when the dimension K=25, the SLGMF recommendation performance is the best, and with the continuous increase of K, the recommendation effect begins to decline, which is because the model has been over-fitting.



Figure 3: Effect of parameters k on precision and recall

When calculating the final similarity of users in formula (8), the parameter α determines the degree of influence of distance similarity and sign-in similarity. To explore the influence of the distance between users and the common sign-in on user sign-in behavior, the experiment selects different α values to make a Top-5 recommendation for user u, and the result is shown in Figure 4. It can be seen that when $\alpha = 0.6$, the recommendation effect is the best, and it also shows that the similarity of user check-in behaviors is more affected by the distance between users.

When constructing the influence model of local geographic factors, each user is divided into a local activity area. The parameter β is used to control the user's activity range, and the parameter γ is used to identify neighbor POIs in the area that has not yet visited the POI I. Therefore, to explore the rules of users' local sign-in, the experiment selects different β and γ for Top-5 recommendations and analyzes the influence of local geographic factors on the user's sign-in behavior. The recommendation results are shown in Figure 5 and Figure 6, respectively. It can be seen that when $\beta = 30$ and $\gamma = 12$, the recommendation effect is the best. It can also be seen that users tend to visit POIs that are closer to them.



Figure 4: Effect of parameter α on precision and recall



Figure 5: Effect of parameters β on precision and recall



Figure 6: Effect of parameters γ on precision and recall

6. Conclusion

Point-of-Interest (POI) suggestion is a crucial personalized service offered by Location-Based social networks (LBSN), which can assist users in finding POIs that interest them and raise the standard of information services. This research suggests a POI recommendation algorithm that incorporates social interactions and local geographic features to address the issue of data sparsity in POI selection. A social influence model is created based on the collaborative filtering approach of users and measures user similarity based on shared sign-ins and the distance between users in social interactions. Divide a local activity area for each user, and establish the influence model of local geographic factors by analyzing the sign-in correlation among POIs in the area. Making POI suggestions involves mining user preferences based on weighted matrix decomposition and incorporating social connections and regional geographic characteristics. Studies reveal that the suggested POI recommendation algorithm outperforms competing approaches in terms of accuracy and recall rate, successfully addressing the issue of data sparsity and enhancing the quality of recommendations. Rich contextual information, such as time, remark language, and photographs, are factors that influence the user's check-in behaviour in real life in addition to the social relationship and geographic aspects taken into account in this article. We'll concentrate on how to incorporate more contextual data in next studies to enhance the effectiveness of POI recommendations.

References

- [1] S. Kim & R. E. Banch. (2014). R-cube: A dialogue agent for restaurant recommendations and reservations. *IEEE*, 1-6.
- [2] C. B. Marques & C. Santos. (2015). Tourist routes strategies of local development. Pasos Revista De Turismo Y Patrimonio Cultural, 12(3), 539-548.
- [3] A. Lapidoth. (2017). A Foundation in digital communication. The Univariate Gaussian Distribution.
- [4] A. Helu, H. Samawi, & H. D. Rochani. (2020). Kernel density estimation based on progressive type-II censoring. *Journal of the Korean Statistical Society*, 1-24.
- [5] R. Mathonat, D. Nurbakova, & J. F. Boulicaut. (2020). Anytime mining of sequential discriminative patterns in labeled sequences. *Knowledge and Information Systems*.
- [6] S. Yali, Z. Z. Fu, & Y. L. Wen. (2019). An adaptive point-of-interest recommendation method for locationbased social networks based on user activity and spatial features. *Knowledge-Based Systems*, 16(1), 267-282.
- [7] H. Gulam, F. Zeng, W. Li, & Y. Xiao. (2019). Deep learning-based sentiment analysis for roman Urdu text. Procedia Computer Science, 14, 131-135.
- [8] M. J. Strickland, C. Siffel, & B. R. Gardner. (2007). Quantifying geocode location error using GIS methods. *Environmental Health*, 6(10).
- [9] M. Ye, P. Yin, & W. Lee. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. *Proceedings of the 34th International ACMSIGIR Conference on Research and Development in Information Retrieval*, Beijing, China, 325-334.

- [10] M. A. Amin & D. Grin. (2014). Probing early-universe phase transitions with CMB spectral distortions. *Physical Review*, 90(8), 083529.
- [11] A. Hossein, A. Mohammad, & A. Sajad. (2019). LGLMF: Local geographical-based logistic matrix factorization model for POI recommendation. *The 15th Asia Information Retrieval Societies Conference Hong Kong, China*.
- [12] D. Bokde, S. Giras, & D. Mukhopadhyay. (2015). Matrix factorization model in collaborative filtering algorithms a survey. *Proceedings of the 4th International Conference on Advances in Computing Communication and Control, Mumbai, India*, 136-146.
- [13] Y. Hu, Y. Koren, & C. Volinsky. (2008). Collaborative filtering for implicit feedback datasets. Proceedings of the 8th IEEE International Conference on Data Mining, Pisa, Italy, 263-272.
- [14] P. Zhao, X. Xu, & Y. Liu. (2017). Exploiting hierarchical structures for POI recommendation. *IEEE International Conference on Data Mining*, 655-664.
- [15] Y. Liu, T. A. N. Pham, & G. Cong. An experimental evaluation of point-of-interest recommendation in location-based social networks. *Proceedings of the VLDB Endowment*, 10(10), 1010-1021.

This page is empty by intention.