Schedule a Rich Sentimental Travel via Sentimental POI Mining and Recommendation

Peiliang Lou[†‡], Guoshuai Zhao[†‡], Xueming Qian[†‡*], Huan Wang[†], Xinsong Hou[†]
†{School of Electronics and Information Engineering}, ‡{SMILES LAB}, Xi'an Jiaotong University
Xi'an, China
peilianglou@sina.com; zgs2012@stu.xjtu.edu.cn; qianxm@mail.xjtu.edu.cn; wang.huan.1006@stu.xjtu.edu.cn; houxs@mail.xjtu.edu.cn

***Abstract*—Trip is an important part of people's lives. We need to obtain interesting location information quickly and conveniently from the huge amount of information. The existing data source of geographic locations can be divided into two categories. The one is users' information that we can utilize to deeply understand their trajectories, the other one is the characteristics and attributes of geographic locations. In this paper, we focus on sentimental attributes of location and propose a POI (Point-Of-Interest) Mining method. Firstly, we use SPM (Sentiment-based POI Mining) algorithm to mine the POIs (Points-Of-Interest) with obvious sentimental attributes, and then recommend the POIs to users by using SPR (Sentiment-based POI Recommendation) algorithm. We conduct a series of experiments in Sina Weibo dataset. The results show the effectiveness of our method.**

Keywords-recommender systems; sentiment analysis; social network; POI

I. Introduction

The social networks have great improvement these years, but the overload information accumulated by social networking applications makes information consumers and information providers confused. Recommender systems can predict user preference by mining the data of user's social networks.

Location-based social networks (LBSN) [1-3] bring new opportunities for recommender systems, including POIs (Point-Of-Interest) recommendation. The data sources the existing location recommender systems use can be divided into 2 categories: 1) user information, which includes user's profiles, user's location histories, and user's trajectories. 2) features of locations such as the number of users visiting a place, and physical attribute such as restaurants or parks[4]. However, in the process of mining POIs, the sentimental features of locations are seldom considered, thus the recommendations may not suit the users' sentimental preference. For example, a user in downtown wants to go to a peaceful place to have a rest, but the system mainly focuses on the geographic attributes of locations near the user, and it may recommend some POIs full of people and noise which will not meet the user's needs. POIs not only have geographic attributes, but also have sentimental attributes. For example, the POIs with many historic spots will be solemn, and the POIs which have many bars and clubs will be lively, whereas the POIs with a lot of trees and pools will be peaceful and relaxing. The sentimental attributes of POIs can be discovered by analyzing the data accumulated from the social network.

Location-based online social networking applications allow users to share their experience on ecommerce platforms or social platforms, and current locations by checking-in on websites such as Sina Weibo[1], and Facebook etc. Shown as in Fig. 1, *Text* is the user's comment about their status or feelings and *Location* shows the GPS position of the user. In a way, the information of *Text* is strongly connected with the *Location*. For example, *Text* could be the comments of the restaurants within *Location,* or the review of the landscape of *Location*. Through sentiment analysis of this information, the sentimental attributes of the *Location* could be discovered.

Figure 1. User's comments on microblog.

After that, POIs with obvious sentimental attributes are mined and will be recommended if they match users' preferences and near users' geographic locations.

Now many researchers pay close attention to the research of users' behaviors and sentiment via social media big data. But in this paper, the sentimental attributes of POIs are detected through the sentiment analysis of microblogs, and then the POIs with certain sentiment attributes are recommended to different users. Firstly, we solve the

This work is supported in part by the Program 973 No. 2012CB316400, by NSFC No.61373113, 61173109, 61332018, and Microsoft Research Asia.

*Corresponding author

[1]http://weibo.com/

problem of mining the POIs which have obvious sentimental attributes by using the proposed SPM algorithm. The sentimental features of POIs are discovered by performing sentimental analysis of Sina Weibo data. According to the result of sentimental analysis and GPS positions of social media data, the POIs which have thick density of social media data and obvious sentimental attributes are searched and regarded as recommended POIs. And then we incorporate sentiment similarity analysis and geographical distance into our recommender system which is based on the widely adopted latent factor model realized by Probabilistic Matrix Factorization (PMF).

The main contributions of this paper are shown as following:

1) We improve the traditional Meanshift algorithm for POI mining [12] with considering sentimental attributes of locations. Through this algorithm, we discover the POIs with obvious sentimental attributes by performing sentimental analysis of Sina Weibo. Our new algorithm is more comprehensive and considerate than traditional algorithms.

2) We incorporate the analysis of sentiment similarity between POIs and users into our recommender system realized by Probabilistic Matrix Factorization. In our recommender system, we also consider the geographic distance between POIs and users.

3) We conducted numerous experiments to evaluate the effectiveness of incorporating the analysis of sentiment similarity and the geographic distance in the recommender system and compared the proposed model with BaseMF, Bias-MF, CF-MF, SPR-S, SPR-L, and SPR. We further show that considering the sentiment attributes of locations can much improve the performance of the recommender systems with more humanization and intelligentialize.

The rest of this paper is organized as follows. We review the related work on the overview of recommender systems in LBSN, POI recommendation and the sentiment analysis of microblog in Section 2. The introduction of our SPM algorithm is reported in Section 3. The details of our SPR algorithm are presented in Section 4, followed by experimental evaluation in Section 5. Finally, Section 6 concludes this paper.

II. Related Work

In this section, we first introduce the overview of recommender systems in LBSN and the POI recommendation, and then we introduce some methods of sentiment analysis of microblogs.

A. *The POI Recommender System*

The recommendation of POIs [6-8] is to provide recommendations of places of interests based on several factors. Geographical influence is one of the major considerations of POI recommendation [9] which indicates that users tend to visit nearby POIs around their homes or offices. The works [6-8] make recommendations by considering the distance between users and locations, and the closer locations have higher probabilities to be recommended. Hu *et al*. [10] observed that there exists weak positive correlation between a business's ratings and its neighbors' ratings, regardless of the categories of businesses. Based on this observation, the recommender system considers geographical neighborhood influences in business recommendation together with influences from other factors including user reviews, business category, and business popularity. User preference is another important consideration of POI recommendation [10-15, 32, 35]. Check-in data is used to study user preference in [10, 12-15], especially in [15], Jiang *et al*. proposed a Geographical-Temporal influences Aware Graph in order to recommend POIs to a user when he or she wants to visit at a given time. What's more, the textual description of photos [12], and social influence among friends [9, 34] can also reflect users' preferences.

Personalized recommendation of POI has attracted significantly attention recently. Collaborative filtering (CF) is a widely used method. The works [8, 9] recommend POIs according to Geographical influence via collaborative filtering. The prediction is an important step of recommendation based on CF method. The works [7, 8] introduce a model to consider geographical influence for POIs by using MF techniques. Cheng *et al*. [6] fuse MF with geographical and social influence for POI recommendation in LBSNs. Qian *et al*. [32, 34] consider more social factors to constrain user and item latent features by using MF techniques.

B. *The Sentiment Analysis of Text*

The first step of researching sentiment analysis of text is the translation of the text into a format that computer can understand, which is based on Natural Language Processing (NLP) methods. The Bag of Words, the N-Grams and the N-Gram Graphs methods are three of the most widely used NLP methods. According to the Bag of Words [17], the sentences of a text are divided into a set of words, and a dictionary correlates each word with a numerical value, showing its sentiment polarity.

The lexicon-based method, one of the major methods of sentiment analysis, computes the sum of numerical value of each word as the overall sentiment polarity of texts. There are several elements of texts considered through this approach. Benamara *et al*. [18] consider the use of adverb and adjectives to detect microblogs' sentiment polarity. Lei *et al*. [31] propose a social user's reviews sentiment measurement approach and utilize it to compute service reputation.

The sentiment analysis of microblogs has attracted significant attention in these years. The works [19-21] explore the use of emoticons. Zhao *et al*. [21] propose a system called MoodLens to perform the sentiment analysis for Sina Weibo which employs an emoticon-based method for sentiment classification. Except for the microblogs itself, the context where microblogs are published is also incorporated into the methods of sentiment analysis such as the author's profile of the microblogs, and the friends of the author. Hu *et al*. [22] consider the social relationship between users, and You *et al*. [23] take the visual content of microblogs into consideration.

Figure 2. The principle of SPM algorithm. The red, green, and blue dots represent microblogs with different sentimental values and GPS positions.

III. OUR SENTIMENT-BASED POI MINING MODEL

In this paper, we incorporate POIs' geographic attributes and sentimental attributes into our SPM model. There are 2 steps of SPM algorithm. The first step is analyzing the sentimental features of locations based on GPS positions and sentimental attribute values of micro-blogs, and the second step is to mine the POIs which have obvious sentiment feature. The Fig. 2 shows the principle of SPM algorithm.

In the Fig. 2, the red, green, and blue dots represent microblogs with different sentimental values and GPS positions. Firstly, the sentimental features of locations are discovered by calculating the sentimental values of microblogs published in these locations. Secondly, based on sentimental values and GPS positions of these microblogs, the locations which have microblogs of thick density and small undulating sentimental values are mined and regarded as POIs.

A. The Sentiment Analysis of Microblogs

The first step of SPM algorithm is performing the sentiment analysis of microblogs. The microblogs published in a certain area, such as users' experience or their reviews about the landscape, restaurants, will not only carry the information about users' preferences and behaviors but also their sentiments like happiness, and sadness. This vast information can represent the sentimental attributes of this location. For example, if happiness is more frequent than other sentiment about the data in one location, it means that this location has a sentimental attribute of making people happy. Through analyzing the sentiment of the microblogs published here, the specific sentimental attributes of this location can be discovered. Thus, we study how to calculate the sentimental values of microblogs in this section.

In this paper, we analyze the sentiment of microblogs based on the lexicon-based method. HowNet Sentiment Dictionary[2] was used to compute sentiment values as in [31], which is well known in the area of Chinese and English sentiment classification.

[2]http://www.keenage.com/download/sentiment.rar

For each microblogs, we divide it into clauses by punctuation marks. Then for each clause, firstly, we find out the positive and negative words. A positive word is initially assigned with value +1.0, whereas a negative word is assigned with value -1.0. Secondly, we take the degree words into consideration as a factor. At last, we add a negation check coefficient that has a default value of 1.0. If the word item is preceded by a negation within the specified location, this coefficient is set to -1.0. Then the review sentiment value could be calculated by:

$$V = \frac{\sum_{c \in m} \sum_{w \in c} I_c D_w S_w}{n_{clause}} \tag{1}$$

where V denotes a microblog's sentiment value. m denotes a microblog. c denotes a clause. w denotes a word. I_c denotes the negation check coefficient. D_w denotes the factor of w's degree word. S_w denotes the sentiment values of word w.

Based on the above-mentioned algorithms, the sentimental values of microblogs can be obtained.

B. Clustering POIs with Combining Sentiment Feature

The second step of SPM algorithm is to discover the POIs which have obvious sentiment feature. If microblogs are dense in a location such as shopping center, school and scenic spot, it shows this location has high population density and has research significance for sentimental attribute. Traditional Meanshift algorithm for POI mining [12, 33] is proved by this method with considering sentimental attributes of locations.

Firstly, GPS positions containing many microblogs are selected as centers. The location of 500 meters away from the center is classified as the initial class.

Secondly, based on the initial class, we adjust GPS positions of the center points and determine the final locations as POIs.

We adopt GPS positions and sentimental values of microblogs as the elements of microblogs' coordinates, which are regarded as their labels. i denotes the i-th micro-blog in this region, that is:

$$x_i = [lon_i,\ lat_i,\ v_i] \tag{2}$$

where lon_i, lat_i respectively denotes the GPS position of the i-th microblog. v_i denotes the sentimental values of the i-th microblog. We utilize M_h to represent the degree of dispersion of the spatial position and sentimental value. M_h could be calculated by:

$$M_h = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x_m) \tag{3}$$

where S_h denotes the whole area of the location of the initial class. In our method, the longest distance between the GPS position of comment and the GPS position of center is no more than 500 meters. x_m denotes coordinates of the center point. k denotes the total amount of microblogs.

We need to decide whether to change the center by calculating the size of vector M_h. The condition of judgment is:

$$\sqrt{D^2 + \frac{\Delta V^2 \lambda}{D^2}} \leq \alpha \tag{4}$$

where D denotes the length of the vector M_h in the space. Physically, D is the geographic distance between two GPS positions which is calculated by Haversine geodesic distance equation [30]. ΔV denotes the sentimental value of vector M_h .We regard $\sqrt{D^2+\frac{\Delta V^2\lambda}{D^2}}$ as magnitude of M_h . Because GPS distance and sentimental value are not in the same reference system, and have different range of value, λ is used to adjust the weight of ΔV in order to balance the importance of GPS distance and sentimental value. At the same time, GPS distance and sentimental value have different volatility, thus $\Delta V^2\lambda$ is divided by D^2 in order to balance the volatility. α is used to determine whether the magnitude of M_h is suitable for clustering. α limits the size of M_h in order to ensure that the distribution of sentimental values of microblogs is concentrated and the geographic distribution of microblogs is dense.

When M_h meets this requirement, the center is determined. And the location within the range of 500 meters to the center is classified into the same cluster as a POI. But on the contrary, M_h coordinate is added to the recent center's coordinate, and the sum is regarded as the coordinate of a new center for seeking the M_h meeting the requirement.

IV. Our Sentiment-based POI Recommendation Model

A. Matrix Factorization

Our proposed model is based on the latent factor model realized by matrix factorization [11, 27, 32]. Using user and POI latent feature vectors $\boldsymbol{U}_u$ and $\boldsymbol{P}_i$, the basic location recommendation model approximates user's preference via solving the following optimization problem:

$$\min_{\boldsymbol{U},\boldsymbol{P},\boldsymbol{b}} \frac{1}{2}\sum_u \sum_{i\in|H_u|} \left(R_{u,i}-\hat{R}_{u,i}\right)^2+\frac{\lambda_1}{2}\|\boldsymbol{U}\|_F^2+\frac{\lambda_2}{2}\|\boldsymbol{P}\|_F^2 +\lambda_3\left({b_u}^2+{b_i}^2\right) \tag{5}$$

where $\hat{R}_{u,i}$ denotes the number of check-ins predicted by:

$$\hat{R}_{u,i}=r+U_u^{\mathrm{T}}P_i+b_u+b_i \tag{6}$$

where r is an offset value, which is empirically set as users' average check-in value. $R_{u,i}$ is the real check-in values of user u to POI i. $\boldsymbol{U}$ and $\boldsymbol{P}$ are the user and POI latent feature matrices which need to be learned, $\|\boldsymbol{Y}\|_F$ is the Frobenius norm of matrix $\boldsymbol{Y}$, and $\|\boldsymbol{Y}\|_F=\left(\sum_{i,j} y_{i,j}^2\right)^{1/2}$. It is used to avoid over-fitting [32]. b_u and b_i are the user and POI bias. This objective function can be minimized efficiently by using gradient descent method as [32]. Once the low-rank matrices $\boldsymbol{U}$ and $\boldsymbol{P}$ and $\boldsymbol{b}$ are learned, the number of check-ins can be predicted according to (6) for any user-POI pairs.

B. The Factor of Sentiment Similarity

Sentimental attribute is one of the main concerns of this paper. It plays a significant role to affect users' decision on where they would like to go. With the concept of item based collaborative filtering, the view of sentiment similarity can be inferred that the POIs which bring similar sentiment to users are relevant. Therefore, sentiment similarity can be utilized to optimize POI latent feature vector $\boldsymbol{P}_i$.

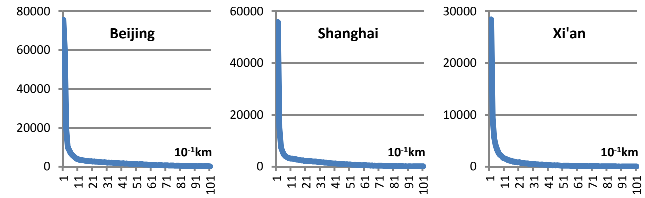

Figure 3. The relevance of users' check-in number and the geographical distances between users and POIs.

The sentiment value of the mined POI is represented by the mean sentimental value of involved micro-blogs, which are ranging from 0 to 1.Exponential function is utilized to compute the un-linear sentiment similarity as follows:

$$E_{i,j}=Exp(-|Avg(i)-Avg(j)|) \tag{7}$$

where $Avg(\cdot)$ denotes the mean sentimental value of involved microblogs in a POI. This formula denotes that the similarity between POI i and POI j becomes higher if the sentimental values are similar.

Considering with sentiment similarity between POIs, the optimization problem can be expressed by:

$$\min_{\boldsymbol{U},\boldsymbol{P},\boldsymbol{b}} \frac{1}{2}\sum_u \sum_{i\in|H_u|} \left(R_{u,i}-\hat{R}_{u,i}\right)^2+\frac{\lambda_1}{2}\|\boldsymbol{U}\|_F^2+\frac{\lambda_2}{2}\|\boldsymbol{P}\|_F^2 +\lambda_3\left({b_u}^2+{b_i}^2\right) +\frac{\beta}{2}\sum_i\left(\left(P_i-\sum_j E_{i,j}^* P_j\right)^T\left(P_i-\sum_j E_{i,j}^* P_j\right)\right) \tag{8}$$

where $E_{i,j}^*$ is the normalization value based on the number of POIs, resulting $\sum_j E_{i,j}^*=1$ .The new added term is utilized to optimize the latent vectors of POIs via exploiting sentiment similarity.

C. The Factor of Geographical Distance

Geographical attribute is another one of the main concerns of this paper. It also plays a significant role to affect users' decision on where they would like to go. For example, when we search a restaurant for our normal dinner considering convenience, we will never choose a faraway eatery. The researchers[3] find that the activity radiuses of 45% users are no more than 10 miles. The relevance of users' check-in number and the geographical distances between users and POIs are shown in Fig. 3. From the global observation, it can be seen that users would like to visit the closer POIs to their activity centers. For a user, the average geographical location of POIs checked by this user is set as his/her activity center. In other words, for a user u, we represent his/her activity center position as $\left(\frac{\sum_{i\in H_u} lat_i}{|H_u|},\frac{\sum_{i\in H_u} lon_i}{|H_u|}\right)$, where i denotes the POI. H_u denotes the set of POIs checked by user u. $|H_u|$ denotes the number of check-ins. lat_i and lon_i are the latitude and longitude of POI i.

Geographical distance between user and POI can be utilized to optimize user and POI latent feature vectors $\boldsymbol{U}_u$ and $\boldsymbol{P}_i$ .The basic idea is that the number of check-ins becomes larger with the distance decrease.

[3] http://www.nytimes.com/interactive/2010/01/10/nyregion/20100110-netflix-map.html

Exponential function is utilized to compute the un-linear factor of geographical distance as:

$$L_{u,i} = Exp[-Distance(u,i)] \quad (9)$$

where $Distance(u,i)$ denotes the geographical distance between user u and POI i via their latitude/longitude coordinates. Note that, the geographical distance between two latitude/longitude coordinates is calculated by using the Haversine geodesic distance equation [30].

Incorporating the factor of geographical distance between user and POI, the optimization problem is updated and can be expressed by:

$$\min_{\boldsymbol{U},\boldsymbol{P},\boldsymbol{b}} \frac{1}{2}\sum_u \sum_{i\in|H_u|} \left(R_{u,i} - \hat{R}_{u,i}\right)^2 + \frac{\lambda_1}{2}\|\boldsymbol{U}\|_F^2 + \frac{\lambda_2}{2}\|\boldsymbol{P}\|_F^2 + \lambda_3\left({b_u}^2 + {b_i}^2\right) + \frac{\beta}{2}\sum_i \left(\left(P_i - \sum_j E_{i,j}^* P_j\right)^T \left(P_i - \sum_j E_{i,j}^* P_j\right)\right) + \frac{\gamma}{2}\sum_u \sum_{i\in|H_u|} \left(L_{u,i}^* - U_u^{\mathrm{T}} P_i\right)^2 \quad (10)$$

where $L_{u,i}^*$ is the normalization value based on the number of POIs, resulting $\sum_i L_{u,i}^* = 1$. The new added term is utilized to optimize the latent vectors via exploiting the factor of geographical distance between user and POI.

V. EXPERIMENTS

Adopting the data of Sina Weibo as the dataset, the proposed methods in this paper are verified by experiments. We compare the performance of our SPM method with that of Meanshift method [5], K-means, and DBSCAN [24]. These are all traditional popular clustering algorithms. Then our SPR method is compared with BaseMF [11], Bias-MF [28], CF-MF [27]. In addition, we present POIs mined by using our model on Bing map.

A. Dataset

The data used for the experimental verification is crawled from Sina Weibo, i.e., the Twitter of China. The number of the microblogs of Sina Weibo is 1.2 million from 388,780 users. Every micro-blog can be treated as a check-in. Table 1 shows the statistics of Sina Weibo dataset. Three cities are involved in our dataset, including Beijing, Shanghai, and Xi'an. Weibo API is utilized to crawl these data. Firstly, we set a GPS position as a center of a city. Generally, the position we set is the administration center of this city. Therefore, it is not guaranteed that the center is the most prosperous area in the city. Secondly, the authorized API is utilized to crawl microblog data involved in the covered area within radius 11.132*km*. At last, the crawled microblogs labeled with GPS positions are mapped. Fig. 4 shows the geographical distributions of Sina Weibo data in Beijing. Note that, the higher the column is, the more check-ins there are.

TABLE I. STATISTICS OF SINA WEIBO DATASET

City	Beijing	Shanghai	Xi'an
Number of Users	192,143	134,154	62,483
Number of Micro-blogs	608,723	408,980	187,838
Number of GPS Positions	242,262	231,769	108,340
Center	39°54'27"*N* 116°23'17"*E*	31°13'46.86"*N* 121°28'24.71"*E*	34°20'29.64"*N* 108°56'24.63"*E*
Coverage Radius	11.132*km*		

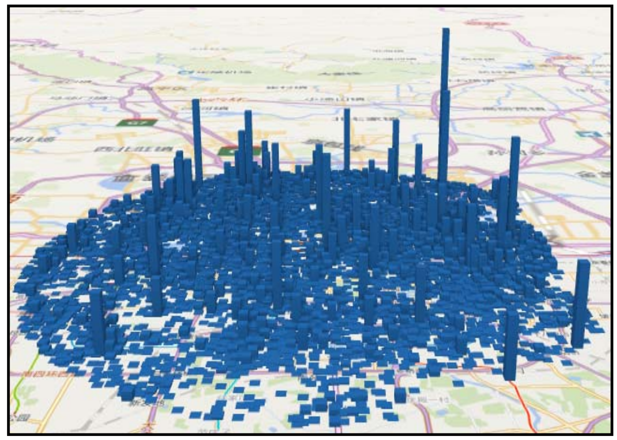

Figure 4. The geographical distributions of Sina Weibo data in Beijing. The higher the column is, the more check-ins there are.

B. Performance Measurements

1) POI Mining

When POIs are mined, the sentimental values of their interior microblogs have wide variances and spatial inhomogeneous distribution. Two kinds of measurement are utilized for performance evaluation of mining results.

The first measurement shown in Table 2 is represented by $Measure_1$ as following:

$$Measure_1 = \frac{1}{M}\sum_{j=1}^{M} \sqrt{\sum_{i\in POI(j)}^{N_j} \left[\varepsilon\left(D_{ij}\right)^2 + \theta\left(V_i - \bar{V}_j\right)\right]/N} \quad (11)$$

where j denotes the j-th POI, i denotes the i-th micro-blog in the j-th POI. $\bar{V}_j$ denotes the average sentimental value of the microblogs in j-th POI. D_{ij} denotes the geographic distance between the GPS position of the i-th micro-blog and the geographic center of the j-th POI.

$Measure_1$ is used to evaluate the distribution of sentimental values and geographic positions of microblogs in POIs, of which the symbols are M-I and M-II and M-III. The lower M-I, M-II and M-III are, the better the performance of the model.

The second measurement is represented by $Measure_2$ as following:

$$Measure_2 = \sqrt{\sum_{j=1}^{M}\left(\bar{V} - \bar{V}_j\right)^2/M} \quad (12)$$

The symbol of $Measure_2$ is M-IV. $\bar{V}_j$ denotes the average sentimental value of the microblogs in j-th POI. $\bar{V}$ denotes the mean of the sentimental values among all POIs. The bigger M-IV is, the larger the gap of sentimental attribute among different POIs. Thus adopting this method

TABLE II. PERFORMANCE MEASUREMENTS ON POI MINING

Measures	Parameter setting	Description	Symbol	Note
$Measure_1 = \frac{1}{M}\sum_{j=1}^{M}\sqrt{\sum_{i\in POI(j)}^{N_j}\left[\varepsilon\left(D_{ij}\right)^2 + \theta\left(V_i - \bar{V}_j\right)\right]/N}$	$\varepsilon = 0,\ \theta = 1$	The average error of each POI on sentimental values of the POI's inner social media data.	M-I	The lower, the better.
	$\varepsilon = 1,\ \theta = 0$	The average fluctuation of each POI on geographic distance from the locations of POI's inner social media data to POI's center.	M-II	
	$\varepsilon = 1,\ \theta = 1$	The average comprehensive fluctuation of each POI on sentimental values and geographic distance.	M-III	
$Measure_2 = \sqrt{\sum_{j=1}^{M}\left(\bar{V} - \bar{V}_j\right)^2/M}$	—	The RMSE of sentimental values of every POI.	M-IV	The higher, the better.

TABLE III. MAE OF ALL METHODS IN OUR WEIBO DATASET, THE LOWER THE BETTER

	BaseMF	Bias-MF	CF-MF	SPR-S	SPR-L	SPR
Beijing	0.105	0.107	0.101	0.100	0.101	0.098
Shanghai	0.107	0.109	0.097	0.097	0.095	0.087
Xi'an	0.241	0.241	0.211	0.207	0.209	0.182
Mean	0.151	0.152	0.136	0.135	0.135	0.122

can find the POIs which have different sentimental attribute and leads to a better mining result.

2) POI Recommendation

The evaluation metric we use in our experiments is Mean Absolute Error (MAE). It is the most popular accuracy measures in the literature of recommender systems [29]. MAE is defined as:

$$MAE = \sum_{(u,i)\in\Re_{test}}\left|R_{u,i} - \hat{R}_{u,i}\right|/|\Re_{test}| \quad (13)$$

where $R_{u,i}$ is the real number of check-ins user u to POIi, $\hat{R}_{u,i}$ is the corresponding predicted value. $\Re_{test}$ is the set of all user-POI pairs in the test set. $|\Re_{test}|$ denotes the number of user-POI pairs in the test set.

Note that, logistic function $\frac{1}{1+(e^t)^{-1}}$ is commonly used in recommender system [25] to map each matrix element into [0, 1]. But for our dataset, some check-in frequencies of a location are large while the function $(e^t)^{-1}$ would result in very small and indistinguishable values [26]. Therefore, as [26], the logistic function is defined as $\frac{1}{1+t^{-1}}$.

C. Evaluation

1) Performance of POI Mining

We compared the performance of our method and the traditional popular methods, including Meanshift [5], K-means, and DBSCAN [24]. These methods only consider the GPS position of data point rather than its sentimental attribute.

As shown in Fig. 5, the M-I and M-II, M-III, and M-IV computed by using the Sina Weibo dataset. The smaller M-I and M-II and M-III, and the larger M-IV, the better the method. Comparing with these methods, M-I and M-II and

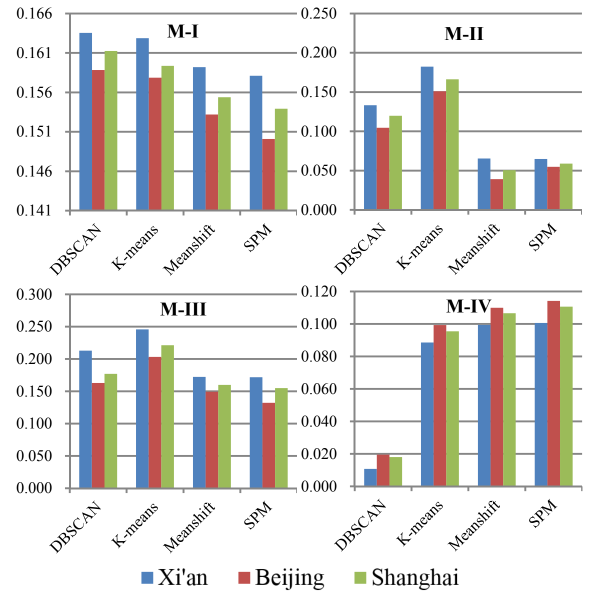

Figure 5. The Performance comparison of POI mining on our Sina Weibo dataset. In M-I, M-II, and M-III, the lower the better. In M-IV, the higher the better.

M-III have different degrees of error decrease of our method, which means that using our method, the geographic distribution of each POI's inner social media data is denser and the distribution of each POI's inner social media data sentimental values is more concentrated. M-IV of our method is larger than that of the compared methods, which means that the gap of sentimental attribute among different POIs has increased by using our method. Therefore, our method is more effective.

2) Performance of POI Recommendation

In this section, we compare the performance of our SPR model with the existing methods including BaseMF [11], Bias-MF [28], CF-MF [27], SPR-S, and SPR-L. The compared methods are overviewed as below:

- Basic Probabilistic Matrix Factorization Model (BaseMF): Matrix factorization is utilized to learn the latent features of users and items for rating prediction. The basic probabilistic matrix factorization (BaseMF) approach [11] learns the latent features by minimizing the objective function

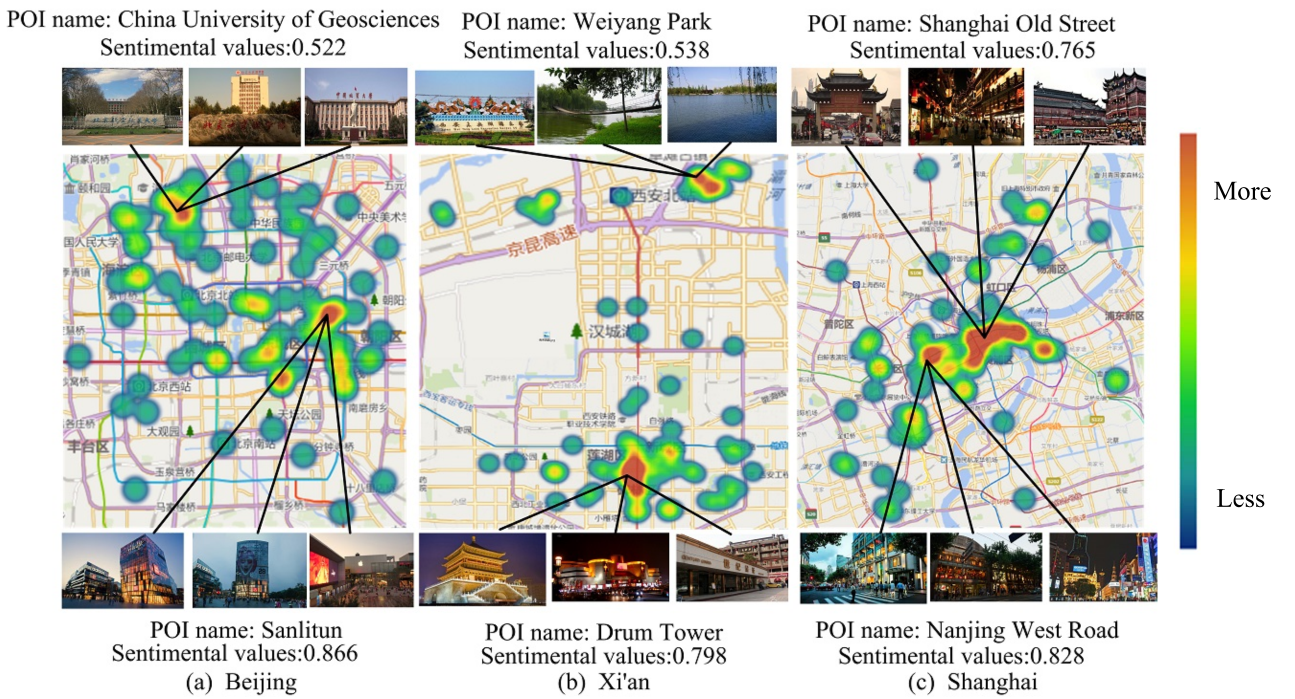

Figure 6. The presentation of the POIs mined by SPM method in Beijing, Xi'an, and Shanghai. The color temperature represents the number of check-ins. The warm color represents more check-ins. The cool color represents less check-ins.

based on the observed rating data. We adopt it for POI recommendation.

- Biases based Matrix Factorization Model (Bias-MF): Biases are utilized in matrix factorization models [28]. Biases based matrix factorization model can be seen as the improved BaseMF.
- Collaborative Filtering based Matrix Factorization Model (CF-MF): Collaborative filtering method is used to compute the similarity between items. We adopt it for latent factor model to recommend POIs as [27].
- Sentiment based Matrix Factorization Model (SPR-S): This is the compact version of our SPR model. It just takes the factor of sentiment similarity into consideration.
- Location based Matrix Factorization Model (SPR-L): This is the compact version of our SPR model. It just takes the factor of geographical distance into consideration.
- The proposed Sentimental based POI Recommendation Model (SPR): This is our SPR model. It takes two factors, including sentiment similarity and geographical distance, into consideration.

There are two factors we take into consideration. However, the impact of the two factors to performance is invisible. Therefore, the methods of SPR-S and SPR-L are performed in this section to demonstrate the effectiveness of our proposed factors. In Table 3, SPR-S denotes the method that considers the factor of sentiment similarity. SPR-L denotes the method that considers the factor of geographical distance. The baselines are the methods of BaseMF, Bias-MF, and CF-MF. It can be seen that, SPR model decreases the prediction error by 19%, 20%, 10% on MAE respectively. Furthermore, it can be seen that both of the factors are effectiveness in our model. The union of two factors is better than the single one.

D. Presentation

The POIs mined by our method are shown in Fig. 6. The color temperature represents the number of check-ins. The warm color represents more check-ins. The cool color represents less check-ins. The sentimental values are normalized. The values below 0.5 show that the sentiment of the POI is negative or the values above 0.5 show that of POI is positive and bigger values mean that the POI is much more positive. As shown in Fig. 6, our method mines not only the POIs with high sentimental values such as Sanlitun in Beijing and Nanjing West Road in Shanghai whose sentiment attributes are happiness and joy, but also the POIs like Weiyang Park and China University of Geosciences whose sentimental attributes are peace and solemn.

VI. CONCLUSIONS

Existing location recommendation systems mainly focus on physical characteristics of locations rather than the sentimental attributes of location. SPM method for POI Mining and SPR method for POI Recommendation based on social media data are proposed in the paper. According to our methods, the POIs possessing obvious sentimental attributes are mined and recommended to users. The methods proposed in this paper are tested by Sina Weibo dataset. The results show that our methods have high effectiveness. The methods are beneficial to discover sentimental attributes of different locations, and they can be used in recommender systems to recommend POIs to users which can suit the users' preference.

REFERENCES

[1] L. Humphreys, "Mobile social networks and social practice: A case study of dodgeball," Journal of Computer-Mediated Communication, Vol. 13, No. 1, 2007, pp. 341–360.

[2] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," Proceedings of Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[3] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks, " Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp.329-336.

[4] J. Bao, Y. Zheng, D. Wilkie, and M. F. Mokbel, "Recommendations in Location-based Social Networks: A Survey," Geoinformatica, Volume 19, Issue 3, 2015, pp 525-565.

[5] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, Issue8, 1995, pp. 790-799.

[6] C. Cheng, H. Yang, I. King, and M. R. Lyu , "Fused matrix factorization with geographical and social influence in location-based social networks," Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012, pp. 17-23.

[7] T. Kurashima, T. Iwata, T. Hoshide, N. Takaya, and K. Fujimura, "Geo topic model: joint modeling of user's activity area and interests for location recommendation," WSDM'13 Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 375–384.

[8] M. Ye, P. Yin, and W.C. Lee, "Location recommendation for location-based social networks. GIS," GIS'10 Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010, pp. 458–461.

[9] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," ACM SIGIR, 2011, pp. 325–334.

[10] H. Gao, J. Tang, X. Hu, and H. Liu, "Content-aware point of interest recommendation on location-based social networks," Proc. 29th Int. Conf. AAAI, Austin, TX, USA, Jan. 2015, pp. 1–7.

[11] R. Salakhutdinov, and M. Andriy, "Probabilistic matrix factorization," Neural Information Processing Systems , NIPS, 2007, pp. 1257–1264.

[12] S. Jiang, X Qian, J Shen, Y Fu, and T Mei, "Author Topic Model-Based Collaborative Filtering for Personalized POI Recommendations," IEEE Trans. Multimedia, vol. 17, no. 6, 2015, pp. 907-918.

[13] Y. Lyu, C.-Y. Chow, R. Wang, and V. C. Lee, "Using multi-criteria decision making for personalized point-of-interest recommendations," SIGSPATIAL'14 Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2014, pp. 461–464.

[14] J. Sang, T. Mei, J.-T. Sun, C. Xu, and S. Li, "Probabilistic sequential POIs recommendation via check-in data," SIGSPATIAL'12, 2012, pp. 402–405.

[15] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," CIKM'14 Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, 2014, pp. 659–668.

[16] B. Hu and M. Ester, "Spatial Topic Modeling in Online Social Mediafor Location Recommendation," RecSys'13 Proceedings of the 7th ACM conference on Recommender systems, ACM, 2013, pp. 25-32.

[17] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs," Proceedings of the International Conference on Weblogsand Social Media (ICWSM), 2007.

[18] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V.S. Subrahmanian, "Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone," Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 2007.

[19] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," SAC'13 Proceedings of the 28th Annual ACM Symposium on Applied Computing, 2013, pp. 703–710.

[20] K. Liu, W. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis," Proceedings of the National Conference on Artificial Intelligence, 2012, pp. 1678–1684.

[21] J. Zhao, L. Dong, J. Wu, and K. Xu, "Moodlens: an emoticon-based sentiment analysis system for chinese tweets," KDD'12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pages 1528–1531.

[22] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," WSDM'13 Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 537–546.

[23] Q. You and J. Luo, "Towards social imagematics: Sentiment analysis in social multimedia," MDMKDD'13 Proceedings of the Thirteenth International Workshop on Multimedia Data Mining, 2013.

[24] M. Ester, H. Kriegel, J. Sander, and X. Xu, "Adensity-based algorithm for discovering clusters inlarge spatial databases with noise," KDD, 1996, pp.226-231.

[25] H. Ma, H. Yang, M. Lyu, and I. King, "social recommendation using probabilistic matrix factorization," CIKM'08 Proceedings of the 17th ACM conference on Information and knowledge management, 2008, pp. 931-940.

[26] H. Gao, J. Tang, X. Hu, and H. Liu, "Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks," RecSys'13 Proceedings of the 7th ACM conference on Recommender systems, 2013, pp. 93-100.

[27] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," KDD'08, 2008.

[28] Y. Koren, "Collaborative Filtering with Temporal Dynamics," Communications of the ACM CACM Homepage archive Volume 53 Issue 4, April 2010, pp: 89-97.

[29] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in KDD'12, 2012, pp. 1267-1275.

[30] M. Strickland, C. Siffel, B. Gardner, A. Berzen, and A. Correa, "Quantifying Geocode location error using GIS methods," Environmental Health 6:1(2007):1-8.

[31] X. Lei, and X. Qian, "Rating Prediction via Exploring Service Reputation," in Proc. MMSP'15, 2015.

[32] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized Recommendation Combining User Interest and Social Circle," IEEE Trans. Knowledge and Data Engineering, vol. 26, no. 7, pp. 1763–1777, 2014.

[33] Y. Ren, X. Qian, and S. Jiang, "Visual Summarization for Place-of-Interest by Social-Contextual Constrained Geo-clustering," in Proc. MMSP'15, 2015.

[34] G. Zhao, X. Qian, and X. Xie, "User-Service Rating Prediction by Exploring Social Users' Rating Behaviors," IEEE Trans. Multimedia, 2016, vol.xx, no.xx, pp.xx-xx. Accepted

[35] S. Jiang, X. Qian, T. Mei, and Y. Fu, "Personalized Travel Sequence Recommendation on Multi-Source Big Social Media," IEEE Trans. Big Data, 2016, vol.xx, no.xx, pp.xx-xx. Accepted