2018 IEEE 3rd International Conference on Big Data Analysis

Research and Implementation of POI Recommendation System Integrating Temporal Feature

Jinxin Liu[1, 2], Xu Jiao[1, 2], Youzhi Jin[1, 2], Xinyu Liu[1, 2] and Li Liu[*, 1, 2]
[1]Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology
[2]Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology
Tianjin, China
e-mail: tjut_ljx@yeah.net, jiaoxu1999@sina.com, jinyouzhi.lyg@gmail.com, liberty2021@163.com, niceliuli@sina.com[*]

***Abstract*—Point-of-Interest is an interesting research topic of personalized recommendation. In this paper, a new position-based recommendation framework is proposing by using the time attribute. We establish a user matrix with user check-in data sets that contain time and location information. The user matrix is divided into the sub-matrices according to the set period and then every sub-matrix is decomposed into the user checking preference matrix and position feature matrix by using the non-negative matrix factorization. The user's attendance preferences at each time are integrated to get the user's final check-in preference for the candidate location to make recommendations more consistent with the needs of users.**

Keywords- POI; LBSN; temporal feature; recommendation

I. INTRODUCTION

Nowadays, Location-based Social Networks (LBSNs) has received much attention in Point-of-Interest (POI) Recommendation System. LBSNs has been used to mine user's explicit and implicit information of preference in order to recommend personalized services to the target user.

The research of POI recommendation usually involves user behavior analysis, movement pattern model, and trajectory sequence prediction. One of the usages of these study is to improve the quality of recommendations.

Mass information push brings convenience to people, but it also leads to the serious problems of information overload for many of these pushes are not required by the user. The demand for personalized recommendation services is growing. At the same period, the rapid development of LBSNs not only enrich people's life experience but also provide multidimensional information data for recommendation services, including social information level, location information layer, time information layer and content information layer data and so on[1].

II. RELATED WORK

At present, there are many popular foreign LBSNs-based applications, such as Foursquare, Facebook, Twitter, Gowalla, and Brightkite. There are also many great location-based online applications, for example, Tencent QQ Zone, WeChat Moments and DIAN PING, developed by Chinese Internet companies. These software or applications can record user activity or event location information and the user pushes this information to friends by sharing.

The research on social information and geographic information is in-depth. M. J. Pazzani[2] proposed the Demographic-based Recommendation. They assume that a user may like objects that are liked by their similar users. Calculate the pre-k users that are most similar to them, and then recommend them with the purchase and scoring records of these users. A. Gunawardana and G. Shani[3] proposed the Content-based Recommendation. Its basic assumption is that a user may like objects similar to what he once liked, and then use the user's historical record to calculate the user's profile. The most in-depth research is the Collaborative Filtering-Based Recommendation, which can be divided into many subclasses, one of which is the Social-Based Recommendation established by H. Kautz, B. Selman and M. Shah[4]. The Point-of-Interest Recommendation is considered to be an important task in the field of Social-Based Recommendation, also known as position recommendation[5]. For large-scale user check-in records on the LBSNs, many research attempts to use and integrate the geographical impact, social impact, content information and popularity impact to improve the effectiveness of the recommendations of interest points. The properties of location-based social networks have been widely studied in geography and society [6]. C. Cheng, H. Yang, I. King, and M. R. Lyu[7] studied the property of the human check-in model and found that the social status was restricted by geographical factors. H. Gao, J. Tang and H. Liu[8] proposed to utilize the social network information for solving the "cold start" location prediction problem, with a geo-social correlation model to capture social correlations on LBSNs considering social networks and geographical distance. M. Ye, P. Yin, WC. Lee and DL. Lee[9] researched on the geographical impact between locations and proposed a system that combines user preferences, social impacts, and geographic impacts. C. Cheng, H. Yang, I. King, and M. R. Lyu[7] studied the geographical influence by combining a multi-center Gaussian distribution model, matrix decomposition, and social influence.

However, the time information is less considered in the Location-based Social Network. The time information in the location recommendation reflects the change of the user's check-in behavior with time. Through the mining of temporal information, we can provide more accurate location recommendation for users at a given time.

III. Analysis and Preprocessing on Data Set

The data set was collected from Gowalla, and we preprocessed the data. Gowalla data stored in the text file format. Each basic data item contains the user ID, the attendance time, the attendance latitude and longitude and the attendance place ID[10].

A. Data Preprocessing Overview

Data preprocessing is mainly to solve some shortcoming of LBSNs check-in data, such as poor quality, redundancy and so on. Therefore, the main problems in the preprocessing include the fact that the frequency of checkpoints is too low to attract attention or that there is a noticeable spatial error in the data; the information of the check-in data is incomplete and cannot clearly refer to the specific feature; as well as massive duplication of check-in data and data redundancy resulting from heterogeneous ontologies. The original LBSNs check-in data we obtained contains attribute data and spatial location data. On this basis, the mark attribute is added as a criterion for dealing with heterogeneous.

The preprocessing of LBSNs check-in data can be divided into four basic parts and an overall operation flow is shown in Fig. 1. The first step is to make a preliminary selection of all acquired check-in data by setting the lowest number of users' check-in records as a threshold and removing data that has no meaning or low in check-in user number. The second step is to determine whether the information is missing or not, and take the corresponding measure to improve the standardization of LBS check-in data. The third step is to merge the signed data and reduce the redundancy of the check-in data. The fourth step is to analysis the distribution of the user's Check-in condition.

1) Dealing with unpopular locations and filter invalid data: A category of geographic sites that are concerned with large numbers of people, known as the Geographic Hot Point(GHP)[12]. Another category of geography that corresponds to a relatively small number of people is defined as Geographic Cold Point (GCP)[12]. The GCP is not added to the map because the number of visitors is too small to be the POI of public interest. Therefore, it must be eliminated.

The main performance of this kind of point is:

- the latitude and longitude exceed the range of the target area;
- locations only meaningful to the few groups, such as "bed" and "kitchen";
- apparent spelling error;
- geographical meaning clear, but almost no attention.
- This type of data is characterized by a small number of users signing in.

2) Eliminate redundant data: There is a kind of heterogeneous data in data processing, that is, there are sometimes different ways of calling the same geographic entity, including the standard name of the geographic entity, commonly known etc. If you do not handle it, it will cause multiple POI data in the database to point to the same geographic entity.

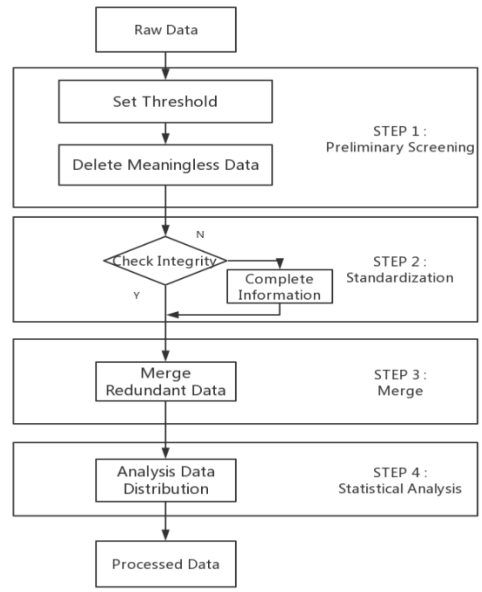

Figure 1. Pretreatment steps

B. Statistical Analysis

Statistical analysis of the data sets can provide a more intuitive understanding of the overall distribution of the data and the basic situation. First, the data is imported into the database to get the distribution of each check - in. Next, people with the same interest are found according to the location information of the user and attributes of the location.

IV. Establishment of Recommendation Framework

The main goal of this project is to study the POI recommendation. The so-called "interest point" refers to the location of the user check-in, such as restaurants, shopping centers, tourist attractions, etc. And the interest point recommendation is to recommend a place that the user may visit in the feature. POI recommendation service not only helps users to find new place effectively but also helps to push advertisements to potential customer groups to increase their popularity and improve their benefits.

POI recommendation is defined as follows: $\mathrm{U}=\{u_1,u_2,\Lambda\ ,u_N\}$ represents the user set, $\mathrm{L}=\{l_1,l_2,\Lambda\ ,l_N\}$ represents the interest point set. The user's check-in information can be transformed into a user-POI frequency matrix, each element c_{ui} represents the user's attendance at the POI. The check-in frequency reflects the user's preference for the POI. The goal of the point of interest is to learn its implicit preference based on the user's check-in history and recommend appropriate points of interest that will be visited by the user.

This paper proposes a point of interest recommendation architecture that combines time characteristics and collaborative filtering, as shown in Fig. 2, including time and user check-in feature analysis, recommended algorithm and recommended generation of three parts.

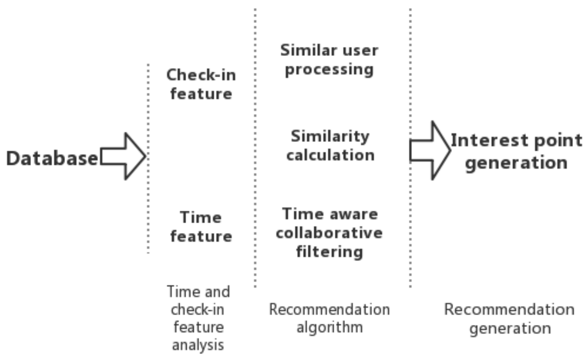

Figure 2. Recommendation framework

A. User 's Check-in Feature Analysis

By counting the number of users in the hourly period of time, we can initially understand the user's check-in habits. There is a gathering of time during the day, that is, most users choose to check-in time. The evening is the peak of the check-in, it infers that most people sign in the gets off work time, and the trough is in the morning, which is consistent with the real-life situation. In the morning, during a period people just get up, the check-in number is relatively small, and with the passage of time, the number of activities gradually increased, more people choose to check online.

B. Modeling from Temporal Attributes

According to the previous description, the user's check-in time distribution is not uniform, in different time periods show different sign in preferences. According to it, time is treated as a property of check-in. To understand the relationship between such attribute, a concept of temporal priority is introduced. Each priority corresponds to one hour of the day. A time variable is defined as $\in [1, T]$, $T = 24$, representing 24 hours a day. For example, "2017-7-22 11.30.00 am", check-in time $t = 11$, indicates the interval belonging to$[11, 12]$.

$U_t \epsilon \mathbb{R}^{m\times d}$ is a matrix which represents the user's preference according to temporal features. With the change of time, the state of the user is changing constantly, but the change of a specific geographical location is very small. We may assume that the effect of the time factor on the location is almost negligible relative to the user variable. Therefore, the location characteristics can be defined as time-independent, represented as $L \epsilon \mathbb{R}^{n\times d}$. The object is to find proper UL^T to approximate the check-in matrix C at each time state t and excavate the aggregation phenomenon. We abstract time-dependent user check-in preferences into the following optimization problems:

$$\min_{U_t\geq 0, L\geq 0} \sum_{t=1}^{T} || N_t \odot (C_t - U_t L^T) ||_F^2 + \alpha \sum_{t=1}^{T} || U_t ||_F^2 + \beta || L ||_F^2 \tag{1}$$

$C_t \in \mathbb{R}^{m\times n}$ is the corresponding User-Location check-in Matrix in the state of time t, N_t and C_t share the same dimensions, at a time state t, $N_t(i,j) = 1$ means a user numbered i has visited location j . $N_t(i,j) = 0$ means opposite. The notation $\odot$ is the *Hadamard product* operation. For two matrices A, B of the same dimensions $m \times n$, the product of $A \odot B$ is a matrix, of the same dimension as the operands, with elements given by $A(i,j) \times B(i,j)$ [13]. $|| \, ||_F^2$ is used to compute the square of the matrix norm[14]. α and β are non-negative parameters of U_t and L.

C. Temporal Signature Model

It is more practical to recommend the user's point of interest at a particular time. On the basis of time analysis, we found the discreteness and continuity characteristics of time, and analyzed the check-in and transfer behavior of users in a specific time according to the time characteristics, so as to get the user behavior preference.

In order to integrate the information of the check-in time, the design of the time segmentation is of the utmost importance. The existing discrete time warp method divides the time into 24 hours, which not only causes the data to be sparse in the unit time, but also causes the loss of the associated information and social attributes between some parts of the time. All check-in time records according to the different days of a week, will be divided into seven parts, each part of a day, and every day will be divided into several time periods. The time of day is broken down as follows:

(0-7) for the before dawn,
(7-9) for the morning,
(9-12) for the forenoon,
(12-14) for the noon,
(14-18) for the afternoon,
(18-24) for the night.

Consider the following two points:

- The time of the rest of the human society is regular;
- The appropriate grouping of time helps to reduce the amount of computation.

V. ALGORITHM ANALYSIS

Fig. 3 shows the workflow of the POI recommendation framework with the integration of temporal features. Fig. 4 is the check-in matrix C where "X" represents the frequency of check-in at the corresponding interest points of the observed user. "?" Represents the login preferences of the user who will be inferred by the recommendation framework at the unreachable interest point. This framework can be divided into three parts: time division, factorization, and aggregation.

- Divide the original POI check-in frequency matrix C into T sub-matrices according to time state t. Each sub-matrix only contains the check-in behavior that occurs at the corresponding time state.
- Each sub-matrix is decomposed into user check-in preference matrix and position feature matrix, and the position feature matrix is shared by all user check-in preference matrices.
- The corresponding low-rank approximation matrices are constructed and aggregated to represent the user's check-in preferences at each location.

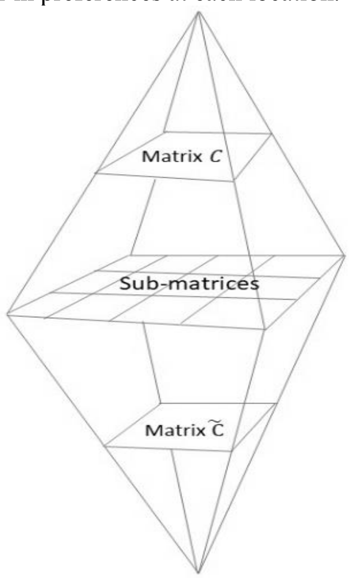

Figure 3. Recommendation Workflow

	L_1	L_1	L_1	…	L_n
U_1	x			X	
U_2		?			X
U_3	x		X	?	X
…		X			X
U_m				X	

Figure 4. Check-in Matrix C

Through the analysis of user social information and check-in time characteristics, this paper proposes a POI recommendation algorithm integrating temporal feature. The algorithm is divided into three main steps, the first step is similar user filtering processing, the second step is similarity calculation, and the third step is to establish recommendation algorithm.

A. Computer Similar Users

We use a similarity detection algorithm FSUA proposed by Y. Song, Y. Si, W. Liu and H. Zhang[15].

Step 1. Statistics target user u signed the location of the collection L_u, for the target user u looking for the same location of the similar user v.

Step 2. Calculate the number of target users u and similar users v check in the same location $num_{u,v}$.

Step 3. If $num_{u,v}$ is greater than the pre-set threshold m, then similar user v meets the requirements, otherwise, similar user v is filtered out.

Step 4. Iterate the above steps to get the filtered similar user set SU.

B. Similarity Calculation

Time is divided into multiple periods, defining more similarities between users with more of the same number of check-in times over the same period. Then calculate the cosine similarity according to the following formula. X_i and Y_i to represent the user's check-in location vector.

$$\cos\ (\theta) = \frac{\sum_{i=1}^{n} X_i \times Y_i}{\sqrt{\sum_{i=1}^{n} {X_i}^2} \times \sqrt{\sum_{i=1}^{n} {Y_i}^2}} \quad (2)$$

C. Location Recommendation with Temporal Effects

The FSUA algorithm is called to get the filtered similar user set SU
Import check-in matrix C, α, β and time states $\{1, 2, \ldots, T\}$
Divide C into $\{C_1, C_2, \ldots, C_T\}$ according to temporal state T
Generate $\{Y_1, Y_2, \ldots, Y_T\}$ from $\{C_1, C_2, \ldots, C_T\}$
Construct $\{\Sigma_1, \Sigma_2, \ldots, \Sigma_T\}$ from $\{C_1, C_2, \ldots, C_T\}$
Initialize $\{U_1, U_2, \ldots, U_T\}$ and L with a random function
while Not Convergent do
    for $t = 1 : T$
        for $i = 1 : m$
            for $k = 1 : d$
$U_t(i,k) \leftarrow U_t(i,k)\sqrt{\frac{[(Y_t \odot C_t)L + \lambda\Sigma_t U_t{-}1](i,k)}{[(Y_t \odot U_t L^\top)L + \lambda\Sigma_t U_t + \alpha U_t](i,k)}}$
            end
        end
    end
    for $i = 1 : n$
        for $k = 1 : d$
$L(i,k) \leftarrow L(i,k)\sqrt{\frac{[\sum_{t=1}^{T}(Y_t \odot C_t)^\top U_t](i,k)}{[\sum_{t=1}^{T}[(Y_t \odot U_t L^\top)^\top U_t + \beta L](i,k)}}$
        end
    end
end
for $t = 1 : T$
    Let $\tilde{C}_t = U_t L^\top$
end
for $i = 1 : m$
    for $j = 1 : n$
        Let $\tilde{C}(i,j) = f(\tilde{C}_1(i,j), \tilde{C}_2(i,j), \ldots \tilde{C}_T(i,j))$
    end
end
return $\tilde{C}$

VI. Experiment and Analysis

A. Experiment Setting

1) Dealing with unpopular locations and filter invalid data: The data set was collected from Gowalla, a large-scale location-based social networking site. It allows users to check in and share their moments in different locations. By analyzing the time and space information associated with these check-in data, we can extract the original data of the points of interest recommendation. All users are grouped into 1083 categories according to their behavioral characteristics, and the locations are divided into 251 categories by region and function. Seven days a week in accordance with social attributes is divided into 42 categories.

TABLE I. Statistical Analysis of Data Set

statistical analysis	**User**	**Location**	**Time**	**Records**
	1083	251	42	227428

2) Assessment Method: The current recommendation algorithm cannot be immediately verified in the large user group, but at the same time to ensure the experimental results have a high credibility, about 85 % of the user check-in data was used as the input of the algorithm, leaving 15 % as the user next real access value. The above training set is used to train the recommended results for each period from high to low, and select *k* points with higher values as the recommended results. Matching the recommended results with 15 % of the validation set, and if the same point exists, the current recommendation is considered valid.

TABLE II. Data Partition

	Number	**Percentage**
Original check - in data	227428	100%
Training data	191702	84.29%
Testing data	35726	15.71%

B. Experimental Result

The sub-matrices are obtained by CP decomposition of the time-divided matrix. Next, the low rank approximation matrix is constructed by sub-matrices. We use these approximation matrices to build a final matrix, where is the recommendation result. In the result matrix, each row represents a time period for a user, each column represents a class of locations. The values in the grid point represent the recommended coefficients for the current time and place for the current user. The recommendation algorithm here excavates the hidden information from the original sparse matrix, which makes the recommended content greatly enriched. The following is a comparison of the number of non-zero points in the User-Location-Time matrix.

TABLE III. Non-zero Rate

	Total points	**Non-zero points**	**Non-zero rate**
Raw matrix	11416986	134778	1.19%
Result matrix	11416986	11247536	98.52%

C. Results Assessment

For each row of the results, select the 5 points with the highest value as the recommended location, so total of 1083*251*5 locations are recommended for all 1083 users. This is the recommended result calculated only through some training sets, and for the rest of the data that is not involved in training, we use it as a user's simulated access over the next time. Match the recommended value with the simulated value, and the current recommended result can be considered as a hit. For the previously recommended results, the accuracy of the calculation is as follows.

TABLE IV. Recommendation Result

Number of testing sets	**Number of accessed**	**Accuracy rate**
35726	30246	84.66%
All recommended quantities	**The number of hits**	**Hit rate**
227430	30246	13.30%

VII. Conclusion

In this paper, a recommendation algorithm integrating temporal features is proposed to model the user's check-in behavior in social networks. The algorithm combines the time effect and the user social effect on the basis of the user's geographic and social information, and this effectively overcomes the problem of low quality data. We first clean the raw data: delete the wrong data, fill the missing data, and merge the redundant location data. For the user data, the corresponding algorithm is used to merge the users with similar behavior to further solve the problem of sparse data. And then divide the processed user-location matrix into time-based sub-matrices according to time. Each sub-matrix is divided into a preference matrix and a location feature matrix using a non-negative matrix decomposition algorithm. The low-rank approximation matrices in each period are calculated by the correlative optimization algorithm. And then gather them into a general approximation matrix to represent the user's check-in preferences at each location. Through experimental verification, this algorithm can provide users a better recommendation by exploring time characteristics from former access records.

As part of our future work, we will further explore the effect of friend attributes on user preferences, thereby improving the performance of the POI recommendation.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61673295), Natural Science Foundation of Tianjin (No. 15JCYBJC46500), National-level College Students Innovative Entrepreneurial Training Plan Program (No. 201710060028).

I would like to thank Yu Wang, Fan Wang, Shujie Li and Qina Fang, who friendly encourage me in writing the thesis.

REFERENCES

[1] X.Y. Ren, M. N. Song, and J. D. Song, “Context-Aware Point-of-Interest Recommendation in Location-Based Social Networks,” Chinese Journal of Computers (2017).

[2] M. J. Pazzani, “A framework for collaborative, content-based and demographic filtering,” Artificial intelligence review. 1999 Dec 1;13(5-6):393-408.

[3] A. Gunawardana and G. Shani, “A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. Journal of Machine Learning,” Volume 10, pp.2935-2962

[4] H. Kautz, B. Selman and M. Shah, “Referral Web: combining social networks and collaborative filtering,” Communications of the ACM, Volume 40 Issue 2, March 1997

[5] J. Bao, Y. Zheng, D. Wilkie and M. Mokbel, “Recommendations in location-based social networks: a survey,” Geoinformatica, 2015, 19(3):525-565.

[6] H. Gao, J. Tang, X. Hu, and H. Liu, “Exploring temporal effects for location recommendation on location-based social networks,” ACM Conference on Recommender Systems. ACM, 2013:93-100.

[7] C. Cheng, H. Yang, I. King, and M. R. Lyu, “Fused matrix factorization with geographical and social influence in location-based social networks,” In Proceedings of the 26th AAAI, 2012.

[8] H. Gao, J. Tang and H. Liu, “gSCorr: modeling geo-social correlations for new check-ins on location-based social networks,” Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.

[9] M. Ye, P. Yin, WC. Lee and DL. Lee, “Exploiting geographical influence for collaborative point-of-interest recommendation,” 2011:325-334.

[10] Download URL: http://snap.stanford.edu/data/loc-gowalla.html

[11] YZ. Wu, J. Lai, and YH. Wu, “Preprocessing of LBS based POI Data Updating. Computer and Digital Engineering,” vol. 40(8), 2012, pp. 87-89.

[12] JX. Cao, Y. D, PW. Yang, T. Zhou and B. Liu, “POI Recommendation Based on Meta-Path in LBSN,” Chinese Journal of Computers, 2016.

[13] https://en.wikipedia.org/wiki/Hadamard_product_(matrices)

[14] https://en.wikipedia.org/wiki/Matrix_norm

[15] Y. Song, Y. Si, W. Liu and H. Zhang, “Point-of-interest recommendation algorithm combining temporal features and collaborative filtering,” Journal of Chinese Computer Systems, 2016.