A Context-Awareness Personalized Tourist Attraction Recommendation Algorithm

Zhijun Zhang¹, Huali Pan², Gongwen Xu¹, Yongkang Wang¹, Pengfei Zhang¹

¹School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China
²Business School, Shandong Normal University, Jinan 250014, China
Emails: zzjsdcn@163.com  panhuali@sina.com  xugongwen@sdjzu.edu.cn
m18366132486@163.com  zpf@sdjzu@sina.com

Abstract: With the rapid development of social networks, location based social network gradually rises. In order to retrieve user’s most preferred attractions from a large number of tourism information, personalized recommendation algorithm based on the geographic location has been widely concerned in academic and industry. Aiming at the problem of low accuracy in personalized tourism recommendation system, this paper presents a personalized algorithm for tourist attraction recommendation – RecUFG Algorithm, which combines user collaborative filtering technology with friends trust relationships and geographic context. This algorithm fully exploits social relations and trust friendship between users, and by means of the geographic information between user and attraction location, recommends users most interesting attractions. Experimental results on real data sets demonstrate the feasibility and effectiveness of the algorithm. Compared with the existing recommendation algorithm, it has a higher prediction accuracy and customer satisfaction.

Keywords: Location services, social network, personalized recommendation, user context, social computing.

1. Introduction

With the development of wireless networks and location technology, it is easier to identify and share personal location information when users are moving. After adding a spatial dimension, the virtual social networks will be returned to the real world. Due to the fact that “position” is one of the most important aspects in daily life, Location Based Social Networking (LBSN) developed rapidly. Some social networking sites appeared, such as Gowalla, Foursquare, Flickr. Many travelers like
to share their travel experiences and to upload photos on these sites, or to share
information about their location and service online.

Compared with traditional online social networks recommendation system,
location based social network recommendation system introduces location
information; it can help recommendation system to find the user’s preferences.
Currently, the type of location based social network recommendation systems
includes friend recommendation, activities recommendation and location
recommendation. Among them, location recommendation has become the focus of
recent research. However, most of the sites in position based social networks are
checked in by very few people, even some newly developed locations are never
checked in. So, position based recommendation algorithm inevitably faces some
problems such as data sparsity and cold-start, which makes the recommendation
accuracy low and cannot meet the personal needs of tourists.

In recent years, researchers have designed a lot of tourism recommendation
systems; some of them have been running for travel websites. User’s query criteria
level of current tourism recommendation system only stays in a syntactic or
semantic layer and does not reach pragmatic level. Most recommendation systems
do not consider the impact of context on the recommendation result, so
recommendation results do not change with the updates of context information,
such as time, weather conditions, location, participants, etc. Therefore this article
will discuss personalized travel recommendation system by introducing context
model.

The remainder of this paper is organized as follows. In Section 2, we provide
an overview of related work. Section 3 introduces the problem description and the
related definitions. In Section 4 is put forward the improvement of the algorithm for
personalized tourist attractions recommendation – RecUFG Algorithm. Section 5
provides the experiment dataset and evaluation metrics, introduces the experiment
results and its analysis, followed by the conclusion and future work in Section 6.

2. Related work

In recent years, there has been presented a lot of research about travel
recommendation and tourism search. Mainly, there are three directions of work as
follows. First, recommending travel routes by analysing photos shared by users of
the social network. Second, to tap popular tourist destination and the route by GPS
track. Third, providing location based route search by analysing check-in record
shared by users in their daily lives. In the area of tourism, the two most successful
personalized travel recommendation technologies are Tripmatcher of Triplehop
(applied by www.ski-europe.com) and specialists recommended platform-MePrint
of VacationCoach (applied by travelocity.com). The two recommendation systems
attempted to simulate traditional travel agency, discussing with users and helping
them to simplify the search process for holiday destinations [1].

Cyberguide put forward by A b o w d et al. [2] is the first mobile context-
aware guide system; it can be designed according to context information, such as
user’s current preferences, and to history location. In the collaborative filtering
approach adopted by Horozov, Narasimhan and Vasudevan [3], the system is dealing with the cold start problem assuming that users who live close to each other like similar attractions; so the system introduced virtual user, sorted and rated the restaurants in advance, and this way it could find the same type of users with rare evaluation. The system could randomly generate option based on similarity drive and recommend user the nearest object to him with the highest average rating.

J. Kim, H. Kim and J. Ryu [4] designed a TripTip system, which could make similarity recommendation by analysing user’s historical sites, and calculate the similarity based on his access label. However, doing that requires the user to give his current location, so it is not entirely suitable for travel plans. Baltuñas et al. [5] proposed a customizable context-based POI (Points Of Interest) recommendation system; the factors such as weather, time, and attractions are made as situation to produce the recommendation. Zheng, Burke and Mobasher [6] thought that if the system introduces less relevant factors, it might have a negative impact on the recommendation effectiveness and efficiency. Ye, Yin and Lee [7] obtained data from FourSquare and provided location recommendation based on collaborative filtering and geography factor. Cheng et al. [8] made similarity estimation on picture with geographic label, and produced personalized recommendations by designing geographical similar model. Song et al. [9] proposed a recommendation for a short trip; it considered the diversity of attractions and sights which popularity combines time limit and the set of categories.

Zhang and Liu [10] proposed personalized recommendation algorithm which merged timing behavior and trusts. The algorithm considered trust relationship between users and time factors, so it ensured the recommendation timeliness. In another literature [11], he proposed a mobile social network recommendation algorithm which merged a variety of context information. The algorithm used user’s location and time information to dig potential users of social relations; it effectively solved the problem of recommendation accuracy.

Li and Lu [12] pointed out that the future development direction of tourism personalized recommendation services should be integration of space-time; it is thought that the integration of the three platforms: the tourism website, LBS (Location Based Service) website, and ESNs (virtual Electronic Social Networks) will become a breakthrough of development. The recommendation system is dependent on user’s information, which includes basic personal information, history and behaviour interaction information. Some recommendation systems will integrate a variety of information to provide a more intelligent recommendation for the users, but will also expose personal privacy. In fact, the threat of user location identification has been considered as one of the biggest barriers in using context-aware services [13]. Therefore, we need to find methods to protect user privacy. These methods should ensure recommendation validity and accuracy without exposing user's privacy and sensitive information. On the other hand, we should achieve tourism informatization and share travel information completely. Some researchers introduced personalization to tourism recommendation and made a personalized recommendation based on existing tour data package [14]. In addition,
some studies [15] mined track from LBSN data, according to user’s travel time and geographical requirements, and automatically generated travel route for the user. However, most of these studies only made recommendations based on history and geography label information; they do not take into account other contextual information of the social network, such as a friendship trust relationships. When people choose travel destinations, they often tend to choose some places where friends had been to or places with a certain degree of popularity.

In conclusion, on the basis of previous studies, considering some context information, such as user preferences, friendships and geographic location, this paper proposes the personalized recommendation algorithm – RecUFG. We make a fully experimental study on a real data set and compare the accuracy of recommendation to verify the validity of RecUFG Algorithm.

3. Description of the problem and related definitions

3.1. Description of the problem

Given user \( u \), pictures shared by \( u \) on LBSN and his friends, according to his geographical location, friendship and collaborative filtering algorithms find similar users to \( u \), estimate other preference attractions for \( u \), and then calculate his best preference attractions. Recommendation algorithm can effectively return to preference attractions with the requirements of the user.

3.2. Related definitions

**Definition 1. Location-Based Services (LBS).** LBS combine mobile communication network and satellite positioning technology. It uses GIS, GPS, network communication and multimedia technology, with the support of the electronic map platform, through the cooperation of mobile terminals and wireless networks, determines the mobile user's location of real time, and timely responses to the user request, to provide users with the appropriate spatial information services.

**Definition 2. Check-in data structure.** Check-in data typically includes user ID, time, location, points of interest, attractions, and other categories of information; the structure is as follows:

\[(\text{UserID}, \text{Time}, \text{Longitude}, \text{Latitude}, \text{POI}, \text{Category}).\]

**Definition 3. Tourism context recommendation model.** Recommendation weight of tourism context recommendation system is calculated as follows:

\[R: \text{User} \times \text{Tourist contents} \times \text{Context} \rightarrow \text{Rating}.\]

After introducing tourism context into recommendation system, the original two-dimensional rating matrix becomes a three-dimensional model; user has a user model with structure as follows:

\[(\text{UserID}, \text{Name}, \text{Address}, \text{Age}, \text{Gender}, \text{Profession}, \text{Interest}, \text{Tourism purpose}).\]

Tourist contents are tourism content model with the structure as follows:

\[(\text{Accommodation}, \text{Food}, \text{Shopping}, \text{Attractions}, \text{Transportation}, \text{Weather}, \text{Location}).\]
Context is contextual information related to the tourism recommendation system; the structure is as follows:

(identity, state, location, time, companion).

3.3. Recommendation framework

The general framework of tourism recommendation system is shown on Fig. 1.

![Fig. 1. A general framework of travel recommendation system](image)

4. Improved location recommendation algorithm – RecUFG Algorithm

Firstly, this paper studies currently popular personalized tourist attraction recommendation algorithms, points out the shortcomings of these algorithms, then proposes an improved personalized attraction recommendation algorithm – RecUFG Algorithm, which is a fusion of a user preference, trust relationships and location factors.

4.1. User-based collaborative filtering algorithm

In the location-based social network, users and location are linked by check-in. User’s check-in reflects his preferences for a variety of locations [16]. For location recommendation, user-based collaborative filtering algorithm mainly includes three steps: calculating similarity between the user, calculating interest degree of user to candidate locations and top-N recommendation. The similarity of user \( u_i \) and user \( u_k \) can be calculated by cosine similarity

\[
W_{i,k} = \frac{\sum_{j \in \mathcal{L}} c_{i,j} c_{k,j}}{\sqrt{\sum_{j \in \mathcal{L}} c_{i,j}^2} \sqrt{\sum_{j \in \mathcal{L}} c_{k,j}^2}},
\]
where \( L \) represents a common check-in set of user \( u_i \) and user \( u_k \), \( c_{i,j} \) and \( c_{k,j} \) represent check-in at location \( l_j \) of user \( u_i \) and user \( u_k \), respectively.

User-based collaborative filtering algorithm primarily recommends according to the similarity between users, without considering the effect of trust between friends, and without consideration of the specific location of interest. So, its recommendation result is less accurate, it cannot meet the tourist’s personal need.

4.2. Friend trust relationship-based collaborative filtering algorithm

Levandoski et al. [17] filtered out the location far away from user's current location, but it does not consider the effect of social network on user in LBSNs. Ye et al. [18] pointed out that the most effective recommendation algorithm is location recommendation algorithm which is based on naive Bayes classification, followed by user-based collaborative filtering, and the last is social network-based collaborative filtering. Farence, Ye and Lee [19] studied the effect of the algorithm on recommendation when a user is far from home. It found that the most effective recommendation algorithm is social network-based collaborative filtering, followed by user-based collaborative filtering, and the last is location recommendation algorithm which is based on naive Bayes classification. It is not difficult to see that the trust relationship between users has a relatively large impact on recommendation result.

Liu and Pan [20] put forward a PTLR; this method calculates user preferences through user attraction pictures matrix, calculates similar neighbour according to intimacy trust friendship, produces recommendation by user interest preference, suitability of attractions time and candidate surrounding attractions. Experimental results show that the algorithm PTLR can effectively improve the recommendation accuracy. There is a weak relationship in LBSN, that is \( u \) adds \( v \) as a friend and \( u \) is not necessarily a friend of \( v \). There is a relationship of concern and to be concerned means concern to each other between users. Given \( SU \) is friend set of \( u \), \( SV \) is friend set of \( v \), \( CU \) is concerned friend set of \( u \), if \( u \) and \( v \) are friends to each other and there is another user who concerns about \( u \) and \( v \) simultaneously, it is considered that \( u \) and \( v \) is likely to know each other in real life. Using (2) to compute the intimacy IF of user and friends in LBSN is

\[
IF = \frac{CU(SU \cap SV) + \tau(SU \cap CU)}{SU \times CU},
\]

\[
\tau = \begin{cases} 
1 & \text{if } u \text{ and } v \text{ both are friends,} \\
0 & \text{if } u \text{ followed } v \text{ but } v \text{ not followed } u. 
\end{cases}
\]

Through data analysis we can find that in social networks, the more influence of friends to each other, the more similar are the attractions. If there are more similarity in age, disposable time and economic conditions between users and friends, there is more reference value. Therefore, TSCF method lays particular stress on evaluation of landscape in calculation of similar user. We can get new user similarity with complex rating similarity:

\[
sim_{UF}(u, v) = \lambda \times \text{sim}(u, v) + (1 - \lambda) \times IF.
\]
According to simUF(u, v) to find top-N of users as similar neighbour set SU. Based on the similar users and trust friendship, TSCF algorithm gets interest rates of attractions that user does not visit, then makes appropriate tourism attractions recommendation.

PTLR takes full advantage of trust relationship between friends and similarity between users, but it ignores the impact of location on the recommendation result, so its recommendation accuracy should be further improved.

4.3. Location-based collaborative filtering algorithm

Chen, He and Liang [21] put forward a personalized recommendation method LTCF which combined LBS and social network tag. On calculating the location similarity of user, this method takes into account users’ experience index. They believe that distance is closer, location similarity between users is higher; distance is farther, location similarity between users is lower. According to the geographical location, they divided area for users, as the target user for any one position with neighbouring location of the user. Set i as target user, j is a user who is near to location of i. Location similarity simD(i, j) of user (i, j) can be represented by

\[
\text{simD}(i, j) = \begin{cases} 
0 & d(i, j) > 20000, \\
0.2 & 10000 < d(i, j) \leq 20000, \\
0.4 & 5000 < d(i, j) \leq 10000, \\
0.6 & 3000 < d(i, j) \leq 5000, \\
0.8 & 1000 < d(i, j) \leq 3000, \\
1 & d(i, j) \leq 1000. 
\end{cases}
\]

LTCF mainly takes advantage of location similarity between users, but it ignores the similarity and trust relationships between users, so its recommendation accuracy is lower.

4.4. Improved algorithm – RecUFG Algorithm

To further enhance tourism attractions recommendation accuracy for users, this paper presents a personalized recommendation algorithm – RecUFG Algorithm, which combines a variety of tourism attractions context. After calculating user interest similarity, use relationship similarity and user distance similarity with the above three methods, RecUFG Algorithm uses weighting method deal with the three similarities, to calculate the integrated similarity simUFG(a, b) between users according to

\[
\text{simUFG}(a, b) = \alpha \cdot \text{simU}(a, b) + \beta \cdot \text{simF}(a, b) + (1 - \alpha - \beta) \cdot \text{simG}(a, b),
\]

where, \(\text{simU}(a, b)\) is interest similarity of user \(a\) and user \(b\), \(\text{simF}(a, b)\) is relationship similarity of user \(a\) and user \(b\), \(\text{simG}(a, b)\) is location similarity of user \(a\) and user \(b\). Weighting factor \(0 < \alpha < 1, 0 < \beta < 1, \alpha\) and \(\beta\) indicate weights of user interest and user relationship when calculating similarity between users, \(1 - \alpha - \beta\)
is weights of user location when calculating similarity between users. If the value of $\alpha$ is greater than 0.5, it means user relationship is more important in the calculation of user similarity, if the value of $\alpha$ is less than 0.5, it means user interest and user location is more important in the calculation of the user similarity. By setting different values $\alpha$ and $\beta$, we can combine the three kinds of similarity to further improve the quality of recommendation algorithm.

Equation (5) may be equivalently expressed as

$$\text{simUFG}(a, b) = \eta \cdot \text{simUF}(a, b) + (1 - \eta) \cdot \text{simG}(a, b).$$

Here, $\text{simUF}(a, b)$ represents a new user similarity that emphasis on intimacy friend rating, $\text{simG}(a, b)$ is location similarity of user $a$ and $b$. Weighting factor $0 < \eta < 1$ represents the weights of user interest and user relationship in calculation of similarity between users, $1 - \eta$ represents the weights of user distance in calculation of similarity between users.

Interest of user $u_i$ to location $l_j$ can be calculated by

$$p_{i,j} = \frac{\sum_{u_k \in U} \text{simUFG}(i, j) \cdot p_{k,j}}{\sum_{u_k \in U} \text{simUFG}(i, j)},$$

where, $\text{simUFG}(i, j)$ is the integrated similarity of user $u_i$ and $u_j$ calculated from Equation (5); according to integrated similarity, calculating the user set of top-$N$ similar with user $u_i$, represented by $U'$, and $U' \in U$. Finally, recommending the location whose interest ranking top-$N$ in the candidate locations to the user $u_i$.

5. Experimental results and analysis

5.1. Introduction to data set

This paper chooses Foursquare, Gowalla, BrightKite and Flickr four social network sites, which are based on the user location information (LBSN), uses the check-ins information in these sites as data set. Foursquare website provides user targeted social network services. Mobile phone users are encouraged to share with others their current geographical location. Gowalla is a mobile Internet application, which provides location-based services, allowing users to know the location of their friends, and share their favourite places, discover new world around them. BrightKite is LBSN application that allows users to check in at any location, its feature is that allows users to use the computer to check in. Brightkite is a pioneer of LBSN application, but with the popular of location-based social network, Brightkite application is behind of its competitors FourSquare and Gowalla. Flickr is photo-sharing website of Yahoo; it provides comprehensive, first-class and efficient pictures service, which includes 9930 pictures and 700 000 video URL with associated data (title, camera type, description, tags), about 49 million pictures are geotagged; various comments, preferences and social network data can be used by Flickr API. In the experiment, we select the user who has at least 10 location check-ins; we extract partial data from February 2015 to January 2016 from the above four sites as experimental data set.
5.2. Effect of location on user check-in behavior

Scellato and Mascolo [22] analyzed data from location-based social networks Foursquare. The results indicated that approximately 40% of friend relationship is within 100 kilometres [23]. DeScioli et al. [24] revealed that users' social relationships and their geographical distance are highly correlated. Users live nearby are more likely to become friends. Therefore, in LBSNs, historical location information of user can increase the effectiveness and efficiency of friends' recommendation.

Next, we analyze the relationship between the check-in probability and the user-friends distance by experiment. Here put check-in frequency of user as a function of user-friends distance, statistics on Foursquare, Brightkite and Gowalla data set are shown in Fig. 2. As we can be seen from Fig. 2, in the three data sets, distribution function is very similar, and they emerge as a power-law distribution function. When the distance is less than 100 km, distribution decreases rapidly. When the distance is more than 100 km, distribution decreases slowly.

![Fig. 2. Relationships between the check-in probability and the user-friends distance](image)

Randomly select two users, observe the function of check-in frequency and user-random distance. Experimental results on Foursquare, Brightkite and Gowalla these three data sets are shown in Fig. 3, check-in probability and the distance of random user do not attenuate completely. When the distance is more than 1000 km, check-in probability increases slightly, this is due to the uneven population density.

![Fig. 3. Relationships between the check-in probability and the distance of random users](image)
With travel distance increases, how changes the probability of their meeting friends. Results on the above three data sets are shown in Fig. 4, when travels distance is over 100 km, their meeting friend’s probability increases 30%. So, we can see that people like do long-distance travel where they have friend there, but short-distance travel is not affected by social relations. In addition, we also find that the impact of friendship on mobility is far greater than that of mobility on new friendships.

If trajectory of two users is similar, how the probability they become friends will change. First, we define user trajectory similarity; user’s trajectory can be seen as a vector, the value of $i$ node in the vector indicates frequency of user check-in in $i$, trajectory similarity can be calculated by cosine similarity as

$$\text{sim}(u_i, u_j) = \frac{f_i \cdot f_j}{|f_i|_2 \times |f_j|_2}.$$  

Set probability of becoming friends among users as function of trajectory similarity, the experimental results are shown in Fig. 5. It can be seen that the more similar trajectory of users, the greater probability of becoming friends.

5.3. Accuracy evaluation of attraction recommendation

In this paper, we use Precision to evaluate performance of recommendation algorithm. Recommendation result accuracy Precision is defined as
\begin{align}
\text{Precision} &= \frac{tp}{tp + fp},
\end{align}

where, \(tp\) represents user’s real rates of attractions in the recommendation list, \(fp\) represents the rates of attractions that user dose not real select in the recommendation list.

To evaluate accuracy of the algorithms, we randomly select 80% users as training set, generating user-attractions matrix. We set 20% users as a test set to determine their interest preferences, calculate the similarity between users, and use RecUFG recommendation algorithm recommend candidate sites for users. We set a comparison experiment, compare with the traditional collaborative filtering methods UserCF, Personalized Context-aware Rank (PCR) method \[25\]. The accuracy of different recommendation algorithm is shown in Fig. 6.

![Fig. 6. Performance comparisons of recommendation algorithm](image)

As it can be seen from Fig. 6, the performance of recommendation algorithm UserCF is the worst, followed by PCR recommendation algorithm, however, RecUFG has a relatively good recommendation performance compared with PCR and UserCF. Compared with precision@5, the recommendation accuracy of precision@10 has significantly improved. It is because when most users choose a few attractions, they mainly select popular attractions, there is no significant difference in individual preferences. When users choose more attractions, especially choose long-distance travel attractions, the accuracy of recommendation algorithm will be higher.

5.4. The effect of parameter \(\eta\) on accuracy

The similarity calculation equation is

\begin{equation}
\text{simUFG}(a, b) = \eta \cdot \text{simUF}(a, b) + (1 - \eta) \cdot \text{simG}(a, b),
\end{equation}

where \(\eta\) is a variable that reconciles PTLR methods and user location similarity. Weights factor \(0 < \eta < 1\), by setting different value of \(\eta\), combining three kinds of similarity, to further improve the quality of recommendation algorithm. The effect of different value \(\eta\) on recommendation accuracy is shown in Fig. 7.
As can be seen from Fig. 7, the accuracy of LTCF method is not high when \( \eta = 0 \); when \( \eta = 1 \), recommendation depends only on the user interest and friends similarity, independent of the location, and accuracy decreased significantly. We found that when \( \eta = 0.7 \), performance is best. That means that \( \eta \) can affect accuracy, it should be a reasonable value.

6. Conclusion and future work

This paper introduced context-aware technology, modelling of context in tourism recommendation system, and described the design of prototype architecture of context-awareness personalized travel recommendation system; the process of context modelling and recommendation mode of personalized space information service and tourism information services were also described. After analysis of the current research, we proposed a recommendation algorithm that is not only taking in account user’s interest preferences but is also considering the relationship of trust between friends; the recommendation algorithm is based on temporal data in a social network taking the geo-tagging photo as a research object. It used the geo-tagging feature to extract similarity of the user’s location, combined with the user’s current context information, filtered and recommended suitable attractions for users. In spite of the improvement on research about tourism recommendation presented in this paper, there are still many issues that need to be studied and solved further. The main data source is Flickr site, and there are numerous social network sites on the Internet. If mining information from more data resources, recommendation effect may be better, so we should make a further study that obtains multiple data sources in future.

Acknowledgments: This paper is supported by the Science and Technology Development Planning of Shandong Province (No 2014GGX101011, No 2015GGX101018), Project of Shandong Province Higher Educational Science and Technology Program (No J12LN31, No J13LN11, and No J14LN14) and Jinan Higher Educational Innovation Plan (No 201401214, No 201303001).
R e f e r e n c e s


