Personalized Recommendation of Tourist Attractions based on LBSN

Huifang Lv*, Wanggen Li, Yang Su, Chengcheng Li and Shixiang Zhang
College of Mathematics and Computer Science, Anhui Normal University, Wuhu 241003, China

ABSTRACT

Photos metadata in Location-Based Social Networks (LBSN) contain rich time and space information, these metadata provide the basis for the research of personalized recommendation of tourist attractions. The existing methods have many problems such as low accuracy of recommendation and single type of attractions recommendation. For those problems, the PRTA-CF algorithm is proposed to improve the method of predicting the user’s preference of attractions that he/she has not been to before, which is used in traditional collaborative filtering. We designed an evaluation model to assess the user’s preference of the attractions he/she has been to. In order to predict the target user’s preference to attractions he/she has not been to before, when we recommended to the target user, we took into account similar users’ recommendation value and the popularity of attractions based on user preference. Experiment shows that compared with the traditional collaborative filtering algorithm and the algorithm only considered similar users’ preference, it can effectively improve the accuracy of personalized recommendation of tourist attractions when considering both.

INTRODUCTION

At present, there is a lot of research using data mining technology for personalized tourism recommendation, and they have made a lot of research results. Memon et al. [1] got user’s travel preference according to his/her past time in one city and recommended another city. They examined their technique on dataset from Flickr which taken from different cities of China. Reference [2] crawled the score and user of attractions of the five cities in China, the tags of scenic spot’s type were introduced into the collaborative filtering algorithm, and then the algorithm based on the social relation and tag and the algorithm based on item and tag were proposed to improve the similarity calculation between scenic spots.

Collaborative filtering [3-5] is the most commonly used recommended technology in the current recommendation system. Tourist attractions recommendation is different from the general recommendation, in real life, the time people travel is often very few, the traditional collaborative filtering couldn’t do well on tourism recommendation. This paper proposed an improved user-based collaborative filtering algorithm named PRTA-CF, which defined the user preference model based on the user’s residence time and the number of photos. It found similar users of the target users. When recommended to the target users, considering both the similar users’ recommendation value of the target attraction and the impact of popular attractions on the user.
PRTA-CF ALGORITHM
The Generation of the Attraction Sets

Images contain a lot of geo-location information, such as the reference [6-8], they proposed different image retrieval methods to identify the geo-location of the images. First we use the Haversine [9] formula to calculate the distance between photos according to the latitude and longitude of each photo, when the distance between the photos are less than the given threshold, photos between them could be regarded as being taken at the same stop point, whose coordinate is represented by the average latitude and longitude of all the pictures taken by the user at this place. Then we use the clustering algorithm named DBSCAN which was used in reference [10] to cluster the stop points of different users, and get the collection of attractions. The density-based spatial clustering of applications with noise algorithm (DBSCAN) is a clustering algorithm, which divides the regions with sufficient density into clusters and discovers clusters of arbitrary shapes in a spatial database with noise, which defines the clusters as the largest set of points of density. After that, we use the TF-IDF algorithm to get the final attractions with the tags. The term frequency-inverse document frequency algorithm (TF-IDF) is a statistical method used to evaluate the importance of a word for a fileset or one of the documents in a corpus.

User Preference Evaluation Model

In general case, a user will take more photos and spend a longer time than other users in the attraction, if he is more interested in the attraction. According to this, we define the user preference evaluation model of the attractions, as shown in equation (1):

$$ r_{u,i} = \frac{t_{b,i} - t_{a,i}}{t_{\text{max},i}} + \frac{n_{u,i}}{n_{\text{max},u}} / \frac{n_{\text{max},i}}{n_{\text{max},v}} $$

Where $r_{u,i}$ represents the user u’s preference rating to the attraction i, we use the spending time and the number of photos of a user to evaluate the preference of the user to an attraction. Where $t_{b,i}, t_{a,i}$ denote the departure time and the arrival time of the user u at the attraction i, $t_{\text{max},i}$ represents the spending time of the user, who spend the longest time of others, $n_{u,i}$ represents the number of pictures taken by user u at attraction i, $n_{\text{max},u}$ indicates that the photos number taken by the user u in an attraction, where the user u has taken the most number of photos of where he have been visited. $n_{\text{max},i}$ indicates the number of pictures taken by the user (with v) who has taken the most number of photos of other users in the attraction i, $n_{\text{max},v}$ represents the number of photos taken by the user v in an attraction, where the user v has taken the most number of photos of where he have been visited.

Personalized Recommendation of Tourist Attractions Based on Improved Collaborative Filtering Algorithm PRTA-CF
THE RECOMMENDATION VALUE OF SIMILAR USERS

When the target user u is required to recommend the attractions that may be of interest, we use the Pearson correlation coefficient to calculate the user similarity, then
the top-N users with the highest similarity to \( u \) are set as similar neighbors, and then find attractions that the similar users have been visited but the target user has not visited. The recommended value of the target attraction \( k \) of the neighbor set can be calculated, as shown in equation (2):

\[
H(u, k) = \sum_{v \in R_u} \frac{\text{sim}(u, v) \cdot r_{v,k}}{r_{\text{max},v}}
\]

(2)

Where \( R_u \) represents the similar neighbor set of the target user \( u \), \( v \in R_u \), \( \text{sim}(u, v) \) represents the similarity between the target user \( u \) and the user \( v \), \( r_{v,k} \) represents the preference evaluation score of the user \( v \) to the target attraction \( k \), \( r_{\text{max},v} \) indicates the largest preference evaluation score of user \( v \) in the visited attraction.

THE POPULARITY OF THE ATTRACTIONS BASED ON USER PREFERENCE

When a user comes to a place, the local popular spots have a certain appeal to the user, under a certain limit of travelling time, the user will filter the attractions according to their own preferences. The popularity of the attractions based on user preference is proposed based on this situation, we calculate the proportion of the number of photos in the target spots accounts for the number of similar spots with the same tags, and the user’s preference for such attractions, then calculate the product of the two, as shown in equation (3):

\[
M(u, k) = \frac{\sum_{u \in U} \sum_{s \in P_k} n_{u,s} \cdot \sum_{u \in U} \sum_{v \in V} n_{u,v} \cdot \sum_{l \in L} n_{u,l}}{\sum_{u \in U} \sum_{s \in P_k} n_{u,s} + \sum_{u \in U} \sum_{v \in V} n_{u,v} \cdot \sum_{l \in L} n_{u,l}}
\]

(3)

Where \( U \) is the set of all users, \( u \in U \), \( n_{u,k} \) represents the number of photos taken by user \( u \) at attraction \( k \), \( P_k \) represents the similar set of attractions with the same tag as attraction \( k \), \( s \in P_k \), \( L \) represents the set of all attractions , \( l \in L \).

PERSONALIZED RECOMMENDATION OF TOURIST ATTRACTIONS ALGORITHM

Considering the user in the choice of tourist attractions, will also be affected by the local popular spots, so we combine the similar users’ recommend value and the popularity of the attractions based on user preference, to predict the preference evaluation score of the target user \( u \) to the attraction \( k \) which he has not been visited and make a recommendation to him, as shown in equation (4):

\[
P(u, k) = \lambda H(u, k) + (1 - \lambda)M(u, k)
\]

(4)

Where \( \lambda \) is the adjustment parameter, which is used to adjust the similarity of the similar users’ recommendation value and the popularity of the attractions based on user preference.

THE RESULTS OF THE EXPERIMENT

The data set used in this article is the 272150 photos metadata captured by 11320 Flickr users in Los Angeles. Each metadata includes the photo ID, the user ID, the
location (latitude and longitude), and the photo tag information. In order to evaluate the accuracy of the proposed algorithm, 20% of the users were randomly selected from the dataset. A number of attractions were randomly selected from the spots which they have visited, and the photos they took at these spots were removed as the test set. The rest of the dataset is training set. In order to facilitate the experiment, the dataset is processed as follows: (1) Delete the users with fewer than 6 photos; (2) Delete the users who has visited less than 3 attractions.

**Evaluation Index**

As the change direction of the accuracy and recall rate may be inconsistent, we introduced the reconciliation of the accuracy and recall rate as F-measure [11], the higher of the value, the better recommendation performance of the algorithm, the formula defined as:

\[ F \text{- measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \]  \( (5) \)

**The Results of the Experiment**

In order to verify the efficiency of the PRTA-CF algorithm proposed in this paper, we compare with the traditional user-based collaborative filtering algorithm (User-CF), the algorithm just consider similar users’ recommended value, and the combination of interest and popularity algorithm (CIAP) from reference [12]. The experiment result is shown in Figure 1.

![Figure 1. Comparison and analysis of experiment result.](image)

From the experiment result we can see that with the increase of the number of similar users, the F-measure value of different algorithms will increase, because more similar users help to find more preferences of users, making the recommendation more accurate. However, the increase of similar users makes the running time of the algorithm increased, which reduces the recommendation efficiency. Therefore, the
number of similar users in this paper is 10. It can also be seen from the figure that the PRTA-CF algorithm proposed in this paper has higher F-measure value than other algorithms, which shows that the proposed algorithm can more accurately recommend the attractions of interest.

CONCLUSIONS

This paper proposes a PRTA-CF method based on LBSN. We designed an evaluation model to assess the user’s preference of the attractions he/she has been to, based on the travel photos uploaded by users on Flickr. We combined similar users’ recommendation value with the popularity of attractions based on user preference, to predict the target user’s preference to attractions he/she has not been to before, so that we made recommendations to the target user. Compared with the traditional user-based collaborative filtering algorithm and the algorithm only considered similar users’ preference, the results showed that the algorithm proposed in this paper can accurately predict the target user’s preference to attractions he/she has not been to, it improved the accuracy of the personal recommendation of tourist attractions. Due to the limited times of users’ travelling, the number of similar users of the user is often few. In the next step, we will consider the trust between users through the target user’s attention. We will combine the similar users and the trusted users of the target user to make the recommendation together, in order to improve the accuracy of recommendation further.

ACKNOWLEDGEMENT

This work is partially supported by the 2016 Annual University Leading Talents Introduction and Cultivation Planning Project of Anhui Province, China (gxfxZD2016014).

REFERENCES


