Research on Personalized Recommendation Algorithm of Intelligent Travel

Wenjing Wang

College of Information, Business College of Shanxi University, Taiyuan 030031, Shanxi, China

Abstract: A cross-city preference extraction method is proposed to extract the tourism preference information of new tourists in this paper. Because new tourists are relative to a city, but the tourist behavior data in other cities can be known. When the travel preference of new tourists is known, the travel route needs to be generated to match the personalized preference and meet the constraint requirements of tourists. In this paper, the history of the tourists visit record to get the user personalized travel preference, through improved PrefixSpan algorithm to generate more accurate backup route, visitors by time constraints, cost constraints, such as screening conditions meet the conditions of the backup route. Finally, the characteristic information of the backup route is matched with the personalized tourism preference of tourists, and the tourism route that matches the personalized tourism preference of users is recommended to tourists.

Keywords: travel preferences; prefixspan algorithm; backup route

1. Introduction

With the rapid popularization of mobile Internet, information sharing and dissemination have become more efficient. As a result, various information fusion of numerous human beings has been produced, and the result of such integration can be called “collective intelligence”. Such information continuously affects all aspects of human life, including the field of tourism. In people’s travel behavior, there is one is extremely important, is the travel route planning. Therefore, it has become a feasible way to provide users with personalized travel route planning services through intelligent travel service robots. Therefore, how to enable intelligent travel service robots to provide personalized travel route planning services for tourists has become an urgent problem to be solved.

2. The Framework

The user’s personalized travel route planning can be divided into two parts, one part is the personalized travel route planning for tourists who visit many times, the other part is the personalized travel route planning for new tourists. The number of new visitors refers to the first time a visitor has visited a city. Therefore, this paper needs to solve the following problems: (1) How to obtain the personalized travel preferences of new and old users. (2) How to generate accurate travel routes; (3) The selected tour routes should not only meet tourists’ personalized travel preferences, but also meet users’ various constraints. In view of these three problems, this paper proposes propose a relatively comprehensive solution, which can quickly, efficiently and comprehensively plan personalized travel routes for tourists and improve their travel experience. The overall process is shown in Figure 1.

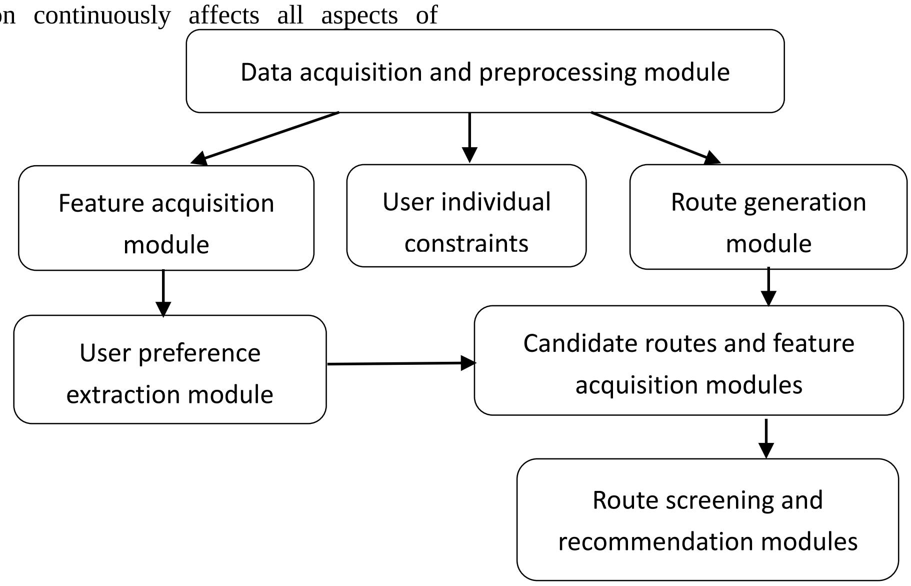

Figure 1. Overall process.

The whole process consists of data acquisition and preprocessing module, feature acquisition module, route generation module, candidate route and feature acquisition module, user preference extraction module, route selection and recommendation module. The whole personalized travel route recommendation method will be explained in detail below.

3. Feature Extraction

Sequence is mapped to a fixed dimension vector space, then levels through a sequence of vector with the context of vector addition and the average final vector to predict the next type information, resulting in the sequence of magnitude of vector can be thought of as the sequence, according to the characteristics of the contents of a whole call it based on the characteristic vector sequence information. The feature vector contains the category information of each attraction. Thus, the unified information representation of the scenic spots is obtained. For the acquisition of scenic spot features based on type information, it is necessary to learn the feature information from the previous statistics of scenic spot type specific gravity table. This paper uses the improved doc2vec method to express its type information in the form of vectors.

The classification sequence of scenic spots is: landscape, landscape, landscape, humanities. Each attraction is composed of a sequence of attractions, and finally a text data composed of a sequence of categories is obtained. However, the feature information of a scenic spot contains not only category information, but also other information. The user’s historical tour route is a sequence of scenic spots generated by different scenic spots in order of time. Due to the principle of the tourists’ choice, the geographical location information of the scenic spots is reflected in the tourist routes. In order to extract the feature information of scenic spots based on routes, this paper borrows the method of word2vec and input all routes into the model as text. The feature information of scenic spots in the routes is learned through the context of each scenic spot, namely the surrounding scenic spots, and is represented as a distributed vector. The context of each attraction can also be used to predict the process of the current attraction, reflecting the correlation between the attractions and the geographical location of the attraction. The overall training framework is shown in Figure 2.

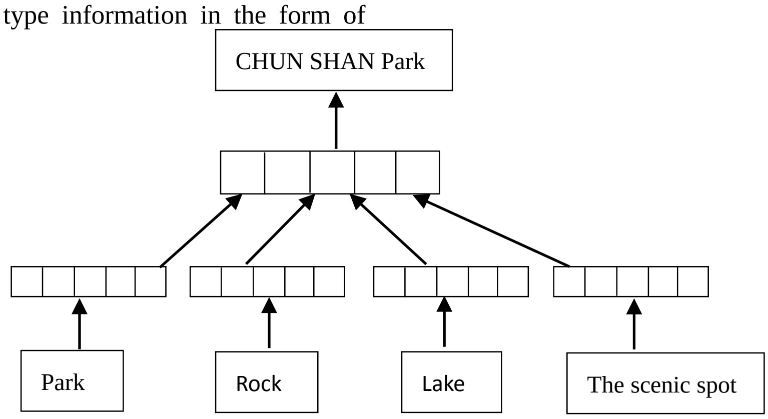

Figure 2. Learning model of scenic spot features.

Through the feature vectors of the scenic spots and the feature vectors of scenic spots based on tourist routes, the sum of them is averaged for the same scenic spots, and finally the feature representation vector of scenic spots based on tourist routes is obtained. Before explaining the extraction of tourists’ personalized preference, several definitions are introduced.

Definition 1 (sequence of tourist attractions) : the sequence of tourist attractions visited by tourist u in city A, which only contains the tourist attractions visited by tourists, and the sequence obtained by sorting them according to the sequence of tourist attractions. Such as < u, $[p_1, p_2, p_3, \ldots, p_n]$ >, the p_i I said visitors u for scenic spot.

Definition 2 (tourist route): On the basis of tourist attraction sequence, the travel time of the attraction and the traffic time between the two attractions are added. Such as $[P_1, T_1]$, T_{1-2}, $[P_2, T_2]$, T_{2-3}, $[P_3, T_3]$ … … , $T_{n\text{-}n+1}$, T_3, $[P_{n+1}, T_{n+1}]$ different tourists visit attractions sequence is not the same, so the arrangement of different attractions combination the personalized features of tourists.

Therefore, the personalized preference of each tourist can be obtained by using the tourist attraction sequence mining. This paper proposes a personalized tourism preference representation model based on the characteristics of tourist attractions. In other words, a dictionary composed of scenic spot and its vector is obtained by combining the feature representation vector of scenic spot with various features obtained in the previous section. Then, according to all attractions in each tourist attraction sequence, it is input into the information fusion layer. The main purpose of this layer is to obtain tourists’ personalized tourism preference information by combining the feature representation vector of the multi-source information with the tourists’ score of each attraction.

According to all scenic spots in each tourist attraction sequence, it is input into the information fusion layer. The main purpose of this layer is to obtain tourists’ personalized tourism preference information by integrating the feature representation vector after the multi-source information fusion with tourists’ score of each scenic spot.

The main method of this layer is similar to the attentional method, which represents the tourists' preference right for each attraction through the tourists' score. The specific calculation expression is as follows:

$$\mathrm{V}_{preson} = \left(\frac{f_1}{5} \cdot V_{p1} + \frac{f_2}{5} \cdot V_{p2} + \frac{f_3}{5} \cdot V_{p3} + \cdots + \frac{f_n}{5} \cdot V_{pn}\right)/n \quad (1)$$

where f_1 is tourists to scenic spots p_1 score, score full marks are five points; v_{pi} represents the fusion of multiple information p_i said the characteristics of a vector. After adding them up and averaging, we get the representation vector of personalized preference characteristics of users with a variety of information.

Finally, the information of each tourist's tourism preference will be obtained. When the data volume is enough to cover the whole China, the grasp of tourists' personalized preference will be clearer and more comprehensive. This personalized preference is a huge help in providing tourists with the services of a new city.

4. Generate and Recommend Routes

The problem of tour route recommendation is to provide tourists with a route that conforms to the order of public visit, and at the same time, this route can meet the tourists' personalized travel preference. For the generation of candidate routes, an improved PrefixSpan algorithm is adopted, which not only considers the sequence of attractions, but also considers the traffic time between the two attractions and the play time in the attractions. For all tourists in a city tour route consisting of a route library, which reflects the geographical location of each attractions between the information, also avoid the repeatability of tourist attractions. Therefore, we adopt the method of frequent pattern mining to generate candidate routes by mining frequent sequences in the route library. This method is described in detail below. The steps of the whole frequent pattern mining are as follows:

Step 1. Judge whether the length of the currently mined frequent tourist route is 0, that is, judge whether the method was called for the first time. Jump to step 2 if 0, or jump to step 5.

Step 2. Prefix all frequent-visit data points that meet the minimum support, build the corresponding projected travel route data set based on the travel route data set corresponding to the travel time of tourists and the categories of tourists.

Step 3. Add all frequent play data points that meet minimum support to the current frequent travel route.

Step 4. Use the new frequent visit data point and its corresponding projection data set as input parameters, and call this method recursively.

Step 5. Create an empty list of "attractions to visit -- journey time" relationships.

Step 6. Process each projection sequence in the projection dataset to build the elements in the relational table. For each of the projection sequence, scanning sequence of each subsequent behavior model, at the same time, the last spot in calculating the current frequent route data point and the distance between the various attractions behind time, journey time as superscript, subsequent spots data points as a subscript, to save all subsequent spots data points support technology to relational tables in the corresponding element.

Step 7. Scan all elements in the relational table, and compare whether the element value is greater than the set minimum support, that is, judge whether the table elements are frequent. Step 8 and step 9 are executed if frequent elements exist, or the execution ends.

Step 8. Add the travel time corresponding to the frequent table elements and the subsequent frequent visit data points to the tail of the current frequent travel route, generate a new frequent visit route, and save it to the frequent travel route database.

Step 9. Construct new projection data set with new frequent travel routes as prefix sequence; and with the new frequent travel routes and the corresponding projection data set as input, recursively call this method.

In order to ensure that tourists can visit enough scenic spots, this paper selects the routes with more scenic spots than the average of the total generated routes as the final candidate routes. Through the route generation module, the candidate tourist routes that meet the constraints of tourists are preliminarily obtained. In order to be able to match with the personalized preferences of tourists, the feature preferences of the generated routes need to be obtained. It is similar to obtaining tourists' personalized tourism preference, because the feature representation of each scenic spot after the fusion of multi-source information is known, the feature representation vector of each route can be obtained by adding up the feature representation vector of all scenic spots in the generated route and calculating the average. The specific calculation method is as follows:

$$V_{seq} = \left(V_{sp1} + V_{sp2} + V_{sp3} + \cdots + V_{spm}\right)/n \quad (2)$$

where v_{spi} is generated by the travel route in the attractions spi feature vector, n is the number of scenic spots in the generated tour route. Vector v_{preson} is Tourists travel preferences, Feature vector v_{seq} is generate candidate route, by the following formula to calculate the matching degree of the two m.

$$\mathrm{m} = v_{person}^{T} \cdot v_{seq} \quad (3)$$

By matching all the candidate travel routes with the tourists' personalized travel preference, the matching value of the same number of candidate travel routes will be obtained. Then, the value of all the matching degrees will be compared, and the tour route with the highest matching value will be selected for recommendation. Such recommended routes can meet tourists' various constraints, and can also meet tourists' personalized tourism preferences, so as to improve tourists' travel experience. Finally, the model is applied to the intelligent travel service robot, so that the robot has the function of personalized travel route recommendation.

5. Conclusions

This paper proposes a travel route recommendation algorithm which can be widely used and applies it to the intelligent travel service robot. In this method, the multi-source information of scenic spots is represented by the fusion of tourists' historical play trajectories and the category information of scenic spots, so as to obtain the distributed representation vector of scenic spots. According to the distributed representation vector of scenic spots, the preference feature representation of each user is obtained by combining the method of the attention-like mechanism. Secondly, the candidate routes are generated according to the improved PrefixSpan algorithm, and the generated routes are screened according to the different constraints of tourists and the personalized preferences of tourists, so as to obtain the optimal route. This is the recommended route for tourists. The methods in this chapter can not only solve the problem of traditional tourist route recommendation, but also be applicable to tourists coming to a new city, which can improve tourists' travel experience.

In the future, more information will be integrated to understand the characteristics of tourists' preferences, such as tourists' review data, photos, punch card data, etc., to enrich tourists' information, so as to better understand tourists' personalized preferences. It will also be considered to obtain more real information about attractions by increasing the behavioral data of most tourists in the attractions, so as to improve the recommendation accuracy. Finally, the algorithm is embedded into the intelligent travel service robot, enabling it to realize personalized travel route recommendation service, saving users' energy and time, and improving their travel experience.

References

[1] Cao, Y.; Fukunaga, A.S.; Kahng, A. Cooperative mobile robotics: Antecedents and directions, 2014.

[2] Autonomous robots, pp. 7-27.

[3] Jung, T.; Lim, J.; Bae, H.; Lee, K.K.; Joe, H.M., and Oh, J. H. Development of the humanoid disaster response platform DRC-HUBO+. IEEE Trans. Robot., 2018, pp. 1-17.

[4] Long, X.; Jin, L.; Joshi, J. Exploring trajectory-driven local geographic topics in foursquare. International Workshop on Location-Based Social Networks, 2012, pp. 927-934.

[5] Farrahi, K.; Gatica-Perez, D. Extracting Mobile Behavioral Patterns with the Distant N-Gram Topic Model. International Symposium on Wearable Computers. IEEE Computer Society, 2012, pp. 1-8.

[6] Yin, P.; Ye, M.; Lee, W.C., et al. Mining GPS Data for Trajectory Recommendation. Advances in Knowledge Discovery and Data Mining, 2014, pp. 50-61.

[7] Murphy, H.C.; Chen, M.M.; Cossutta, M. An investigation of multiple devices and information sources used in the hotel booking process. Tourism Management, 2016, pp. 44-51.