

## Collaborative Filtering Recommendation Algorithm Based on label of tourist spots and User Preference

Ya Zhou<sup>1</sup>, Cailin Hu<sup>1, a</sup>, Han Xiong<sup>1</sup>, Ling Li and Xiafei Wei<sup>1</sup>

<sup>1</sup>Gugangxi Key Laboratory of Software

Guilin University of Electronic Technology, Guilin 541004, China.

<sup>a</sup>963127935@qq.com

**Abstract:** The traditional collaborative filtering recommendation algorithm relies on the user's scoring relation to the item. However, user's behavior data in the field of tourism industry is sparsely, and the interaction between the data are few. These problems lead to the traditional algorithms are lacking of ability to acquire the users' preference, and influence the recommendation quality of tourist spots. In this paper, a collaborative filtering recommendation algorithm based on tourist spots labels and user preferences were proposed. By using the label information of the tourist spots, to extract visitors' interest factors of spots and preferences weights, used the adaptive algorithm of neural network to optimize the users' preference feature vector, and computed the similarity between users to get the recommended result. The results show that compared with the traditional recommendation method, the method of this paper can ameliorate the accuracy of user similarity relationship and has a great improvement in the recommendation quality of tourist spots.

**Keywords:** Recommendation, Tourist Spots, User Preference, labels, Similarity Relationship.

### 1. Introduction

With the rapidly development of Internet technology, the information is exploding, there has been "information overload[1]" phenomenon, especially in field of the tourism industry, a wide range of tourism information staggered, users prone to "choice of obstacles[2]" situation. Personalized recommendation technology can select and analysis the massive complex information, to extract user preferences for tourism users and recommend the appropriate tourism spot.

User preferences reflect the user's interest in specific tourism spots, therefore, improving the accuracy of user preferences to improve the recommendation quality. But the tourism industry is not like the traditional electronic commerce industry, due to the particularity of the scene, there are two factors that reduce the accuracy of user preferences. Firstly, the traveler's effective preference information is difficult to obtain, such as a user will not often go to the same tourism spots or the number of spots are few, Secondly, the tourist data is sparseness, the interaction between the data are few, it is difficult to obtain the relationship between users, and the same tourist attraction between users is very little, and the number of visitors of spots in a year are also few, the hot spots of type is not enough, causing the spots unable to be recommended.

In order to solve the above two problems, FENZA[3] using the FCM method to fuzzy clustering users and spots information, then coupled with the users' context matching information to generate association rules, According to user history information take users into classification, As user's personal preference about the tourist spots to recommend by different categories. HWANG[4] conveying tourist users' preference by a  $h$  (number of points) dimension of the score vector, and using the collaborative filtering algorithm to predict the spots' score, taking into the distance of the users and spots to recommend. Massa[5] proposing an trust propagation algorithm, Every item have many attributes, combining these item attributes with trust propagation algorithm can improve the accuracy of the users' preference, and the algorithm could construct a variety of improved similarity calculation method. Wang Xianfei[6] proposed a knowledge-based tourism recommendation algorithm, using the preferences and current needs of users and tourism information to recommended tourism products for users, tourism information including tourism knowledge, tourist information, recommended rules

and personalized rules. B Yu[7] Considering the advantages and disadvantages of personalized travel and classic travel, and taking the the interaction between user preferences and classic attractions into the recommendation algorithm to propose a CIAP method which based on user preference and spots popularity.the algorithm extract the user preference matrix from the user's GPS trajectory information to build the model of user preference,based on the model to define the user similarity function and spots popularity function,According assign weights to merge them to compute best recommendation results;Chaolun Xia[8] considering the influence to weights of score by time, and proposed a similarity compute method based on time attenuation. Yueping Wu [9]proposed a similarity compute method combine the Jaccard similarity coefficient of items attributes and scoring similarity,the method can improve the accuracy of calculation results.

It's partly the above methods can improve the recommendation quality.However,the existing recommendation algorithm focuses on the similarity between the user and the item but ignores the dynamic change of the user's interest, without improving the accuracy of the user's preference.In this paper, proposing a collaborative filtering recommendation algorithm based on tourist spots labels and user preferences. By using the label information of the tourist spots,to extract the interest factors of the visitors 'spots and the users' preferences weights, and using the adaptive algorithm of neural network to optimize the user preference feature vector, and computing the similarity between users to get the recommended result. The results show that compared with the traditional recommendation method, the method of this paper can ameliorate the accuracy of user similarity relationship and has a great improvement in the recommendation quality of tourist spots.

## 2. Collaborative Filtering Recommendation Algorithm Based on label of tourist spots and User Preference

### 2.1 Algorithm Framework.

In the traditional recommendation system, users and tourist spots are important entities, by analyzing the relationship between the two entities to express the user's preferences.Next, we will introduce the definition of matrix and vector in algorithm.The sets of users:  $U = \{u_1, u_2, u_3, \dots, u_i, \dots, u_n\}$ , tourist spots sets:  $S = \{s_1, s_2, s_3, \dots, s_j, \dots, s_m\}$ , the sets of spots labels:  $L = \{l_1, l_2, l_3, \dots, l_k, \dots, l_p\}$ . And we need the tourist spots-labels matrix:  $L_{s_j} = \{l_1, l_2, l_3, \dots, l_x\}$ , It represents the labels information of each tourist spots, and  $L_{s_j} \subset L$ .

The overall flow chart of the framework model proposed in this paper is shown in Fig. 1.

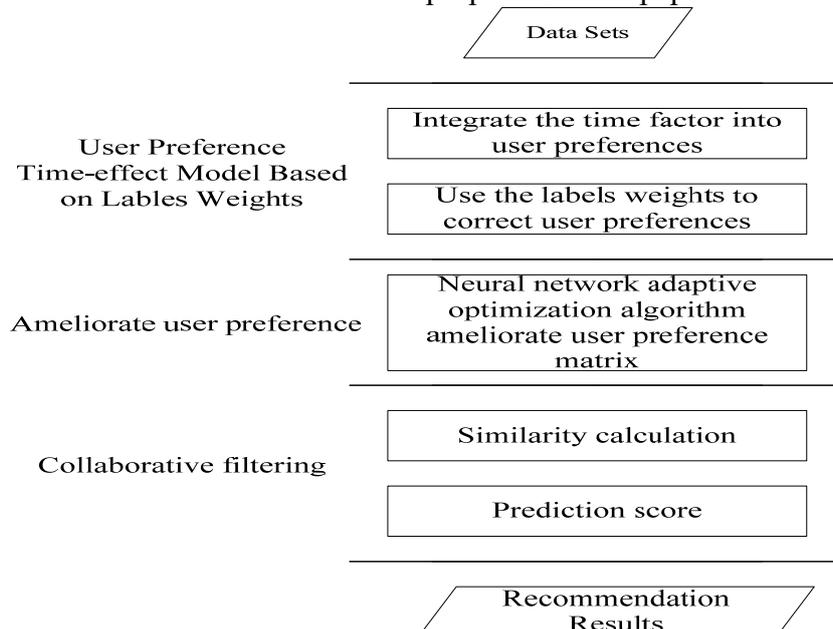


Fig. 1 The overall flow chart of improved recommendation algorithm.

After the original data processing, we got the users-tourist spots scoring matrix, the matrix include the active time and tourist spots labels, the score is the initial value of user preferences. Firstly, we integrating the travel time factor into the user preferences, and we could get the different user preferences at different travel time, Then taking the tourist spots labels into the user preference instead of the tourist spots; Secondly, we use the neural network adaptive optimization algorithm to ameliorate the user preference matrix; Next step is to calculate the similarity user and prediction score, and we could get the recommendation results.

## 2.2 User Preference Time-effect Model Based on label weights.

User preferences will dynamically change in the recommended time, and it will affect the recommendation quality, and we will analyze the spots labels information to enhance the user preference, Therefore, Taking the users' active time and spots label into user preference to build the user preference time-effect model on label weights.

### (1) Integrate the time factor into user preference

User tourist spots scoring reflect the user's preference for the tourist spots. Then, how to calculate the result while there are multiple scoring information for the same spot, and the score is different at different time, Therefore, we want to find a way to count the total score which is useful for user preference. For example, user A scored tourist spot S at time  $t_1$  and  $t_2$  ( $t_2 > t_1$ ), and current time is  $t$ , we consider that the shorter the time and the greater the impact. And every user have a preference period, if the differential value of scoring time and current time is smaller than preference period, the scoring has positive effects, on the contrary, the scoring has negative effects. Then, The formula for define effects of the scoring time as follow:

$$P_s(t) = \sum_{i=1}^m \alpha_i e^{-\Delta(t-t_i)} \quad (1)$$

$$\alpha_i = \frac{T_p - (t - t_i)}{|T_p - (t - t_i)|} \quad (2)$$

$P_s(t)$  is preference the scoring at time  $t$ ,  $\Delta$  is the declining factor to adjust the scoring at different time,  $\alpha_i$  is represent the control factor, and  $T_p$  is represent the preference period,  $t$  is the current time,  $t_i$  is the scoring time. Combine the time-effect model with user-tourist spot scoring matrix to calculate the user preference matrix:  $R_{us}' = R_{us} * P_s(t)$ ,  $R_{us}'$  is represent users' preference matrix.

### (2) Use the spot labels weights to correct user preference

Tourist spot labels can reflect the style of tourist spots, If user give a high score to tourist spot, we consider the user like the style of spots, It means that the labels is also the important role to reflect the user preference. Then, we want to transfer the scoring of tourist spots to labels, using the user-tourist spots scoring matrix  $R_{us}'$  and the tourist spot-labels matrix  $T_s$  to calculate the user-tourist spot labels scoring matrix  $R_L$ , and integrate the time factor into the user-tourist spot labels scoring matrix:

$$R_L = R_{us}' \bullet T_s.$$

Many tourist spot have different labels, every label have different weights, According to analysis the important proportion of labels, we can get the  $imp(L_k)$ :

$$imp(L_k) = \log \frac{\sum n_{l_k}}{n_{l_k}} \quad (3)$$

As show in formula (3),  $n_{l_k}$  is represent the number of times using label  $l_k$ , Use the important proportion to calculate the labels weights. The formula for calculating the labels weights as follow:

$$O_{U_{l_k}} = \frac{n_{ul_k}}{N} impor \tan(L_k) \quad (4)$$

$n_{ul_k}$  is represent the number of user scoring label  $l_k$ , and  $N$  is total number of user,  $O_{U_{l_k}}$  is the weights of label  $l_k$ . Then we can get the user preference matrix with active time and label weights:  
 $R'_L = R_L * \overrightarrow{O_L}^T$ , and  $R'_L$  is the result matrix of user preference time-effect model based on label weights.

**2.3 Neural Network Optimization Process.**

We will use the neural network adaptive optimization algorithm to ameliorate the user preference matrix. And the formula is  $UL: \{UL | R \rightarrow UL \bullet R'_L + b\}$ ,  $UL$  is represent the user preference matrix,  $R'_L$  is represent every user score for every label, and the  $R$  is represent the total tourist spots labels scoring matrix,  $b$  is the threshold matrix of the network nodes, the weigh of network node is  $\Delta w_i(t+1) = -\eta \partial_i * x_i + \alpha \Delta w_i(t)$ , Compare the result with  $R$  to adjust the weights and threshold of network node to calculate the user preference matrix.

The flow chart of neural network optimization process as follows:

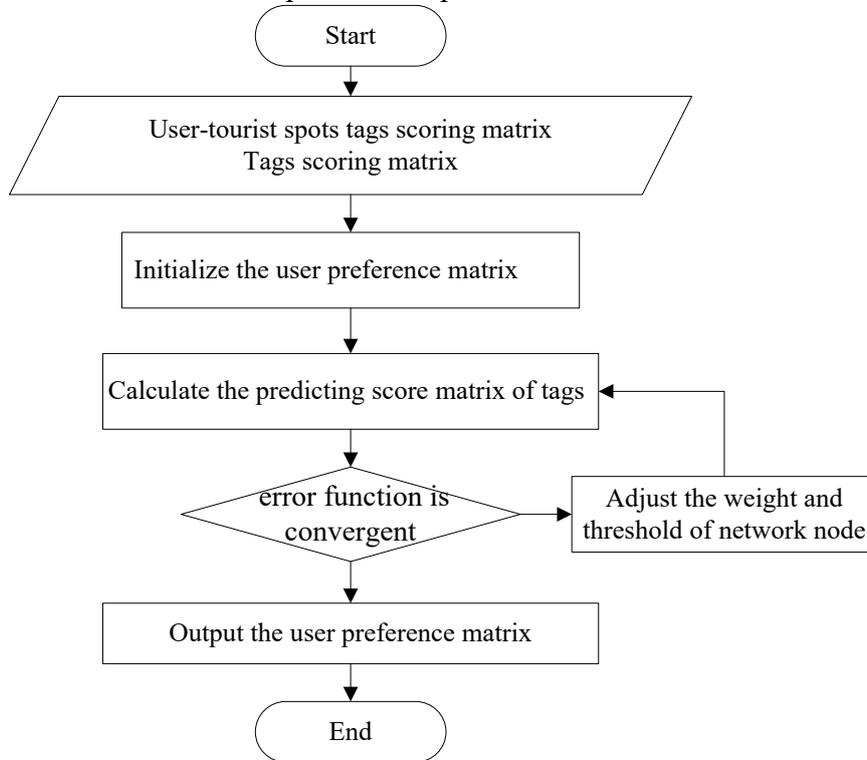


Fig. 2 The training process of neural network algorithm.

**2.4 Steps of Algorithm.**

According to the user preference time-effect model based on labels weights and neural network optimization process, we can get accurate user preference, next step is to calculate similarity of users and predict score.

Then we choose the modified cosine similarity algorithm as:

$$sim(u, v) = \frac{\sum_{i=1}^n (r_{ui} - \bar{r}_i) \bullet (r_{vi} - \bar{r}_i)}{\sqrt{\sum_{i=1}^n (r_{ui} - \bar{r}_i)^2} \sqrt{\sum_{i=1}^n (r_{vi} - \bar{r}_i)^2}} \tag{5}$$

$$pred(u,i) = \frac{\sum_{v \in U} sim(u,v) \cdot (r_{vi} - \bar{r}_v)}{\sum_{v \in U} |sim(u,v)|} \quad (6)$$

In the formula,  $r_{ui}$  and  $r_{vi}$  represents the user  $u$  and  $v$  predictive score for tourist spots  $i$ ,  $\bar{r}_i$  represents the average score of tourist spots  $i$ ,  $sim(u,v)$  represents the similarity between user  $u$  and user  $v$ , and  $n$  represents the number of tourist spots in the data set.

Suppose users-tourist spots matrix is  $user\_spot(u,S)$ , and  $S$  represents the list of tourist spots. tourist spot-labels matrix is  $spot\_lables(s,L)$ , and  $L$  is represent the labels of spot.  $user\_spots(u,l)$  is represent the spots frequency of user. user-label thermal matrix is  $user\_label(u,l)$ , it represent the frequency of user  $u$  using the label  $l$ . users-tourist spots scoring matrix is  $user\_spots\_rank(u,s,t)$ , it represent user  $u$  scored for spot  $s$  at time  $t$ . The flow of algorithm as Table 1.

Table 1 Steps of Algorithm

<b>Input:</b>	users- tourist spots scoring matrix, tourist spots - label matrix, users - tourist spots matrix
<b>Output:</b>	Tourist spots Matrix(recommendation results)
<b>Process:</b>	<ol style="list-style-type: none"> <li>1. Calculate <math>user\_label(u,l)</math> through <math>spot\_label(s,L)</math> and <math>user\_spot(u,S)</math>;</li> <li>2. Calculate the ranking of every spots with <math>user\_spots\_rank(u,s,t)</math>, and combine the <math>spot\_lables(s,L)</math> to calculate the user-tourist spots labels scoring matrix <math>R_L</math>;</li> <li>3. Correct <math>R_L</math> with the weights of labels <math>O_l</math> and redefine as <math>R'_L</math>, Calculate the labels scoring matrix <math>R</math>;</li> <li>4. Use the neural network adaptive algorithm to initialize the user preference matrix <math>UL</math> with user-tourist spots labels scoring matrix <math>R'_L</math> and the labels scoring matrix <math>R</math>;</li> <li>5. Calculate the predicting score matrix of labels <math>R\_1</math>;</li> <li>6. Use the error function to judgment the matrix <math>R\_1</math> and <math>R</math>, and adjust the weights and threshold of network node to calculate the user preference matrix;</li> <li>7. Iterate steps 5 and steps 6 until the error function is convergent;</li> <li>8. Calculate the similar neighbors and predict score for every user.</li> <li>9. Rank the tourist spots with the predicting score, and take the top-k spots as the final recommendation results.</li> </ol>

### 3. Experimental Results and Analysis

#### 3.1 Experimental Data Set.

In order to evaluate the property of Collaborative Filtering Recommendation Algorithm Based on label of tourist spots and User Preference, We uses real travel data set, and the data set is the real data climb from the tourism website(<http://www.mafengwo.cn/>). It contains the score records of tourist users, basic data of tourist spots and labels, This data contains 1228 tourist users, 130 tourist spots, 113 labels and 24481 reviews, 60898 score records. And the data sparsity degree is bigger than 9%.

In the paper, we setting the test of experimental method according Ten cross validation, namely, we divided the data sets into 10 parts in random, which 8 parts be selected as training sets in every time. And the remind part as test set (formally, 80% training sets and 20% testing sets). The assess result will be come out as five average final evaluation results.

### 3.2 Experimental Content and Evaluation Criteria.

The experiment consists of three parts:(1)Analysis the user preference decline factor and user preference period;(2)Comparison the user preferences calculation of based on tourist spots and labels;(3)Contrast with traditional methods to analysis the result of RMSE.

In this paper we use the mean absolute error(MAE) and root-mean-square error(RMSE) to judge the accuracy of the results.Them namely:

$$MAE = \frac{\sum_{r_{ui} \in T}^n |r_{ui} - r'_{ui}|}{n} \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{r_{ui} \in T} (r_{ui} - r'_{ui})^2} \tag{8}$$

In the formula,  $r_{ui}$  represents the user's predictive score for tourist spots  $i$ ,  $r'_{ui}$  represents the user's true score for the tourist spots  $i$ ,  $T$  is the test data set, and  $n$  represents the number of tourist spots in the data set.

### 3.3 Results and Analysis.

#### (1)Analysis of User Preference Decline Factor and User Preference period

First, In order to guarantee the quality of the experiment,we will discuss some of the parameters in the test.They are the user preference decline factor  $\Delta$  and user preference period  $T_p$ .

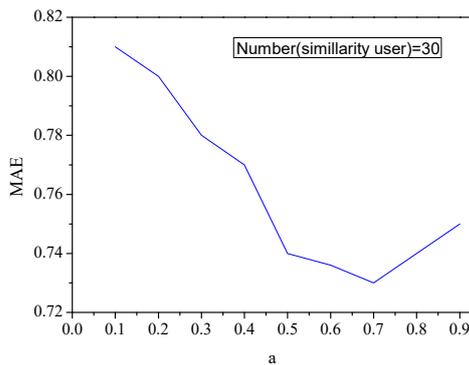


Fig. 3 User preference declining factor

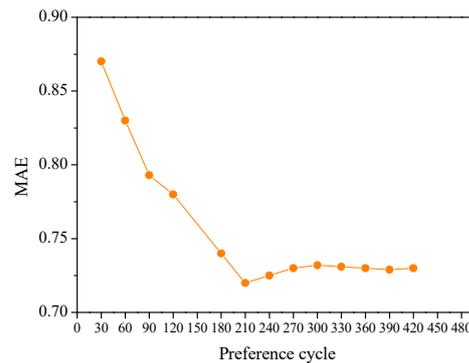


Fig. 4 User preference period

In the user preference model, changing the user preference decline factor can affect the range of user preferences, also affect the recommendation quality.The number of users' similarity user is 30,And analysis the effect of user preference decline factor on MAE value.As shown in Fig 3,When  $\Delta$  range from 0.1 to 0.9, the MAE is also changing,And when  $\Delta=0.7$ ,the MAE is minimal,Therefore, in the case of no changing in the data set,we will set the value of user preference decline factor( $\Delta$ ) is 0.7 in others experimental.

The user preference period( $T_p$ ) is the key parameter to control the gaining of user preference. Therefore, it is necessary to get the most suitable parameter through experiment analysis so that the user preference is the most favorable.As shown in Fig 4, the MAE value is also changing when the user preference period changes, the user preference period is range from 30 to 210, The MAE decrease progressively,when the value of user preference period is 210, the MAE is the lowest.and when the value range from 210 to 270,the user preference begin to weaken,and the value is bigger than 270, the user score data has no value.Then,the best value of user preference period is 210.

#### (2)Comparison the user preferences calculation of based on tourist spots and labels

In this paper, we use the user users-spot labels scoring to define the user's preference, and compare it with the traditional method which use user users-tourist spots scoring to define the user's preference.

Compare the performance of the two methods by calculating the number of similar users. As shown in Fig 5, It is difficult to find the similar relationship between users with users-tourist spots scoring matrix, and 80% of the users do not have the common score; On the contrary, after extract the multi-dimensional nature of the tourist spots, use the users-tourist spots labels scoring matrix to define the user preference to calculate the similarity user can greatly improve the practicality of the algorithm.

Next step is to compare the recommendation quality on MAE. As shown in Fig 6. As the value of user similarity is increasing, the MAE is decreasing, and when the the value of user similarity less than 0.5, the performance of the two methods is basically the same, while the value is greater than 0.5, Similarity computational methods with the tourist spots labels is better than the tourist spots.

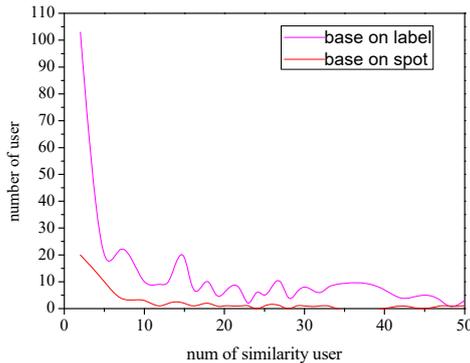


Fig. 5 Compare the number of similar users

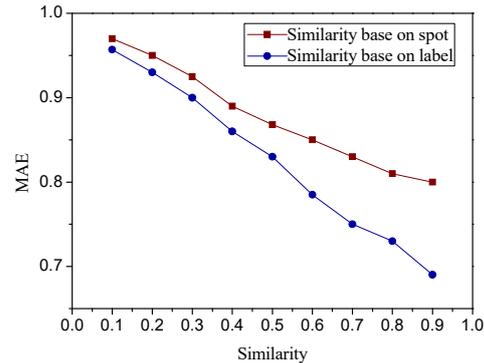


Fig. 6 Compare the MAE

(3) Contrast with traditional methods to analysis the result of RMSE

Next, we will contrast with traditional methods to analysis the result of RMSE. the traditional methods include ICF[10], UCF[11], and P-CF[12] which could adapt to changing of user interest.

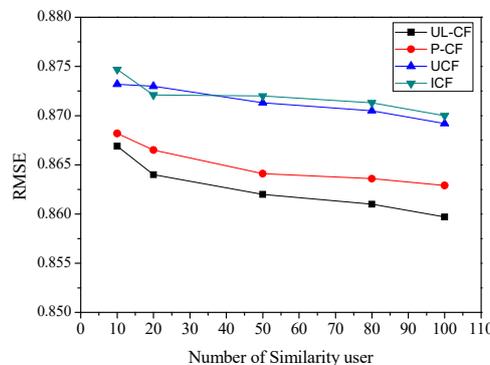


Fig. 7 Performance comparisons on RMSE

We use the root-mean-square error (RMSE) to judge the accuracy of the results. As shown in Fig 7, As the number of similar neighbors increases, the performance of ICF is not good, and the value of RMSE is still great, so we can find that collaborative filtering based on the user is more effective than the collaborative filtering algorithm based on the tourist spots, and takes user preference into collaborative filtering is better than ICF; Contrast with UCF, the algorithm of us made the RMSE much small, It is a better way to define user preference with active time factor and tourist spots labels; The last, Compare with P-CF, The gap between two method in RMSE are not very large, but we still have advantages, the neural network algorithm plays a important role. In conclusion, the collaborative filtering recommendation algorithm based on label of tourist spots and user preference can get the better recommendation results.

4. Conclusion

User similarity calculation occupies a very important position in the recommendation algorithm, Due to user's behavior data in the filed of tourism industry is sparsely, and the interaction between the

data are few, the method of similarity calculation based on user or item is not well adapt to the field of tourism recommendation. And the spots label information is not being fully utilized in the process of recommendation. This paper proposed a collaborative filtering recommendation algorithm based on label of tourist spots and user preference to solve the sparseness problem between the user and the spot, to extract the interest factors and preferences weights of users, and use the adaptive algorithm of neural network to optimize the user preference feature vector to get better recommendation result. The results show that compared with the traditional recommendation method, the method of this paper can ameliorate the accuracy of user similarity relationship and has a great improvement in the recommendation quality of tourist spots. For the later part of the study, We will be aim to study the user's internal relevance, to further improve the accuracy of user preferences, making the recommended algorithm more suitable for the tourism industry.

### **Acknowledgments**

This work is supported by Guangxi Key Laboratory of Trusted Software of Guilin University of Electronic Technology. And this work is supported by the National Natural Science Foundation of China (No.61662015), key projects of science and technology development of Guangxi science and technology department (No.1598089), NSFC-Guangdong joint fund key project (No.U1501252).

### **References**

- [1]. Pilli L E, Mazzon J A. Information overload, choice deferral, and moderating role of need for cognition: Empirical evidence[J]. *Revista De Administracao Publica*, 2016, 51(1):36-55.
- [2]. Lin Y C, Chen T, Wang L C. Integer nonlinear programming and optimized weighted-average approach for mobile hotel recommendation by considering travelers' unknown preferences[J]. *Operational Research*, 2016:1-19.
- [3]. FENZA G, FISCHETTI E, FURNO D, et al. A hybrid context aware system for tourist guidance based on collaborative filtering[A]. FENZA G. *The IEEE International Conference on Puzzy Systems*[C]. New Jersey: IEEE Press, 2004. 259-262.
- [4]. HWANG S, YAN W. On-tour attraction recommendation in a mobile environment. *IEEE International Conference on Pervasive Computing and Communications*. 2012.
- [5]. P. Massa, P. Avesani. Trust-aware Collaborative Filtering for Recommended Systems [C], *Proceedings of International Conference on Cooperative Information Systems*, 2004: 492-508.
- [6]. Xian-fei WANG, Mei CHEN, Xiao-tian LI. Research and Design of Constraint-Based Travel Recommendation System[J]. *Computer Technology and Development*, 2012(22):141-145.
- [7]. B Yu, F Liu, T Li. Recommendation of tourist attractions based on user preferences and attractions popularity. *Journal of Computational Information Systems*, 2014, 10(20):8661-8668
- [8]. Chaolun Xia, Xiaohong Jiang, Sen Liu, et al. Dynamic Item-Based Recommendation Algorithm with Time Decay [J]. *Sixth International Conference on. Natural Computation*, 2010:242-247.
- [9]. Yueping Wu, Jianguo Zheng. A Collaborative Filtering Recommendation Algorithm Based on Improved Similarity Measure Method[J]. *IEEE International Conference on Progress in Informatics and Computing*, 2010:246-249.
- [10]. Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques[J]. *Advances in artificial intelligence*, 2009, 2009: 4.
- [11]. Bokde D K, Girase S, Mukhopadhyay D. An Item-Based Collaborative Filtering using Dimensionality Reduction Techniques on Mahout Framework[J]. *Computer Science*, 2015, 134(3):561-576.
- [12]. LI Kechao, LIANG Zhengyou. Exponential forgetting collaborative filtering recommendation algorithm incorporated with user interest change. *Computer Engineering and Applications*, 2011, 47 (13) : 154-156.