Personalized Recommendation Method for Tourist Attractions Based on User Information Mixed Filtering

Hongshen Liu(✉) and Honghong Chen

Heilongjiang Polytechnic, Haerbin 150080, China
okjhn12300@126.com

Abstract. In order to improve the effectiveness of tourist attraction recommendation, this article proposes a personalized recommendation method for tourist attractions based on mixed filtering of tourist information. This method includes two parts: the construction of a tourist attraction information database and the personalized recommendation method for tourist attractions. Among them, the construction of the tourist attraction information database includes three steps: mining tourist attraction information based on association rules, updating tourist attraction data, and constructing a tourist attraction information feature vocabulary based on topic similarity clustering. The personalized recommendation methods for tourist attractions mainly include two aspects: describing the semantic association of tourist attraction information and selecting the optimal personalized recommendation path for tourist attraction information. The experimental results show that the proposed method improves the accuracy and efficiency of personalized recommendation for tourist attractions.

Keywords: User Information · Mixed Filtration · Scenic Spot · Theme Similarity Clustering · Personalized Recommendations

1 Introduction

Nowadays, tourism has become a common way of leisure and entertainment for people. However, tourists usually need to spend a lot of time and effort planning their itinerary before traveling, and due to a lack of sufficient travel information, they often find it difficult to make the best choice [1]. Meanwhile, with the continuous growth of personalized user needs, traditional tourism recommendation systems are no longer able to meet the needs of tourists. In response to these issues, tourist attraction recommendation systems have emerged [2]. It can improve users' travel experience and satisfaction by analyzing their travel preferences and behavioral data, recommending scenic spots that match their interests and needs [3].

The reference [4] method utilizes convolutional neural networks (CNN) to extract text comment features for sentiment classification, and calculates similar user groups using Pearson similarity formula to achieve tourism recommendation. Reference [5]

L. Yun et al. (Eds.): ADHIP 2023, LNICST 549, pp. 166–178, 2024.
https://doi.org/10.1007/978-3-031-50549-2_12

proposed a multi-objective tourism route recommendation method that integrates user features and group intelligence. Firstly, obtain scenic spot information and corresponding group intelligence data through websites such as Ctrip, Wanglu Travel, Baidu Index, etc.; Secondly, combining user characteristics and group intelligence data, construct the comprehensive attractiveness of scenic spots to users with different characteristics and calculate the attractiveness index of tourism routes; Finally, define a multi-objective optimization function for tourism route recommendation and generate a route recommendation list using the multi-objective genetic algorithm NSGA2. Reference [6] Method By summing up the impact of 11 situational elements on scenic spot recommendation and discussing the difference in their impact degree, a tourism scenic spot recommendation model integrating situational awareness and random forest algorithm is proposed, and the situational elements are modeled as the characteristic attributes to be considered when the decision tree splits in random forest to achieve tourism scenic spot recommendation.

The personalized recommendation system for tourist attractions places more emphasis on user satisfaction with the recommendation results, and can flexibly and accurately recommend based on user behavior and interest information. Personalized recommendation of tourist attractions can not only provide better tourism choices for tourists, but also promote the development of the tourism industry. By accurately recommending scenic spots, tourists can improve their stay time and consumption level, and promote local economic development. Therefore, this paper proposes a personalized recommendation method for tourist attractions based on mixed filtering of user information. On the basis of constructing a tourist attraction information database, personalized recommendation of tourist attractions based on user information mixed filtering is achieved by describing the semantic association of tourist attraction information and selecting the optimal personalized recommendation path for tourist attraction information. In order to enhance the experience of tourists, promote the development of the tourism industry, and promote the management and development of tourist attraction information through this study.

2 Construction of Tourist Attraction Information Database

2.1 Mining Tourist Attraction Information Based on Association Rules

Using association rule mining algorithm [7–9] to mine tourist attraction information, based on the principle of association rule mining algorithm, strong association rules with the minimum support and minimum confidence are found in numerous tourist attraction information databases to achieve tourist attraction information mining.

The definition of HGD refers to the percentage of the thing containing the Information set of tourist attractions in the whole tourist attraction information database, recorded as $h(i)$, and DKL refers to the union of two Information set of tourist attractions in the whole transaction data. Suppose that in the Information set $I = (i_1, i_2, \cdots, i_n)$ of tourist attractions, there is a weight value for any tourist attraction information i. This weight value is used to measure the importance of tourist attraction information in the whole set. The greater the weight value, the more important the tourist attraction information is. On this basis, the tourist attraction information in the set is sorted according to the weight value, and a combination of ranking from large to small is obtained to form a linear ordered set.

Use z, x to represent the element in the Information set I of the tourist attraction. If $z < x$, it means that z is in front of x. If the weighted support of z is defined as $\mathrm{H}(z)$, then the minimum weighted support of the tourist attraction information is:

$$\mathrm{H}(z) = \frac{s - f(z)}{D} \tag{1}$$

In formula (1), D represents the number of tourist attraction information database data, s represents the weighted frequent tourist attraction Information set, and $f(\cdot)$ represents the calculation factor of weighted support.

According to the above formula, calculate the minimum weighted support of the tourist attraction information database. Based on this, calculate the data confidence using the association rule mining algorithm [10] as follows:

$$Z = \frac{k(z \cup x)}{n(x)\mathrm{H}(z)} \tag{2}$$

In formula (2), $k(z \cup x)$ represents the number of times two tourist attraction information simultaneously appears in the tourist attraction information database, and $n(x)$ represents the degree of data correlation.

According to the above calculation, find out the problem of frequent Information set of tourist attractions in the tourist attraction information database, determine the reliable relationship rules in the tourist attraction information database, so as to complete the information mining of tourist attractions, and determine the process of information mining of tourist attractions as shown in Fig. 1:

2.2 Update of Tourist Attraction Data

2.2.1 Differences in Tourist Attractions

Firstly, the evaluation of different tourist attractions by the same user is calculated by subtracting the scores to determine the degree of dissimilarity of tourist attractions, which is a measure of the dissimilarity of tourist attractions.

Set x as the evaluation set after recommending tourist attractions, i and j as any two tourist attractions, with the same user U rating them U_1 and U_2 respectively. For the currently selected tourist attractions i and j, set $S(x)$ as the user who overrated, N as the number of users, and the similarity between tourist attractions i and j is:

$$d = \frac{Z(U_1 - U_2)}{N \times S(x)} \tag{3}$$

According to formula (3), if the user's ratings for i and j are very close, the degree of difference between tourist attractions is very small, while the opposite is true. The dissimilarity calculation of tourist attractions is the result of analyzing from the user rating dimension. If analyzed from the tourist attraction rating dimension, the dissimilarity calculation of users will be obtained.

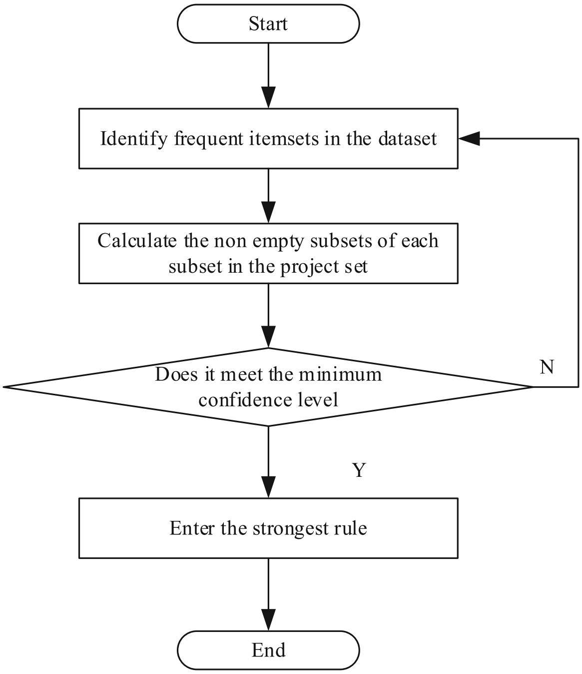


Fig. 1. Process of Mining Tourist Attraction Information

2.2.2 User Dissimilarity

Extract absolute values by subtracting the ratings of different users for the same tourist attraction, and obtain the degree of user dissimilarity. This calculation can be used to measure the degree of difference in user preferences.

Set X as the training set, A and B as any two users, and the same tourist attraction i is rated by users A and B respectively. User A has a rating of A_1, user B has a rating of B_1, and the tourist attraction that is overrated by A and B is $S'(x)$, V is the number of tourist attractions. The difference between user A and user B is:

$$d' = \frac{A_1 - B_1}{M \times S'(X)} \tag{4}$$

According to formula (4), if the ratings of users A and B in the same tourist attraction are very close, then the dissimilarity of users is small, while the dissimilarity is large.

2.2.3 Tourist Attraction Update Process

From the above calculation, it can be concluded that the dissimilarity matrix can perform local online updates on newly added data, and only one rating vector is used for each update, without the need for mutual operations with other vectors, thus saving computational time. The specific update method is as follows:

After calculating the tourist attraction i, several new user ratings are added. If n_1 is the new user of the tourist attraction, n_2 is the scoring set of old users, A_1 and B_1 are selected as any two users, and users A_1 and B_1 score tourist attraction i as a_1 and b_1, then the update result of the dissimilarity between A_1 and B_1 is:

$$d'' = d' + \frac{d}{|b_1 - a_1|} \tag{5}$$

In summary, because updating only requires processing one vector at a time and does not require the use of all rating data for updates, it has the characteristics of less workload and fast processing speed.

2.3 Constructing a Tourism Attraction Information Feature Thesaurus Based on Theme Similarity Clustering

Separate tourist attraction information from conventional information and design an algorithm that can filter such information. Using topic similarity clustering as the core algorithm [11, 12], design a model that can aggregate relevant information and filter irrelevant information. Based on this model, establish a feature vocabulary for tourist attraction information, and cluster based on the similarity between these vocabulary and tourist attraction information. In this process, it is necessary to first establish a suitable parameter interval using the existing vocabulary, which can directly affect the clustering effect of tourist attraction information and the accuracy of the feature vocabulary. Among them, the selection of clustering centers is the most cautious step, and the feature lexicon obtained through different clustering centers has completely different characteristics. At this point, vectors can be used as the basis for discriminating information similarity and as clustering tools for tourist attraction information, and the frequency of tourist attraction information occurrence can be treated as independent semantic units. Among them, the K value of topic similarity clustering is a very important calculation node, and the topics of different clustering centers are cross connected to calculate the information similarity in pre training through clustering algorithms. This article uses the method of extracting feature categories to establish a feature text space, and appropriately introduces the feature factor vocabulary of tourist attraction information. By increasing the weight of low-frequency words, a cross channel of word frequency is established, enabling the algorithm to quickly collect vocabulary of tourist attraction information with different weights. In the dataset of the feature vocabulary, the frequency calculation expression of a vocabulary can be obtained as follows:

$$\lambda = \frac{d'' \sum_{x=1}^{n} b_x}{\sum_{y=1}^{a} \sum_{x=1}^{b} b_{yx}} \tag{6}$$

In the formula, λ represents the characteristic frequency coefficient of a certain vocabulary; y represents the number of irrelevant information related vocabulary; b_x represents the number of times feature words appear in the category of tourist attraction information. The element feature similarity of each word frequency can be obtained through formula (6) as follows.

$$w = \frac{n(p+q)}{(p+v)(q+v)(u+v)} \times \lambda \tag{7}$$

In the formula, p represents the number of documents containing information about tourist attractions and belonging to irrelevant information; q represents the number of documents containing information about tourist attractions that are not irrelevant; u represents the number of documents that do not contain tourist attraction information and belong to irrelevant information; v represents the number of documents that do not contain tourist attraction information and do not belong to irrelevant information.

The larger the word frequency coefficient obtained, the higher the correlation between the word and tourist attraction information, and the more suitable it is to be included in this feature vocabulary; The smaller the word frequency coefficient obtained, the smaller the correlation between the word and tourist attraction information, and the less suitable it is to be included in the feature vocabulary. By traversing all the information vocabulary in the tourist attraction information database, the latest feature words in the tourist attraction information can be automatically obtained and stored in the feature vocabulary database.

3 Personalized Recommendation Methods for Scenic Spots

3.1 Describe the Semantic Association of Tourist Attraction Information

Using the related technologies of Ontology Semantic Web, realize the digitization of tourist attraction information stored in the tourist attraction information database, describe terms, descriptors, titles and other normative documents, obtain descriptive metadata, and send it to users as associated data. First, according to the Cool URIs naming specification formulated by the Semantic Web, URI naming is carried out for the tourist attraction information of the tourist attraction museum. With the help of various description methods provided by FRBR, an associated data vocabulary set is created to describe the semantic ontology of tourist attraction information. The specific types are shown in Table 1:

Based on Table 1, construct a progressive transformation mechanism for the semantic association of tourist attraction information, and select the corresponding subnet type involved. Then describe the associated semantics of tourist attraction information, use entity extraction mechanism, and use D2RQ transformation tool to transform tourist attraction information into RDF metadata form. Based on this, create new semantic descriptive metadata. Finally, based on the characteristics of user dissimilarity, the publishing mode of associated data is selected to expand the information of tourist attractions, gather the Open APIs provided by local tourism, and build a network environment with stronger correlation and scheduling of tourist attraction information. By combining services and associated data, the integrated system links internal and external tourist attraction information. This completes the description of the semantic association of tourist attraction information.

Table 1. Semantic Classification Standards for the Association of Tourist Attraction Information

Semantic name	Subnet type involved	Segmentation criteria
Hierarchical relationship	P-P; K-K; M-M	Genus
Citation relationship	P-P; M-M	Entity
Related relationships	P-P; K-K; M-M	Overall part
Equivalence relationship	P-P; K-K; M-M	Synonymous
Attribute relationship	P-P; P-K; K-K; M-M	Synonymous
Discussing Relationships	K-K; K-M; M-M	Synonym
See relationship	P-K; K-K; K-M; M-M	Antonym

3.2 Selecting the Personalized Recommendation Path for the Best Tourist Attraction Information

After publishing the associated data of tourist attraction information, it is necessary to optimize the associated data network based on user access and retrieval information, select the optimal personalized recommendation path for tourist attraction information, that is, the link path of the associated data, and perform operations such as adding and modifying tourist attraction information data. Firstly, standardize the association semantics [13], count the frequency of different semantic query words, determine the core query words, determine the four attributes of the core query words, classify them, and calculate the similarity between different tourist attraction information and the word. Using the similarity distance formula, the similarity Q calculation formula is:

$$Q = w\sum_{j=1}^{4}\left(a_k - a_j\right) \tag{8}$$

In the formula, j is the four attributes of the core query word, a is the associated data of tourist attraction information, and k is the visual spatial dimension of this data [14, 15]. When $k = 1$ is used, it represents the true distance between the core query word and the spatial dimension. When $k \neq 1$ is used, it is the definite distance, representing the total absolute wheelbase on the spatial dimension.

Transform formula (8) to obtain the optimal path S for the transformation boundary distance in the visualization spatial dimension as follows:

$$S = \frac{1}{c^{Q}} \tag{9}$$

In the formula, c is the frequency at which the associated data of tourist attractions and core query words appear together. When $0 < c < 1$, the optimal path S value is between (0, 1), the smaller c, the closer S value approaches 0, and when $c > 1$, the smaller c, and the closer S value approaches 1.

Based on the frequency of the occurrence of data related to different tourist attraction information, determine the values of the k, j two link parameters, so that the S value reaches the limit value, and obtain the final path of the transformation data boundary distance. Use this path to transform the related data of tourist attraction information, and obtain personalized recommendation results of tourist attraction information.

4 Experiments

4.1 Experimental Setup

The experimental environment settings required for the experiment are shown in Table 2.

Table 2. Experimental Environment Settings

Classification	Facilities	Parameter
Computer hardware	CPU	Lenovo i5-2340@3.40 GHz
	Running memory	4 GB
	Hard drive	2T
	Graphics card	GTX 1080
Computer software	Background development language	Python
	Information processing framework	Keras
	Tourist attraction information database	SQL
	Operating system	Windows 10
	Result calculation	MATLAB

The methods used in this article, reference [4], and reference [5] were used for testing.

4.2 Result Analysis

(1) Comparing the time required to recommend tourist attraction information using different methods under the same tourist attraction information, the test results are shown in Fig. 2.

Analyzing Fig. 2, it can be seen that under the same amount of data information, the testing time of this method is shorter than that of the two comparative methods, and the highest recommended time for tourist attraction information is 4 s. In summary, it can be seen that the method in this article accurately excavates the information of tourist attractions by constructing a tourist attraction information database, reduces the time for recommending tourist attraction information, and improves recommendation efficiency.

(2) Under the same amount of information data, three methods were tested for the accuracy of recommending tourist attraction information, and the test results are shown in Fig. 3.

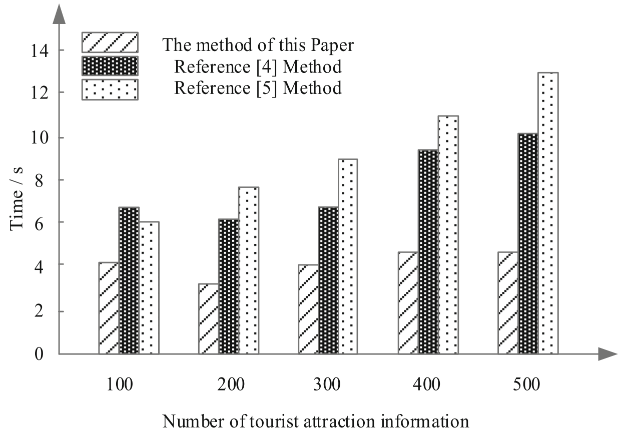


Fig. 2. Test Results of Recommended Time for Tourist Attraction Information

By analyzing Fig. 3, it can be seen that in the same amount of data information, compared with the two comparison methods, the accuracy of the proposed method in recommending tourist attraction information is higher and the detection accuracy tends to be stable. Among them, the accuracy of the proposed method for recommending tourist attraction information is higher than 88%. This is because the proposed method constructs a tourist attraction information feature vocabulary based on topic similarity clustering, which describes the semantic association of tourist attraction information and improves the accuracy of tourist attraction information recommendation.

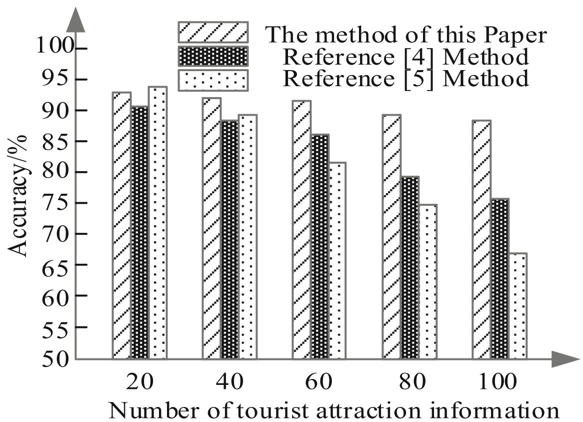
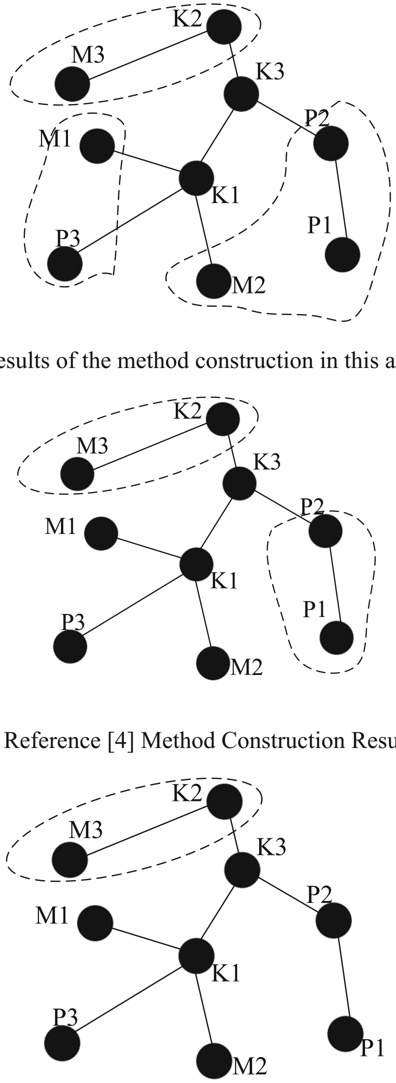


Fig. 3. Test Results of the Accuracy of Tourist Attraction Information Recommendation

(3) Three sets of experiments were conducted on 9 different tourism node elements, and their recommended tourist attraction information and comparison results are shown in Fig. 4:



(a) Results of the method construction in this article

(b) Reference [4] Method Construction Results

(c) Reference [5] Method Construction Results

Fig. 4. Experimental Comparison Results

It can be seen from Fig. 4 that the node set of the Information set of tourist attractions integrated by the reference [4] method and the reference [5] method is relatively loose. The reference [4] method constructs the association semantics of P2, M2 and K2, M3, and the reference [5] method value only constructs the association semantics of K2, M3, and the nodes P3, M2 do not establish a clear relationship with other nodes, while the association relationship between the nodes of this method, There are more methods than those in reference [4] and reference [5], and three associated semantics of P1P2M2, K2M3, and M1P3 have been constructed. The results indicate that the method proposed in this paper reveals the relevance of tourist attraction information to a higher extent than the two comparative methods, improving the personalized recommendation effect of tourist attraction information.

To further test the personalized recommendation effect of the proposed method for tourist attractions, the satisfaction level of 100 tourists with different indicators of the recommended route results using the three methods was calculated using a 10 scale scoring system. The scoring results are shown in Table 3.

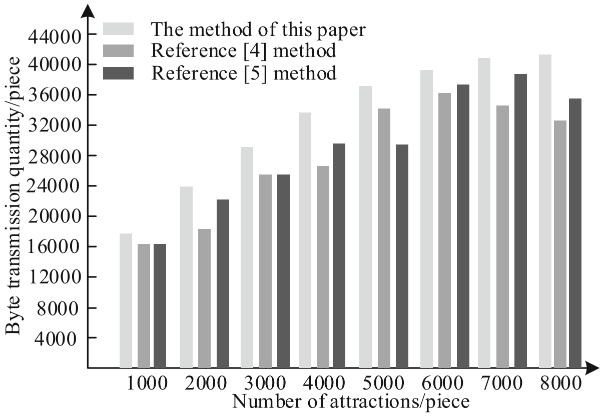
Table 3. User satisfaction results

Evaluating indicator	The method of this paper	Reference [4] method	Reference [5] method
Overall satisfaction	8.9	6.8	7.4
Overall coordination	8.6	6.7	6.8
Reflectance of Scenic Area Characteristics	9.2	7.8	6.7
Personalization level	9.3	7.4	6.9
Average	9.0	7.2	7.0

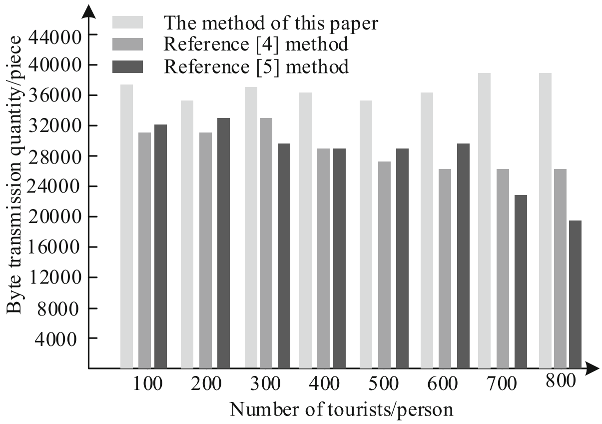
The above results indicate that the overall satisfaction, overall coordination, reflection of scenic spot characteristics, and degree of personalization of the personalized recommendation results of tourist attractions using this method are all rated at least 8.5 points by the respondents, with an average score of 9.0 points for each indicator; The average scores for various indicators in the methods of reference [4] and reference [5] are 7.2 and 7.0 respectively. The above results show that the recommended tourist attractions using this method can satisfy more users and have better design effects.

Carrying capacity is an important indicator for measuring the operational performance of corresponding methods. Complete the bearing capacity test from two aspects: different number of scenic spots and different number of tourists. Three methods are used to test the carrying capacity of different tourist attractions and numbers of tourists. The comparison results are shown in Fig. 5.

Analyzing Fig. 5, it can be seen that under the same number of scenic spots, the number of bytes transmitted per second by our method is higher than that by the methods in reference [4] and [5]. As the number of scenic spots increases, the number of bytes transmitted shows a steady growth trend, and the number of bytes transmitted changes steadily; The methods in reference [4] and reference [5] show significant fluctuations in the number of byte transfers as the number of tourist attractions increases. As the number of tourists continues to increase, the byte transmission quantity of the method proposed in this article has remained stable. In addition, the byte transmission quantity of the methods in reference [4] and reference [5] has shown a downward trend with the increase of tourists and shows significant fluctuations. The experimental results indicate



(a) Results of the method construction in this article



(b) System Carrying Performance under Different Visitor Numbers

Fig. 5. Load bearing performance test results of three methods

that the method proposed in this paper has a good carrying capacity, and the carrying capacity is not affected by the increase in the number of scenic spots and tourists.

5 Conclusion

This paper proposes a personalized recommendation method for tourist attractions based on mixed filtering of user information, and draws the following conclusion: this method can effectively improve users' selection experience and satisfaction with tourist attractions. This method combines user behavior data with user personal information and utilizes a hybrid filtering algorithm for user information to achieve personalized recommendation of tourist attractions. At the same time, this method optimizes recommendation results and filters user personal information to improve the accuracy of recommendations. The experimental results show that this method has higher recommendation accuracy and better user satisfaction. In summary, this method will be one of the important development directions for future tourist attraction recommendation systems and is expected to be widely applied and promoted. In the future, with the continuous development of big data and artificial intelligence technology, personalized recommendations for tourist attractions will become more accurate and intelligent. Recommendations based on location. Personalized tourist attraction recommendations can be further optimized based on information such as the user's current location, travel time, and real-time traffic conditions, combined with user history and social network information.

References

1. Qin, P., Jia, H., Huo, X., et al.: User personalized POI recommendation method integrating big data mining. Comput. Simul.**39**(6), 355–358,385 (2022)
2. Wu, X.: An ecological environmental carrying capacity estimation of tourist attractions based on structural equation. Int. J. Environ. Technol. Manag. **25**(4), 310–323 (2022)
3. Syahputri, J., Dharmowijoyo, D.B., et al.: Effect of travel satisfaction and heterogeneity of activity-travel patterns of other persons in the household on social and mental health: the case of Bandung Metropolitan area. Case Stud. Transp. Policy **10**(4), 2111–2124 (2022)
4. Zhang, J., Bai, S., Liu, S.: Design of travel recommendation model based on convolutional neural network. J. Terahertz Sci. Electron. Inf. Technol. **18**(6), 1128–1132 (2020)
5. Ma, Z., Chen, C., Huang, Z.: Multi-objective travel-route recommendation method based on integration of user features and group-intelligence. J. Geo-Inf. Sci. **24**(10), 2033–2044 (2022)
6. Wu, X., Yang, X., Zhu, F.: Tourist attractions recommendation model integrating context-awareness and random forest. Mod. Electron. Tech. **46**(6), 154–160 (2023)
7. Zhao, Z., Jian, Z., Gaba, G.S., et al.: An improved association rule mining algorithm for large data. J. Intell. Syst. **30**(1), 750–762 (2021)
8. Chen, J., Becken, S., Stantic, B.: Assessing destination satisfaction by social media: an innovative approach using importance-performance analysis. Ann. Tour. Res. **93**(Mar.), 103371.1–103371.19 (2022)
9. Lakshmi, N., Krishnamurthy, M.: Association rule mining based fuzzy manta ray foraging optimization algorithm for frequent itemset generation from socialmedia. Concurr. Comput. Pract. Exp. **34**(10), e6790.1–e6790.9 (2022)
10. Kaur, G., Datta, R.K.: Predicting the formation of tornadoes using association rule mining by studying a real life tornado event: Georgia, USA January, 2013. Mausam: J. Meteorol. Dept. India **72**(4), 813–820 (2021)
11. Guo, H.X., Wang, J.R., Peng, G.C., et al.: A data mining-based study on medication rules of Chinese herbs to treat heart failure with preserved ejection fraction. Chin. J. Integr. Med. **28**(9), 847–854 (2022)
12. Binding, C., Gnoli, C., Tudhope, D.: Migrating a complex classification scheme to the semantic web: expressing the integrative levels classification using SKOS RDF. J. Doc. **77**(4), 926–945 (2021)
13. Wan, S., Kamis, N.H., Ahmad, S., et al.: Similarity–trust network for clustering-based consensus group decision-making model. Int. J. Intell. Syst. **37**(4), 2758–2773 (2021)
14. Law, R., Pylkkanen, L.: Lists with and without syntax: a new approach to measuring the neural processing of syntax. J. Neurosci. Off. J. Soc. Neurosci. **41**(10), 2186–2196 (2021)
15. Martin, M.B.D.J., et al.: Novel methodology to visualize biomass processing sustainability & cellulose nanofiber product quality. ACS Sustain. Chem. Eng. **10**(11), 3623–3632 (2022)