https://doi.org/10.31449/inf.v49i5.6002 Informatica **49** (2025) 1–18

Personalized Self-Guided Tour Strategy by Integrating Random Forest Preference Model and Attractions Association Model

Huili Duan
Modern Service College, Zhengzhou Tourism College, Zhengzhou, 451464, China
E-mail: dhuili2022@163.com

Keywords: random forest algorithm, frequent pattern growth algorithm, personalization, self-guided tour, preference selection

Received: April 10, 2024

With the rapid development of tourism, personalized self-guided tours have become an important trend in the tourism market. Since traditional personalized recommendation methods often ignore elements such as users' personal preferences and characteristics of travel destinations, in order to improve the reliability of personalized self-guided tour strategy research. The study fuses the random forest algorithm preference model with the frequent pattern growth algorithm association model to provide personalized self-guided tour strategies. The study firstly utilizes the random forest algorithm to predict the attraction preference selection, and then utilizes the frequent pattern growth algorithm to mine the association relationship among the attractions, so as to provide effective data for strategy formulation. The results indicated that the prediction accuracy, recall, and data processing precision of the strategy model for attraction data were 92.07%, 93.07%, and 81.06%, respectively. The coverage of the strategy model was, however, much higher than that of the comparison model for regional attractions, featured attractions, popular attractions, and specialized attractions, at 78.09%, 85.61%, and 63.26%, respectively. This suggests that the model developed in the study can help meet the goal of the trip and greatly increase the accuracy of attractions in the process of developing a personalized self-guided tour strategy. The study intends to increase user happiness and travel suggestion accuracy in order to strongly promote the growth of the tourism industry.

Povzetek: Raziskava združuje model naklonjenosti z algoritmom naključnega gozda (RFA) in model povezav atrakcij s FP-Growth za personalizirane strategije samostojnih potovanj. Rezultati izboljšajo kvaliteto priporočil in povečajo zadovoljstvo turistov.

1 Introduction

Personalized self-guided tours (PSGT) strategy refers to customizing the most suitable travel solutions for users based on their personal preferences and preferences to enhance user satisfaction and experience. Traditional personalized recommendation algorithms mainly make recommendations based on the user's historical behavioral data or interest labels, but they often ignore the user's personal preferences and the characteristics of travel destinations [1-2]. Random forest algorithm (RFA) is an integrated learning method that can effectively solve the overfitting problem and has high prediction accuracy by combining multiple decision tree (DT) models for prediction. However, the application of RFA preference models in PSGT strategies often focuses only on the user's behavioral data and characteristics, ignoring the user's personal preferences and destination characteristics [3-4]. In addition, frequent pattern growth algorithm (FP-Growth) is an association rule mining algorithm (ARMA) that can be used to discover frequent item sets and association rules in a dataset. In the field of tourism, FP-Growth can be used to mine association rules between tourists' consumption behaviors and preferences to discover common characteristics and behavioral patterns of different types of tourists [5-6]. By mining and analyzing these association rules, useful insights and guidance can be provided for the formulation of PSGT strategies. Therefore, the study fuses the RFA preference model with the FP-Growth association model to enhance the effectiveness of PSGT strategies. RFA is first utilized to construct a prediction model of users' preference for attraction selection. Then FP-Growth is utilized to mine association rules of tourist destinations to provide more accurate recommendations. The study aims to use the fusion of RFA preference model and FP-Growth association model to enhance the recommendation effect of PSGT strategy, to provide tourists with a more personalized and high-quality tourism experience, and to promote the development and progress of the tourism industry.

The study's first section provides an overview of the state of the art in terms of machine learning (ML) research in the tourist industry. The second section of the study uses FP-Growth to mine the correlation relationship between attractions based on attraction PS after first using RFA to forecast the attraction preference selection (PS). Through simulation studies, the application performance of the developed PSGT-strategic model (SM) is confirmed in the third section. The experimental results are compiled in the fourth section, which also examines the benefits and drawbacks of the study's methodology.

2 Related works

PSGT strategy research is particularly important in the context of the current increasingly diverse tourism market. To achieve this objective, a large number of specialists and academics have focused on the fields of ML and data mining. Among them, RFA and FP-Growth, as two representative technical tools, have shown great potential and value in the research of PSGT strategies. Deng et al. proposed an edge computing-based travel destination selection preference prediction method to improve the accuracy and prediction duration of travel destination selection preference prediction. The study utilized edge computing to select preference features, and then utilized the RFA to sort and get the weight value, and then performs preference prediction based on the weight value. The results showed that the method can significantly improve the prediction accuracy and also shorten the time-consuming process of selecting tourist destinations [7]. Vieira et al. analyzed the techniques of acceptance extension model, perceived usefulness, and perceived ease of use in social networks in order to assess the behavior of travelers in choosing tourist attractions. The study was conducted by analyzing social data from different social software and considering the above three factors. The study of tourists' behavioral intentions was greatly improved by this strategy, according to the results [8]. Huda et al. conducted a study on sustainable digital strategies for hotels in order to enhance the market expansion of budget hotels. The study was conducted to enhance the operational practices of hotels by restoring financial sustainability through surveys of different customers and enhancing digital market strategies in resilient sectors of the tourism industry. The results showed that the study was able to consistently gain a small advantage in the sustainability of frequent guest maintenance [9]. Yen et al. proposed a new approach based on network data envelopment analysis and social network analysis to realize the potential research on the accessibility of different tourist destinations. The study realized the sharing of network data by classifying and statistically analyzing the data of tourism intermediaries. The results showed that the method can effectively help destination managers and governments to identify tourism opportunities and implement appropriate strategies [10].

Liang used RFA to build a regression model based on data mining and data learning to investigate the growth of rural tourism. RFA was used to analyze and summarize the tourism service strategies of rural revitalization regions. The results indicated that RFA was able to weigh the importance of the variables in the service strategy, and thus could be applied to the development of the service strategy [11]. Zhang et al. applied ML methods such as RFA to the strategy research of customized tourism through the study of big data customized tourism. This suggested that users could receive customized travel search services through data mining and classification utilizing several big data elements related to tourism [12]. To provide tourists with effective visit information, the frequent item set method was employed to mine the points of interest in different attractions, as proposed by Höpken et al. The study analyzed tens of thousands of tourist image information to cluster the attractions that are of interest to potential tourists, and then used association rules for data mining. The results showed that the method is able to increase the support level to 4.6%. [13]. Shabtay et al. constructed a new bootstrap frequent pattern for the mining of different program interests through a multi-objective mining algorithm. The study utilized the bootstrap frequent pattern to be able to find the interest relationships of different item sets quickly while using a small amount of memory. The results indicated that the bootstrap frequent pattern can significantly improve the problem of counting class rules among different items [14]. In order to conduct an effective analysis of the relevant research, the study analyzed and summarized the research conducted by the experts and scholars mentioned above, as shown in Table 1, which is the result analysis table obtained by citing the methods used in the literature.

Table 1: Analysis of results obtained from the method of using citations

Literature	Main methods	Main findings	Performance index
Deng et al	Edge algorithm	Significantly improve prediction accuracy and shorten time consumption	Improved prediction accuracy and reduced time consumption
Vieira et al	The acceptance expansion model in social networks	Analysis of significantly enhancing tourist behavioral intentions	Improving the accuracy of behavioral intention analysis
Huda et al	Research on sustainable digital strategy for hotels	Continuously gaining small advantages in sustainable development through frequent customer maintenance	Financial sustainability recovery and improved customer maintenance effectiveness

Yen et al	Based on network data envelopment analysis	Effectively help identify tourism opportunities and execute strategies	Enhancing the ability to discover tourism opportunities
Liang	Data mining and random forest algorithm	The random forest algorithm can balance the importance of variables in service strategies	Improving the accuracy of service strategy formulation
Zhang et al	Research on big data customized tourism	79.84% of customers are willing to purchase customized big data services again	Users have a high willingness to purchase again, and personalized search services have good results
Höpken et al	Identifying frequent item sets and mining interest points	Support increased to 4.6%	Improving the accuracy of interest point mining
Shabtay et al	Constructing guided frequent patterns using multi-objective mining algorithms	Significantly improve the numerical rule problems between different projects	Improved speed in mining interest relationships between projects and reduced memory usage

Through the analysis of the above research, it is found that the random forest preference model and the scenic spot association model play a very important role in the development of PSGTs. Using the RFA can effectively improve the accuracy in the prediction of attraction PS, and the FP-Growth algorithm can significantly improve the mining of the association relationship between different attractions. Additionally, the study will provide a clear comparison of the performance metrics of each reference [6], [8], and [12] with the model approach in terms of performance. The approach of Deng et al. [6] is particularly effective in real-time and prediction accuracy. The approach of Vieira et al. [8] is highly specialized in understanding travelers' social behaviors and psychological preferences. Finally, the approach of Zhang et al. [12] is particularly effective in providing personalized services. In contrast, the RFPAP model is particularly adept at the scientific assessment and analysis of tourism resources, while the FP-Growth model is highly effective in mining association rules from large amounts of data. The integration of these methods may result in more comprehensive and accurate solutions in the field of personalized travel recommendation. In summary, the use of RFA can effectively improve the accuracy in the prediction of attraction PS, and the FP-Growth algorithm can significantly improve the mining of correlations between different attractions. Based on this study, the two algorithms are integrated and used to construct a PSGT strategy model. The study aims to provide more and more favorable data support for the development of tourism.

3 Self-guided tour strategy by integrating RFPAP model and FP-growth model

To meet the personalized needs of tourists, the study designs a self-guided tour strategy based on data mining by fusing the RFPAP model with the FP-Growth model. This strategy will utilize advanced data analysis methods to deeply mine tourists' tourist attraction preferences (PTA) and the correlations between attractions, so as to provide tourists with more accurate and personalized travel route suggestions. The study utilizes RFA to predict PTA and FP-Growth algorithm to mine the correlations among attractions on the basis of prediction. The design of PSGT strategy is accomplished through the fusion of the two models.

3.1 RFA-based attraction preference prediction model construction

To analyze the attraction selection of PSGT, the study utilizes RFA for predictive analysis of PTA. In the prediction of tourism PTA using RFA, the data of attraction selection by itself is of little value, but it is effectively categorized and processed by RFA. This is used as a basis for obtaining information about attraction PS from a large amount of tourism data, and it also improves the prediction accuracy when multivariate prediction is performed. Through extensive analysis of RFA, the algorithm is composed of a large number of DTs [15-16]. This DT is the predictive classifier in the algorithm, and by effectively analyzing the DT, the sensitivity and accuracy of data classification in self-guided tour attractions can be improved. The flow chart of DT construction based on attraction selection is shown in Fig. 1.

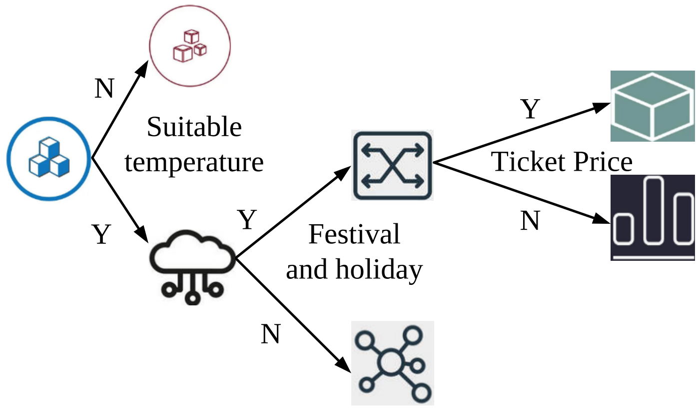


Figure 1: Process diagram for constructing a decision tree based on scenic spot selection

Fig. 1 illustrates the decision-making process employed by the research team. It can be observed that the suitability of weather conditions is initially assessed, and subsequent decisions are then made based on ticket prices. If the ticket price is deemed appropriate, further consideration will be given to whether the scenic spot has any festivals or activities. Conversely, if the price is deemed unsuitable, or for other reasons, users will not choose the attraction. The decision-making process of scenic spot selection is clearly displayed by the clear definition of nodes and branches. This enables users to make more suitable choices. In the process of predicting tourism PTA using RFA, the specific operation to improve the accuracy of classification is divided into 3 steps. The first step is to use the sampling technique to extract the original data, the extraction process is a put-back operation, and the corresponding training set can be obtained through random extraction. These datasets are classified, and each extraction will have about 36.8% of the data not sampled, and these unsampled data forms the out-of-bag (OOB) [17-18]. The OOB error can be calculated to evaluate the performance of the model. Therefore, OOB can be used for performance analysis of RFA. Then on the basis of obtaining the OOB, the study sets the features in each sample to M. At this point, a variable is extracted for each node in each tree. This variable will be analyzed to find the strongest variable of classification ability, and combined with the threshold value is used for the determination and analysis of classification nodes. Finally, the study utilizes the generated random forest to classify the tourist attraction data and the classification result is based on the number of votes of the classifier. A schematic diagram of attraction prediction by RFA is shown in Fig. 2.

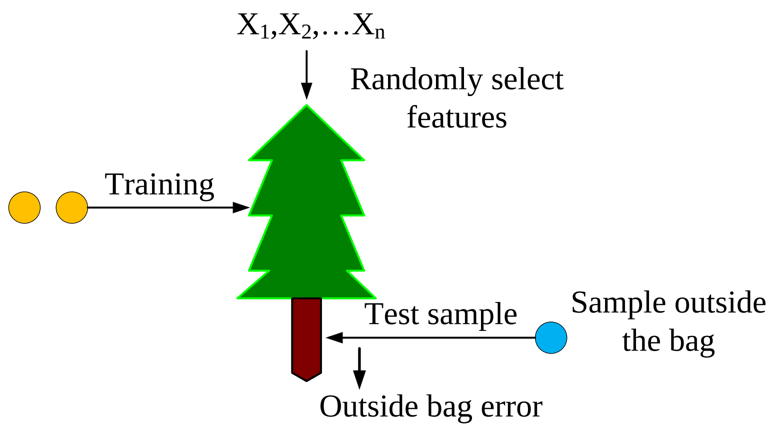


Figure 2: Schematic diagram of scenic spot prediction using random forest algorithm

Fig. 2 illustrates the sample data points utilized for training and testing. It can be observed that two-thirds of the data points are employed for training, while the remaining one-third are utilized for testing. The green tree icon represents the remaining samples that have not yet been included in the current training or testing process. The random selection of features serves to mitigate the risk of overfitting and enhance the model's capacity for generalization. The yellow circle represents "out-of-set samples," which, along with their corresponding "out-of-set errors," are employed to assess the efficacy of the model on previously unseen data, thereby providing an additional metric for evaluating the model's generalizability. The study must obtain a classification matrix in order to examine the expected values of attraction selection and guarantee the accuracy of the classification results. The classification matrix constructed by the study can be represented by Equation (1).

$$A=\begin{Bmatrix} a_1 f_1 & a_1 f_2 & \cdots & a_1 f_p \\ a_2 f_1 & a_2 f_2 & \cdots & a_2 f_p \\ \vdots & \vdots & \ddots & \vdots \\ a_n f_1 & a_n f_2 & \cdots & a_n f_p \end{Bmatrix} \quad (1)$$

In Equation (1), A denotes the classification matrix and f_p is the feature p value corresponding to the classification process. $a_n f_p$ is the feature p value corresponding to the n th tourist. After calculating the obtained classification matrix, the eigenvector values of tourists in attraction selection can be expressed by Equation (2).

$$X=\left(X_1,X_2,\cdots,X_n\right) \qquad X\in A \quad (2)$$

In Equation (2), X_n denotes the feature vector of the n th visitor at the time of selection. The predicted value of attraction selection at this time can be expressed by Equation (3).

$$Y=f(X)=\left(y_1,y_2,\cdots,y_n\right) \quad (3)$$

In Equation (3), $f(X)$ denotes the classification function of RFA, and y_n denotes the predicted value of PTA for the n th tourist. It is discovered that the sorting of DT would significantly affect the prediction outcomes in the PTA prediction process. The sorting judgment according to the majority principle can be employed to obtain the importance sorting results of the selection variables. Furthermore, the credibility of the sorting results of the majority principle is considerably higher than that of the sorting results of the selection of single factors. In the process of feature sorting, it is necessary to judge the metric importance of feature variables to determine the sorting principle [19-20]. The study uses the OOB data as a test set based on the previous OOB data as a way to judge the generalization ability when DT classification. The corresponding sorting principle at this point can be expressed by Equation (4).

$$I_k(i)=OOB(E_k)_i-OOB(E_k) \quad (4)$$

In Equation (4), $OOB(E_k)_i$ denotes the i th incorrectly assessed eigenvalue in the incorrect assessment result. $OOB(E_k)$ denotes the reordered misassessment value. Equation (5) can be used to calculate the DT importance measure after obtaining the ranking principle.

$$I(i)=\frac{\sum_{i=1}^{N} I_k(i)}{N} \quad (5)$$

In Equation (5), N denotes the number of DTs. Combining the above studies, the study constructed the random forest preferred attraction prediction (RFPAP) model on the basis of RFA. The prediction flowchart of this model is shown in Fig. 3.

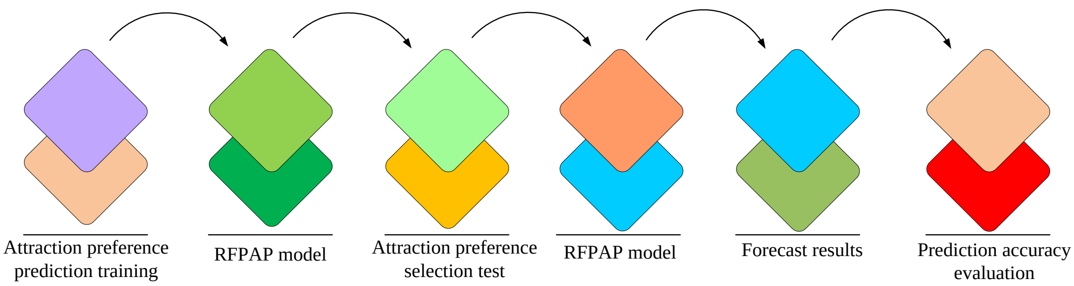


Figure 3: Process diagram of preference prediction model for random forest scenic spot selection

Fig. 3 illustrates the two-stage structure of the scenic spot selection preference prediction model developed in the study. The model comprises a training stage and a testing stage. Firstly, during the training phase, the study employs the RFA to classify and process the raw data, with the objective of selecting representative training datasets and identifying the optimal parameter configuration through algorithm iteration. Subsequently, during the testing phase, the selected and optimized dataset will be applied to the RFPAP model for the actual prediction of attraction selection preferences, and the prediction results will be recorded. Finally, in order to evaluate the performance of the model, the predicted results are compared in detail with the actual selection values of tourists. This comparison allows for the calculation of the prediction accuracy of the model, thereby ensuring its accuracy and reliability.

3.2 Design of attraction association model based on FP-Growth algorithm

Through the study of PTA prediction, it is found that PSGT tourists do not choose a single attraction in the process of attraction selection. Instead, they will choose based on the association between attractions and attractions. To better analyze the self-guided tour strategy, the study utilizes ARMA to mine the association relationship between different attractions, so as to get more accurate PS prediction results [21-22]. The study chooses FP-Growth as the ARMA between attractions. This algorithm is a class of mining algorithms obtained by improving on the basis of Apriori algorithm. The effectiveness of correlation mining between attractions can be greatly increased by reducing the need for repetitively scanning the attraction database while also preventing the production of candidate datasets throughout the mining process. In the process of attraction correlation mining using the FP-Growth algorithm, the attraction database needs to be compressed first, and then the compressed database is divided into a complete set of correlation database [23-24]. Through the compression of the database, it can effectively reduce the amount of computation when attraction data relevance mining. The flowchart of attraction relevance data mining based on FP-Growth algorithm is shown in Fig. 4.

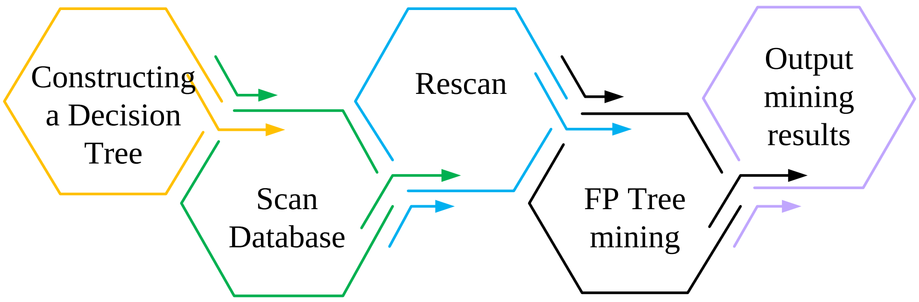


Figure 4: Flowchart of attraction association data mining based on FP-Growth algorithm

In Fig. 4, the FP-Growth algorithm mines the attraction association data in the process of traversing the database twice to search for the data information it needs. This can significantly reduce the database scanning time and improve the efficiency of the relationship mining. The quantity of attraction data information will have a big influence on how accurate associative mining is in the attraction associative relationship mining process. Therefore, the study uses web crawler technology to obtain the attraction data in a tourism website under authorized circumstances. By obtaining the data of tourists in selecting multiple attractions to visit, the FP-Growth algorithm is able to obtain the likelihood of tourists selecting the next tourist attraction through data mining [25-26]. The research of PSGT attraction recommendation strategy is carried out on this basis. When organizing and correlation mining the data of tourist attractions, it is necessary to design the corresponding correlation mining table according to the goal of mining. As shown in Table 2 is the research based on a city's tourist attractions designed inter-attraction association rule mining information table.

Table 2: Information on mining association rules between scenic spots

Attraction and tourist selection association number	Attraction ID	Attraction associated markers
G1	J1	Y
G2	J2	Y
G3	J3	Y
G4	J4	Y
G5	J5	Y
G6	J6	Y
G7	J7	N
G8	J8	Y
G9	J9	N
G10	J10	N

Equation (6) illustrates how the study created an evaluation matrix utilizing the association rule mining data in Table 1 in order to assess the correctness of association connection mining through the collection of mining information.

$$V=(V_{ik})_{m\times n}=\begin{bmatrix} V_1 & V_{11} & V_{1n} \\ V_2 & V_{22} & V_{2n} \\ \vdots & \vdots & \vdots \\ V_m & V_{m2} & V_{mn} \end{bmatrix} \quad (6)$$

In Equation (6), V indicates each indicator value, $i=1,2,\cdots,m$, $k=1,2,\cdots,n$, $m=4$, $n=17$. After the indicators are constructed, they must be normalized in order to be compared and analyzed. Equation (7) can be used to calculate the standardized values.

$$X_{ik}=\frac{V_{ik}-\min_k V_{ik}}{\max_i - \min_i V_{ik}} \quad (7)$$

In Equation (7), V_{ik} is the normalized k value corresponding to the i th indicator. After normalizing the data, the correlation of the data can be calculated and analyzed on the basis of the normalization process, and the formula for the correlation can be expressed in Equation (8).

$$\zeta_{ik}=\frac{\min_i\min_k\left|X_{0k}-X_{ik}\right|+\rho\max_i\max_k\left|X_{0k}-X_{ik}\right|}{\rho\max_i\max_k\left|X_{0k}-X_{ik}\right|+\left|X_{0k}-X_{ik}\right|} \quad (8)$$

In Equation (8), $\left|X_0k-X_ik\right|$ denotes the absolute value corresponding to the sequences X_0 and X_i at point k. $\min_i\min_k\left|X_0k-X_ik\right|$ denotes the minimum value of the bipolar absolute value of the two sequences. $\max_i\max_k\left|X_0k-X_ik\right|$ denotes the maximum value of the bipolar absolute value of the two sequences. ρ denotes the coefficient. At this point, it is necessary to confirm the weight value of each evaluation index, which can be expressed by Equation (9).

$$\begin{cases} W1=(\omega_{11},\omega_{12},\omega_{13}) \\ W2=(\omega_{21},\omega_{22},\omega_{23}) \\ W3=(\omega_{31},\omega_{32},\omega_{33}) \end{cases} \quad (9)$$

After obtaining the weight values of the assessment indicators at all levels, the attraction association model can be constructed by utilizing the mining results of the correlation rules between attractions. The purpose of constructing the attraction association model is to improve the accuracy of the correlation judgment between attractions and attractions, and the attraction association process of the model is shown in Fig. 5.

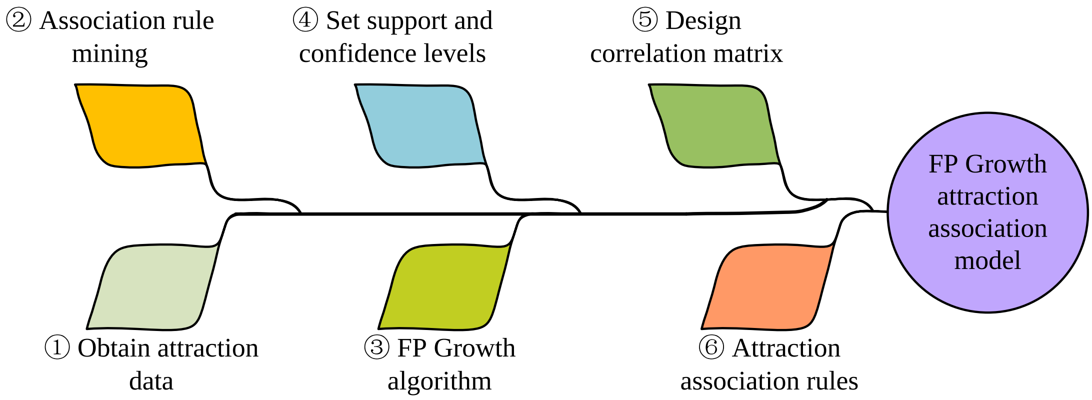


Figure 5: Flow chart of attraction association based on FP-Growth algorithm

Fig. 5 illustrates the process of constructing a scenic spot association model using the FP-Growth. The data processing and setting stage involves obtaining scenic spot data, analyzing the correlation degree betweenscenic spots, using the algorithm to mine the correlation degree, setting confidence and support, designing the correlation coefficient matrix, mining scenic spot rules, and finally outputting the correlation degree model.

3.3 SM design of personalized self-guided tours based on RFPAP and FP-Growth

By utilizing the RFPAP model for the preference study of attraction selection and the FP-Growth model for the mining of attraction association relationships, the study completes the design of a tourism strategy for PSGT. The goal of the plan is to offer travelers a more precise and customized travel experience to cater to their individual needs and tastes. First, the RFPAP model is used to study tourists' preferences in the attraction selection process. The model mines the characteristics and influencing factors of tourists' preferences for attractions by analyzing tourists' historical behavioral data. When using the model to analyze personalized self-help tours, it is necessary to set the relevant parameters, and the steps of parameter setting are as follows: ① Data preprocessing: cleaning and integrating tourism data, and conducting necessary transformation and standardization. ② Determine the FP-Growth model parameters: Set the minimum support to filter out frequent and complex items. ③ Set RFPAP model parameters: Set relevant parameters based on model definition and business requirements. ④ Fusion strategy parameter setting: Determine the fusion method and set weight or priority parameters. ⑤ Model training and validation: Train and validate models using historical data, evaluate parameter settings. ⑥ Parameter optimization: Utilize automation tools to optimize parameters and improve model performance. ⑦ Record: Record parameter settings and optimization processes, and write documentation to explain. ⑧ Continuous adjustment: Monitor model performance after going live and adjust parameters based on actual conditions [27-28]. By understanding tourists' preferences, attraction recommendations and travel route suggestions can be provided to tourists that better meet their needs. Second, the FP-Growth model is used to mine the associative relationships among attractions. The model analyzes a large amount of attraction data to discover potential connections and common features between attractions. By understanding the associative relationships among attractions, it can provide tourists with more systematic and comprehensive travel route suggestions, and promote tourists' in-depth visit and experience of attractions [29-30]. Finally, based on the output of the RFPAP model and FP-Growth model, the study designed a PSGT strategy. The strategy takes into account tourists' preferences and the correlation between attractions to provide tourists with personalized tour itinerary suggestions. By optimizing the tour sequence, attraction recommendations and activity arrangements, it improves tourists' travel experience and satisfaction. The specific strategy content is: according to the tourists' personalized needs and time schedule, develop a self-guided tour strategy that meets the needs of tourists. The strategies include tour order, attraction recommendations, and activity arrangements to improve tourists' travel experience and satisfaction. In the practical application, feedback and evaluation of tourists' self-guided tour strategies are collected. The preference model and association model are updated and adjusted according to the feedback data to adapt to the changing market demand and tourists' preferences. Regularly evaluate the performance of the models and make necessary optimizations and improvements to maintain the accuracy and effectiveness of the models.

In the PSGT strategy formulation process, in order to evaluate the strategies effectively, the study used strategy support, strategy confidence and strategy enhancement as measures for the performance evaluation of the formulated strategies. Strategy support is the percentage of the formulated strategy that can satisfy the sample, and the support can be calculated using Equation (10).

$$\sup port(C \to D) = \frac{q(C \cup D)}{N} \tag{10}$$

In Equation (10), C and D denote data that are not intersected in the strategy. $q(C \cup D)$ denotes the data that contain the same attractions in the strategy at the same time. Strategy confidence refers to the proportion of different attractions appearing in the strategy. The higher the confidence rating, the more likely it is that the same attractions would recur, and hence the more reliable the approach that is developed. The strategy confidence can be expressed by Equation (11).

$$confidence(C \to D) = \frac{q(C \cup D)}{q(C)} \tag{11}$$

In Equation (11), $q(C)$ denotes the minimum confidence level. The strategy enhancement degree can be expressed in Equation (12).

$$life(C \to D) = \frac{q(C \cup D)}{q(D)} \tag{12}$$

In Equation (12), $q(D)$ denotes the probability value of the degree of enhancement. By analyzing and summarizing the design process of PSGT strategy, the study successfully completed the design of PSGT strategy. The flowchart of SM design of PSGT based on RFPAP and FP-Growth is shown in Fig. 6.

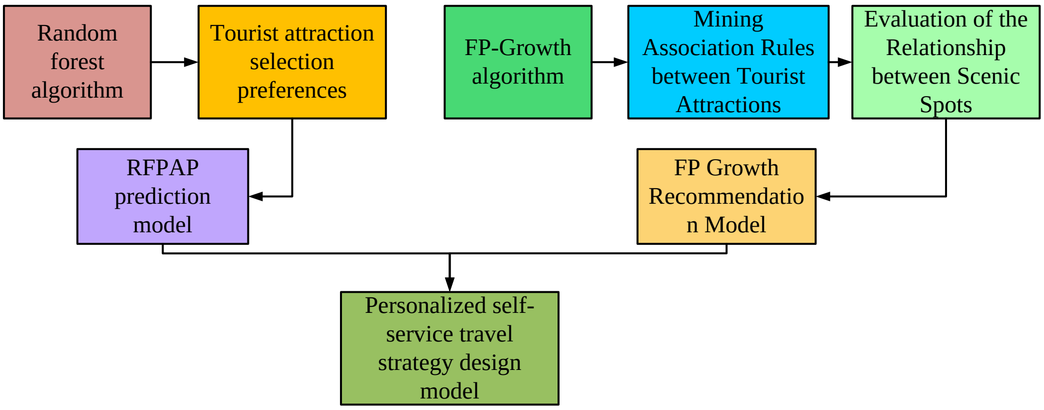


Figure 6: Design flowchart of personalized self-service travel strategy model based on RFPAP and FP-Growth

Combined with the analysis of Fig. 6, it can be noticed that the research utilizes the crawler technology to obtain the data information of tourist attractions, and uses the RFPAP model to complete the PS of attraction selection. The attraction association model is then created by mining the association relationship between attractions using the FP-Growth algorithm. Finally, the two models are fused and used to construct the design of PSGT strategy and complete the self-guided tour strategy output.

4 Performance analysis of self-guided tour strategic model integrating RFPAP model and FP-Growth model

To validate the performance of the PSGT strategy design that incorporates the RFPAP model and the FP-Growth model, the study compares the support vector machine (SVM), back propagation neutral network (BPNN) with the RFPAP+FP-Growth model. The comparison of the three models is used to validate the specific performance of the SM for PSGT constructed by the study. To design a reliable, PSGT strategy, the study obtains data from tourism websites and official government websites. The content of the tourism website reflects the preferences of tourists, while the official government website provided accurate basic information. To ensure the accuracy of the data, measures are taken to remove duplicates and erroneous data, as well as to deal with missing values. Finally, a dataset comprising 100 data points on attractions in the same province is constructed for the purpose of evaluating the performance of the PSGT strategy model. These measures ensure the accuracy, consistency, and professionalism of the data, thereby providing a solid foundation for the design of PSGT strategies.

4.1 Performance analysis of RFPAP model for attraction preference prediction

The study compares SVM, BP, and SM in order to validate the prediction models' performance in the PTA prediction process. As demonstrated by the comparison findings of predictive accuracy and recall in various extracted data ratios for the three models in Fig. 7, these metrics are employed as validation measures for performance testing.

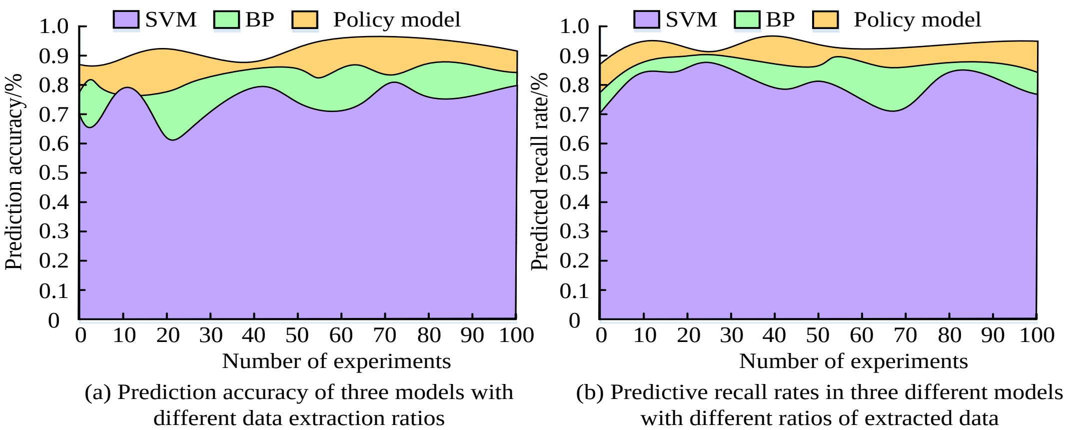
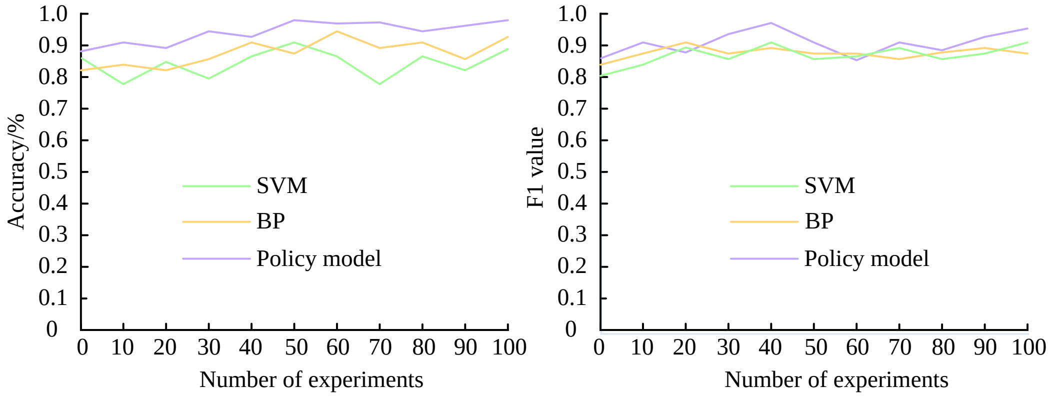


Figure 7: Comparison results of prediction accuracy and recall in three different models with different data extraction

ratios

In Fig. 7(a), the SM has the highest prediction accuracy of 92.07% for attraction data among different attraction extracted data. In addition, the prediction accuracy of SVM and BP model for attraction data is 78.95% and 83.76% respectively. In Fig. 7(b), in the comparison of the recall of data extracted from different attractions, the recall of attraction data of SM is 93.07%. The attraction data recall of SVM and BP models are 81.26% and 88.61%, respectively. This suggests that the prediction model is more resilient and reliable in terms of recall rate and accuracy of data prediction. In order to further validate the processing ability of SM in tourist attraction data, the study takes the precision rate and F1 value of PTA prediction as the validation indexes. Fig. 8 displays the comparative findings of the three models' F1 value and precision rate in PTA prediction data processing.



(a) The accuracy of three models in processing tourist attraction preference prediction data

(b) F1 values of three models in processing tourist attraction preference prediction data

Figure 8: Comparison of accuracy and F1 value of three models in processing tourist attraction preference prediction data

There are some variations in the three models' capacities to handle PTA data when processing PTA projection data, as shown in Fig. 8(a). The SVM model has the lowest accuracy, while the SM model has the best accuracy. SVM, BP, and SM have data processing accuracy of 81.06%, 89.73%, and 93.58%, in that order. The F1 values of SVM, BP, and SM are 0.81, 0.86, and 0.92 in Fig. 8(b), in that order. This suggests that the study's SM may provide more useful data and has a greater accuracy in PTA data processing. Additionally, the model's performance is noticeably superior to that of the comparative model. To verify the attraction coverage ability of the SM in the attraction PS process, the study used the attraction coverage as a validation index and selected regional attractions, specialty attractions, popular attractions and niche attractions. In addition, it is used to verify the coverage ability of SM during PS. The comparison results of attraction coverage of the three models in attraction PS are shown in Fig. 9.

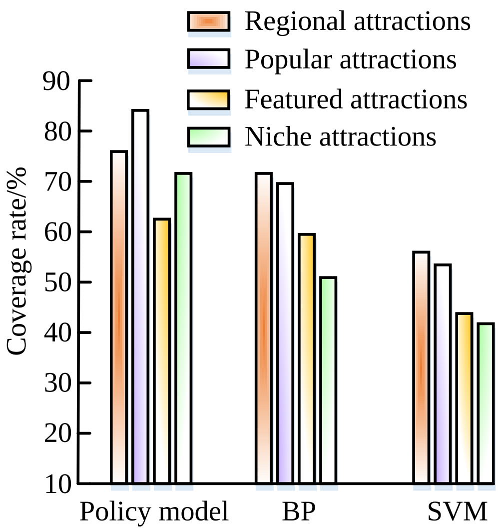
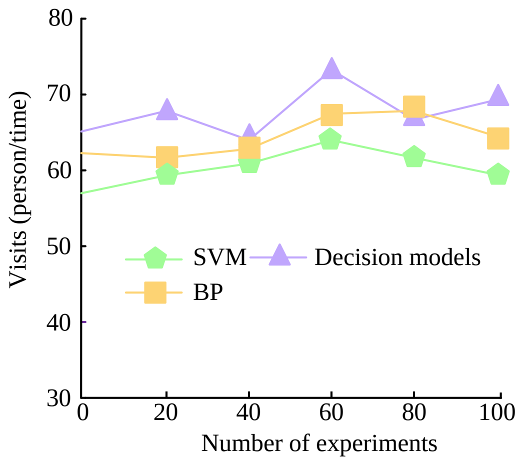


Figure 9: Comparison of attraction coverage rates among three models in attraction preference selection

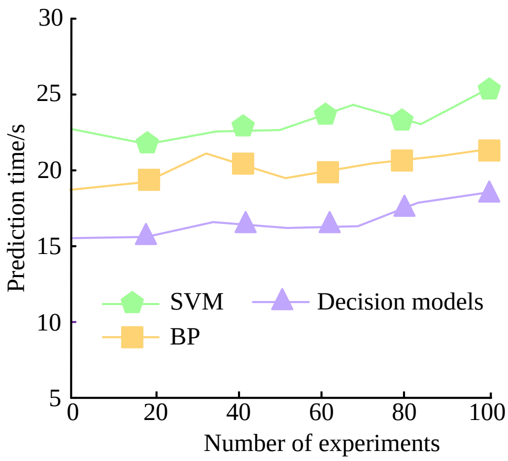
Attraction coverage plays an important role in forecasting preferences for tourist attractions. It can represent the diversity of attractions and shows how many tourist attractions there are in an area relative to all the attractions. Attraction coverage can influence the attraction recommendation results, help synthesize user needs, and plan travel routes. By considering attraction coverage, user preferences for tourist attractions can be better met. This indicates that the higher the attraction coverage, the greater the tourists' preference for choosing attractions may be when predicting preferences for tourist attractions. In Fig. 9, the attraction coverage of SM in the prediction process of PTA is significantly higher than that of BP model and SVM model. The coverage rates of SM, BP model, and SVM model in regional attractions, specialty attractions, popular attractions, and niche attractions are 78.09%, 85.61%, 63.26%, and 72.05%, 72.55%, 69.83%, 58.66%, and 47.81%, 57.0%, 54.22%, 43.93%, and 40.15%, respectively. This suggests that the study's constructed SM performs more steadily in PTA prediction.

4.2 Performance analysis of attraction association for FP-Growth modeling

To validate the performance of SM in attraction association mining, the study uses the visitor's visit and stay time as validation metrics for exploring the performance of association mining. The comparative findings of the three models in the attraction association process are displayed in Fig. 10 with regard to visit volume and association prediction time.



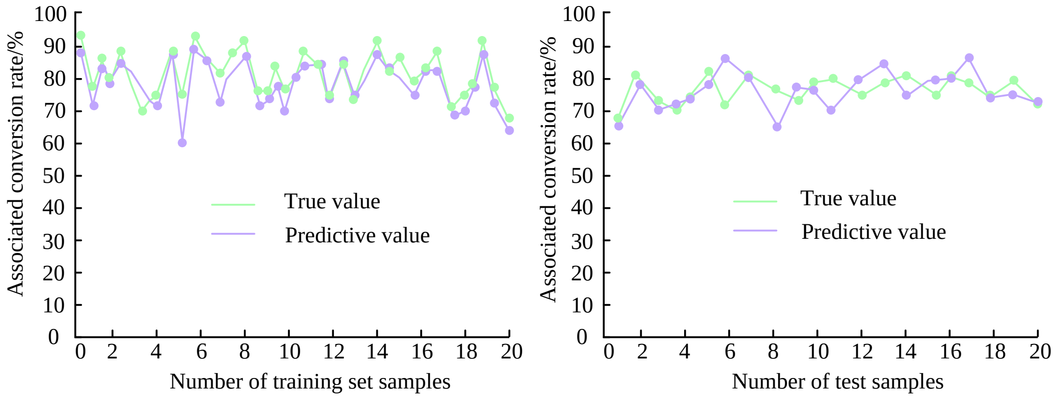
(a) Comparison of Visits Related to Scenic Spots in Three Models



(b) Comparison results of three models for predicting time related to scenic spots

Figure 10: Comparison results of visit volume and association prediction time of three models in the process of scenic spot association

In Fig. 10(a), there is still a gap between the mining ability of the three models in the process of mining the visits to the attractions' associative relationships. The average visits through the correlation mining SVM model is 62 people/time, the average number of visits of BP network is 68 people/time, and the average visits of SM is 73 people/time. In Fig. 10(b), in the process of attraction correlation mining, the three models' correlation prediction time shows an increasing trend of time consuming with the increase of the experiments. The average of prediction time consuming of SM in the process of attraction correlation prediction is 15.8 s. The average of prediction time consuming of BP model is 18.9 s. The average of prediction time consuming of SVM model is 22.9 s. This shows that the SM designed by the research is capable of finding the correlation between attractions in correlation mining to find the association between attractions, thus increasing the number of attraction visits and reducing the prediction time consumed. To verify the application effect of the SM, the real value of the attraction association conversion rate is compared with the predicted value of the SM. As shown in Fig. 11, the results of the comparison between the real and predicted values of the association conversion rate are shown.



(a) Prediction results of training samples
(b) Prediction results of test samples

Figure 11: Comparison results between actual and predicted values of association conversion rate

In Fig. 11(a), in the training set, the average of the true association conversion rate is 88.32% and the average of the SM predicted association conversion rate is 80.05%. In Fig. 11(b), in the test set, the average value of the true association conversion rate is 77.29% and the average value of the SM predicted association conversion rate is 73.61%. The comparison reveals that although the performance evaluation index value of the modeling approach in the test samples is not as good as that in the training samples, it has some performance advantages. In addition, the difference with the real value is not big, 8.27% in the training set and 3.68% in the test set. This suggests that SM is highly applicable. SVM, BP, and SM are used for comparison in order to confirm the accuracy of SM in the attraction association relationship mining process. Fig. 12 displays the findings of a comparison of the three models' associative relationship accuracy.

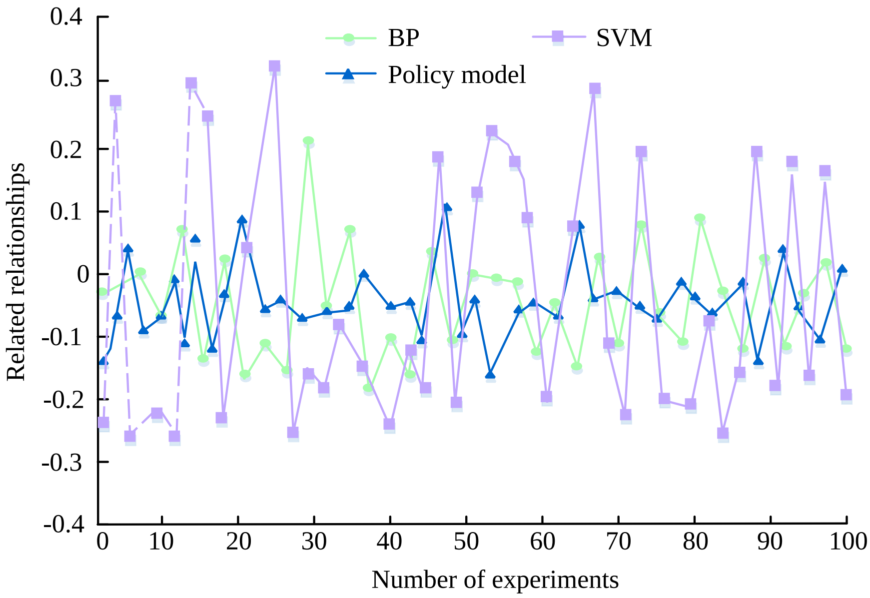


Figure 12: Comparison of accuracy of correlation relationships among three models

The errors of all three models in the correlation prediction process in Fig. 12 exhibit notable changes as the number of iterations rises. The SVM method has a -0.29 negative prediction error and a 0.34 positive prediction error. The BP model's negative prediction error is -0.21 and its positive prediction error is 0.24. The SM has a -0.15 negative prediction error and a 0.08 positive prediction error. This suggests that the SM has the smallest error range of the three models in performing correlation mining has the smallest error range, followed by BP, and the worst is SVM. The positive and negative errors of SM are 0.26 and 0.14 lower than those of SVM. This indicates that SM has a stronger mining ability in the correlation mining process, which can provide richer data results for attraction strategy development.

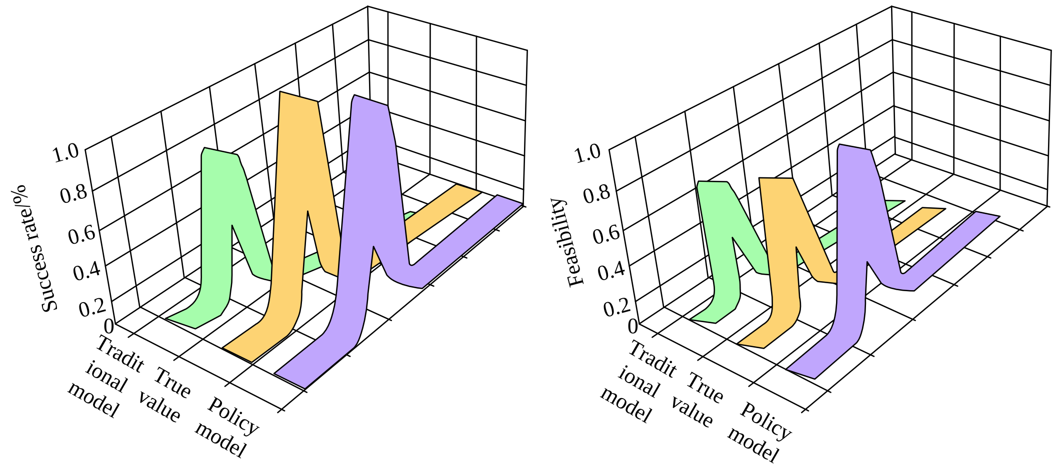
4.3 Strategic model application performance analysis of personalized self-guided tours

To verify the application performance of the SM of PSGT, the study takes the policy support, policy confidence, and policy enhancement as the verification indexes, and compares the analyzed values of the SM with the real values. Additionally, Table 3 displays the comparison results.

Table 3: Comparison of support, confidence, and improvement results

Attraction and tourist selection association number	Strategy model analysis value			True value		
	Support	Confidence level	Enhancement degree	Support	Confidence level	Enhancement degree
1	0.18	0.65	1.08	0.16	0.71	1.37
2	0.16	0.61	1.13	0.4	0.68	1.26
3	0.21	0.6	1.09	0.13	0.69	1.29
4	0.19	0.61	1.07	0.12	0.66	1.4
5	0.21	0.59	1.08	0.15	0.63	1.35
6	0.16	0.57	1.09	0.12	0.67	1.28
7	0.17	0.60	1.12	0.13	0.65	1.31
8	0.18	0.57	1.05	0.14	0.62	1.28
9	0.21	0.59	1.07	0.13	0.70	1.33
10	0.16	0.58	1.15	0.11	0.65	1.29

In Table 3, the mean support, confidence and enhancement of the analytical value of the attraction-tourist choice association of the SM are 0.183, 0.597 and 1.093, respectively. The mean support, confidence and enhancement of the real value of the attraction-tourist choice association of the SM are 0.159, 0.666 and 1.316. The difference between the analytical value and the real value of the support, confidence and enhancement is 0.024, 0.069 and 0.223. The differences are 0.024, 0.069 and 0.223. This indicates that the gap between the SM of PSGT designed by the study using the RFPAP model and the FP-Growth model and the tourists' actual appearance of attraction choices is not large, and has a strong feasibility. The study contrasts the traditional model, the SM with the real value, and uses the success rate of attraction strategy suggestion and the feasibility of excursion strategy recommendation as the validation indexes in order to further validate the practical application effect of the SM of PSGT. Fig. 13 shows the results of attraction strategy recommendation success rate and feasibility comparison.



(a) Success rate of attraction recommendation strategy (b) Feasibility of attraction recommendation strategy

Figure 13: Successful recommendation of tourist attraction strategies and feasibility comparison results

In Fig. 13(a), during self-guided tour, the success rate of real recommendation is 79.26%, the predicted value of SM recommendation is 82.17%, and the predicted value of recommendation of traditional model is 73.68%. In Fig. 13(b), in the feasibility of attraction tour, the real feasibility is 63.86%%, the predicted value of SM feasibility is 65.94%, and the predicted value of feasibility of traditional model is 59.87%. The difference between the predicted value and the feasibility prediction value of SM and traditional model is smaller compared with the true value, which indicates that SM has some enhancement performance and can improve the success probability of attraction selection. The comparison results of adaptability and robustness validation for different scenic spots are shown in Table 4.

Table 4: Comparison results of adaptability and robustness validation for different scenic spots

Types of attractions	Sample quantity	RFPAP+FP-Growth model prediction accuracy/%
Cultural attractions	100	93.15
Natural attractions	100	92.63
Urban attractions	100	89.94

Table 4 analysis indicates that the prediction accuracy of the strategy model for cultural, natural, and urban attractions is 93.15%, 92.63%, and 89.94%, respectively, for the same number of attractions. With regard to the adaptability of the RFPAP+FP-Growth model, it is evident that the model demonstrates excellent adaptability, enabling it to adjust prediction strategies in accordance with the distinctive characteristics of different types of attractions. This allows for the provision of personalized self-service travel suggestions for tourists. In terms of robustness, the RFPAP+FP-Growth model is capable of maintaining stable predictive performance in a variety of scenarios. The results of the RFPAP+FP-Growth model in the personalized self-service travel process are presented in Table 5.

Table 5: Feedback results of RFPAP+FP-Growth model in personalized self-service travel process

Feedback metrics	Prediction accuracy/%	User Satisfaction/Score	Average feedback time/day	Repeated booking rate/%	Complaint rate/%	Recommendation rate/%
RFPAP+FP-Growth model	90.67	8.7	2.8	75.92	2.08	86.15

In Table 5, the model exhibits an accuracy of 90.67% in predicting user preferences and travel needs, while concurrently achieving a high user satisfaction score of 8.7. Users provided feedback on the model's recommendations with a median time of 2.8 days. Conversely, the model's complaint rate is a mere 2.08%, indicating its high reliability and stability. In conclusion, the RFPAP+FP-Growth model demonstrates efficacy in PSGT planning, consistently providing users with accurate and satisfactory personalized recommendations.

5 Discussion

The preceding research indicates that the RFPAP model prioritizes user behavior analysis and personalized recommendations, whereas the FP-Growth model emphasizes the extraction of frequent item sets and association rules. The integration of the two models allows for the analysis of the correlation between services while considering the user's personalized needs, thereby enhancing the comprehensiveness of the recommendation system. In self-guided tours, RFPAP is based on users' historical recommendations, while FP-Growth analyzes destination and activity associations to discover users' potential points of interest and provide accurate recommendations. The discrepancies in data can influence the outcomes of the models, including the user behavior, destination, and activity data. When integrating models, it is essential to ensure the effective integration of different data sources and features to guarantee optimal performance. In the field of research pertaining to the development of personalized self-service travel strategy models, the innovative integration of the RFPAP model and the FP-Growth model has the potential to bring about revolutionary progress in the realm of personalized services within the tourism industry. The RFPAP model provides tourists with comprehensive resource information by employing scientific analysis of the mechanical properties and evolutionary laws of tourism resources. The FP-Growth model employs big data mining techniques to accurately capture the travel preferences and consumption habits of tourists. The combination of the two not only addresses the shortcomings of traditional methods in terms of the provision of personalized services, but also offers tourists a more accurate and efficient travel experience through the precise evaluation of tourism resources, the provision of personalized recommendations regarding tourism plans, and the optimization of tourism services. This integration method not only reflects the scientific and progressive nature of the approach, but also meets the needs of the personalized tourism era, injecting new vitality into the development of the tourism industry.

6 Conclusion

The study aims at the problem of ignoring the user's personal preference and the low accuracy of the feature selection of tourist destinations in traditional personalized recommendation algorithms. The study constructed the SM for PSGT by integrating the RFA preference model and FP-Growth association model. Through the construction of the SM, it is expected to formulate a PSGT recommendation strategy for tourists that is more in line with their needs, and to improve the travel experience and satisfaction. The results indicated that the mean value of support, mean value of confidence and mean value of enhancement of SM in attraction associative relationship research were 0.183, 0.597 and 1.093, respectively. The recommended predictive value was 82.17% and the feasibility predictive value was 65.94%, which were all small gaps from the true values, which indicated that all the SM constructed in the study had a strong practicality. The method not only significantly improves the PS capability of attractions, but also enhances the mining capability of correlations, thus improving the formulation and achievement of PSGT strategies. Although the study achieved good results, there are still some shortcomings. The dataset used in the study is small, covering only a limited sample of self-guided tour destinations and users. The next step of the study can consider expanding the size of the dataset to improve the accuracy and generalization ability of the model.

References

[1] Ravinder Jangra, S. P. Kaushik, and Surjit Singh Saini. An analysis of tourist’s perceptions toward tourism development: study of cold desert destination. Geography and Sustainability. 2(1):48-58, 2021. https://doi.org/10.1016/j.geosus.2021.02.004.

[2] I Gusti Bagus Rai Utama, Christimulia Purnama Trimurti, Jaya Pramono, and Ni Putu Dyah Krismawintari. Tourist behavior destination selection model. International Journal of Advanced Science and Technology. 29(5):7352-7375, 2020. https://doi.org/orcid.org/0000-0003-0517-5355.

[3] Miriam Edith Pérez-Romero, Martha Beatriz Flores-Romero, and Víctor G. Alfaro-García. Tourism and destination competitiveness: an exploratory analysis applying the forgotten effects theory. Journal of Intelligent & Fuzzy Systems.

40(2):1795-1804, 2021. https://doi.org/10.3233/JIFS-189186.x.

[4] Hamam Mokayed, Tee Zhen Quan, Lama Alkhaled, and V. Sivakumar. Real-time human detection and counting system using deep learning computer vision techniques. Artificial Intelligence and Applications. 1(4):221-229, 2023. https://doi.org/10.47852/bonviewAIA2202391.

[5] Weiguang Xie, and Yunchi Ma. Tourism resource evaluation and countermeasures based on network communication and topsis algorithm. Wireless Communications & Mobile Computing. 21(5):1-13, 2021. https://doi.org/10.1155/2021/8172282.

[6] Bin Deng, Jun Xu, and Xin Wei. Tourism destination preference prediction based on edge computing. Mobile Information System. 21(2):1-11, 2021. https://doi.org/10.1155/2021/5512008.

[7] Shengsha Xu. An apriori algorithm to improve teaching effectiveness. International Journal of Performability Engineering. 16(5):792-799, 2020. https://doi.org/10.23940/ijpe.20.05.p13.792799.

[8] Bruno Miguel Vieira, Ana Pinto Borges, and Elvira Pacheco Vieira. The role of social networks for decision-making about tourism destinations. International Journal of Internet Marketing and Advertising. 18(1):1-27, 2023. https://doi.org/10.1504/ijima.2023.10053065.

[9] Miftachul Huda. Digital marketplace for tourism resilience in the pandemic age: voices from budget hotel customers. International Journal of Organizational Analysis. 31(1):149-167, 2022. https://doi.org/10.1108/IJOA-10-2021-2987.

[10] Huang-Ping Yen, Po-Chi Chen, and Kung-Cheng Ho. Analyzing destination accessibility from the perspective of efficiency among tourism origin countries. SAGE Open. 11(2):1-14, 2021. https://doi.org/10.1177/21582440211005752.

[11] Jianhui Liang. Problems and solutions of art professional service rural revitalization strategy based on random forest algorithm. Wireless Communications and Mobile Computing. 22(1):1-11, 2022. https://doi.org/10.1155/2022/9752512.

[12] Heqing Zhang, Tingting Guo, and Xiaobo Su. Application of big data technology in the impact of tourism E-commerce on tourism planning. Complexity. 21(13):1-10, 2021. https://doi.org/10.1155/2021/9925260.

[13] Wolfram Höpken, Marcel Müller, Matthias Fuchs, and Maria Lexhagen. Flickr data for analysing tourists’ spatial behaviour and movement patterns: a comparison of clustering techniques. Journal of Hospitality and Tourism Technology. 11(1):69-82, 2020. https://doi.org/10.1108/JHTT-08-2017-0059.

[14] Lior Shabtay, Philippe Fournier-Viger, Rami Yaari, and Itai Dattner. A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. Information Sciences. 553(2):353-375, 2021. https://doi.org/10.1016/j.ins.2020.10.020.

[15] Yanling Dong, Shufen Zhang, Jingcheng Xu, Haoshi Wang, and Jiqiang Liu. Random forest algorithm based on linear privacy budget allocation. Journal of Database Management. 33(2):1-19, 2022. https://doi.org/ 10.4018/JDM.309413.

[16] Krit Sriporn, Cheng-Fa Tsai, and Paohsi Wang. Analyzing tourist behavior in virtual museums using intelligent approach with feature selection. Journal of Information Science and Engineering. 38(3):681-695, 2022. https://doi.org/10.6688/JISE.202205_38(3).0011.

[17] Sérgio Moro, and Joaquim Esmerado. An integrated model to explain online review helpfulness in hospitality. Journal of Hospitality and Tourism Technology. 12(2):239-253, 2021. https://doi.org/10.1108/JHTT-01-2020-0026.

[18] Jiaman Ding, Cuihua Liu, Runxin Li, Jinguo You, and Lianyin Jia. An improved deep forest classification algorithm. International Journal of Modeling Identification and Control. 40(4):305-314, 2022. https://doi.org/10.1504/IJMIC.2022.125548.

[19] Hong Li, Man Qiao, and Shuai Peng. Research on the recommendation algorithm of rural tourism routes based on the fusion model of multiple data sources. Discrete Dynamics in Nature and Society. 22(6):1-10, 2022. https://doi.org/10.1155/2022/2262148.

[20] Jiaqi Zhang. A novel recommendation algorithm based on multi-objective evolution. Applied and Computational Engineering. 22(6):29-34, 2023. https://doi.org/10.54254/2755-2721/2/20220524.

[21] Duaa Hamed AlSaeed. LOCUS: a mobile tourism application and recommender system for personalized places and activities. Informatica. 47(2), 2023. https://doi.org/10.31449/inf.v47i2.4351.

[22] Zakria Mahrousa, Dima Mufti Alchawafa, and Hasan Kazzaz. Frequent itemset mining based on development of fp-growth algorithm and use mapreduce technique. Association of Arab Universities Journal of Engineering Sciences. 28(1):83-98, 2021. https://doi.org/10.33261/JAARU.2021.28.1.008.

[23] Indra Griha Tofik Isa. Sales strategy by determining tenant position at Sriwijaya state polytechnic through association rules analysis. International Journal of Entrepreneurship and Business Development. 5(3):563-573, 2022. https://doi.org/10.29138/ijebd.v5i3.1855.

[24] Jianzhou Cui. Tourism English based on association rules mining algorithm. ICAIIS 2021: 2021 2nd International Conference on Artificial Intelligence and Information Systems. 28(6):1-5, 2021. https://doi.org/10.1145/3469213.3472785.

[25] Houria Laaroussi, Fatima Guerouate, and Mohamed Sbihi. Deep learning approach and topic modelling

for forecasting tourist arrivals. International Journal of Electrical and Computer Engineering Systems. 14(4):411-419, 2023. https://doi.org/10.32985/ijeces.14.4.5.

[26] Liyan Shen. Analysis of immersive virtual reality tourism resources based on navier-stokes equations. Informatica. 48(6), 2024. https://doi.org/1031449/inf.v48i6.5514.

[27] Zhaoyu Wang. Ecological protection design strategy for mountain tourism highway slopes. Journal of World Architecture. 5(3):9-12, 2021. https://doi.org/10.26689/jwa. v5i3.2176.

[28] U. Umbetov, A. B. Shinykulova, and G. S. Morokina. Algorithm of solving goal in tourism for hierarchical management with limited necessary resources. Herald of Science of S Seifullin Kazakh Agro Technical Research University. 1(108):190-201, 2021. https://doi.org/10.51452/kazatu.2021.1(108).155.

[29] Purwanto Purwanto, Retnowati Retnowati, and Hari Suryanto. Strategy for enhancing community economy through optimization of tourism areas (a study on rawa pening lakes in central java province-indonesia). Journal of International Conference Proceedings. 3(1):183-193, 2020. https://doi.org/10.32535/jicp.v2i4.794.

[30] Córdoba Mariano, and Balzarini Mónica. A random forest-based algorithm for data-intensive spatial interpolation in crop yield map. Computers and Electronics in Agriculture. 184(6):1-9, 2021. https://doi.org/10.1016/j.compag.2021.106094.