

# Attraction Recommendation and Itinerary Planning for Smart Rural Tourism Based on Regional Segmentation

Ruiping Chen<sup>1\*</sup>, Yanli Zhou<sup>2</sup>, Dejun Zhang<sup>3</sup>

Compliance Office, Foshan Polytechnic, Foshan, 528137, China<sup>1</sup>

School of Culture Tourism and Creativity, Foshan Polytechnic, Foshan, 528137, China<sup>2,3</sup>

**Abstract**—As the rural tourism industry develops, effective attraction recommendations and planning are crucial for the tourist experience. Then, a rural scenic spot tourism recommendation and planning technology based on regional segmentation was proposed. The scenic area was divided into multiple grids based on tourist check-in behaviour, and the interest and influence of the scenic area were associated with the grid check-in behaviour. Content recommendation was achieved through two factors: popularity and regional location. And considering the sparsity of data in the recommendation, clustering algorithms were introduced to model tourist check-in behaviour based on factors such as time and regional location, and content recommendation was achieved through tourist preferences. In the performance analysis of recommendation models, the proposed model has an accuracy of 0.965 and 0.956 on the Gowalla and Yelp datasets, respectively, which is superior to other models. Comparing the recommendation loss performance of different models, the proposed model has an RMSE loss of 0.120 on the Gowalla dataset, which is superior to other models. In practical application analysis, when the recommended number is 5, the accuracy and recall of the proposed model are 0.138 and 0.069, respectively, which are superior to other models. In tourism itinerary planning, the overall planning time of the model is the shortest. Therefore, the proposed model has excellent application effects, and the research content provides important technical references for tourist travel and rural tourism destination planning.

**Keywords**—Regional division; trip planning; recommended tourist attractions; clustering algorithm; time factor

## I. INTRODUCTION

According to data released by the National Tourism Administration, the average annual growth of rural tourism tourists has exceeded 10%, becoming an important industry supporting local economic development. However, traditional recommendation techniques mainly rely on user historical preferences and ratings to make recommendations, ignoring the changing interests of tourists at different times and in different regions [1]. In addition, due to the relatively small amount of data on rural tourist attractions and the sparsity of data, traditional recommendation systems couldn't meet the tourism needs of tourists [2]. The existing tourism recommendation methods may ignore user preferences, and their recommendation content may be inaccurate. To address this issue, Tourist Sign-in Area Segmentation (TSAS) has been proposed. This technology uses the check-in data of

tourists to divide the scenic area into multiple grids and associates the interest and influence of the scenic area by analyzing the check-in behaviour of tourists [3]. In addition, to address data sparsity, this technology introduces clustering algorithms to model the check-in behaviour of tourists and recommends content based on their preferences, improving the recommendation accuracy. There are several innovations in the recommendation technology studied. Firstly, it combines geographical location and time factors to achieve more accurate recommendations and planning of rural tourism attractions. Through in-depth analysis of tourist check-in behaviour, this technology can accurately capture the interests and preferences of tourists and provide personalized recommendations and planning solutions based on factors such as time and geographical location. The research content will provide reference for recommending rural tourism attractions and tourist itinerary planning and accelerate the development of the rural tourism industry.

The research content includes four sections. Introduction is given in Section I, Related works is given in Section II. The construction of rural tourism recommendation and itinerary planning model based on regional segmentation is given in Section III. Section IV apply the mentioned technology to specific scenarios and verify the effectiveness of the proposed recommendation model in practical scenarios. Finally, Section V gives the summary and analyses of the entire article are conducted, and the direction of technological improvement in the future is elaborated.

## II. RELATED WORKS

Recommendation technology is mainly used to solve information data overload, which can help users find suitable information content faster and more accurately. At present, recommendation technology has been widely applied in various fields, and relevant scholars have conducted extensive research on it. Cui Z et al. found that traditional recommendation systems may overlook the inherent relationship between user preferences and time. To address this problem, a new fusion recommendation model based on time correlation coefficients was proposed. This model further improved the accuracy and efficiency of recommendations by clustering similar users together. In addition, the study also proposed a personalized recommendation model based on preference patterns, mainly analyzing user behaviour to optimize content recommendation. The effectiveness of the

proposed model was validated using two datasets, MovieLens and Douban. Compared to other models, the overall recommendation performance of the proposed model was better [4]. Zhou X et al. focused on modeling and analyzing patient doctor generated data using an ensemble-based deep learning framework. So a fusion extraction model was proposed in the study, which could extract and highlight semantic information in patient inquiries. Then, the study proposed an intelligent recommendation method that refined the learning process through clustering mechanisms, providing patients with automatic clinical guidance and diagnostic recommendations. The accuracy of online patient queries could be effectively improved by applying the proposed technology to specific scenarios [5]. Cho J et al. focused on the impact of recommendation algorithms on user opinions on the video sharing platform YouTube. Therefore, traditional recommendation models were improved by processing information based on experimental results and filtering un-healthy content information. Through testing, the researched technology had better recommendation performance in practical scenarios and could filter out the impact of harmful information on users [6].

Interest-based recommendation technology has a wide range of applications in tourism services and other fields. Interest-based recommendation technology focuses more on factors such as user preferences and behavioural habits, which is closer to the actual needs of users. Nitu P et al. conducted research on tourism recommendation technology based on social media activities. To improve the recommendation effect, personalized recommendations of the model were achieved by analyzing user Twitter data as well as research friend and follower data to identify travel-related tweets. Time-sensitive nearest degree weights were introduced to improve recommendation accuracy. The proposed technology applied to practical tourism recommendation scenarios had excellent recommendation performance, which was superior to other recommendation techniques [7]. Giglio S et al. conducted research on urban tourism recommendation technology. Clustering analysis was used to collect and analyze image data from multiple cities in Italy to improve the recommendation accuracy of the model, and Wolfram Mathematica was used to automatically identify clusters around points of interest. New tourism scenarios and more information for the interest point recommendation process could be provided by applying this technology to tourism recommendation scenarios, which was superior to other recommendation models [8]. Huang F et al. found that existing tourism route planning methods were mainly targeted at specific tasks and couldn't be applied to other tasks. To address this issue, a multi-task deep travel route planning framework was proposed, which integrated rich auxiliary information to construct a flexible planning model. These results confirmed that this method exhibited flexibility and superiority in travel route planning, outperforming relevant recommendation models [9]. Wang et al. focused on the importance of location in recommendation systems. The study proposed a multi-objective recommendation framework based on location and preference perception, modeling location-based recommendations as a multi-objective optimization problem. The study considered the performance of recommendation algorithms in recommending similar and

different items, and a new multi-objective evolutionary algorithm was proposed. These results confirmed that this model could generate better recommendation solutions and overcome data sparsity and cold start issues compared to other recommendation models, resulting in better overall recommendation performance [10].

In summary, recommendation technologies mainly analyze the feature associations between objects and targets to achieve effective content recommendation. The above studies have analyzed the application of recommendation technology in different fields and discussed its effectiveness based on interest points. However, existing research has problems such as neglecting changes in user dynamic interests, insufficient understanding of user behaviour at a deeper level, and insufficient explanation of recommendation systems. Therefore, a tourism recommendation technology based on regional segmentation is proposed to provide important technical support for the development of the tourism industry and the promotion of tourism destinations.

### III. CONSTRUCTION OF RURAL TOURISM RECOMMENDATION AND ITINERARY PLANNING MODEL BASED ON REGIONAL SEGMENTATION

This section proposes a recommendation technique based on regional segmentation to segment rural areas and establish a recommendation model based on attendance. Simultaneously considering factors such as data sparsity and tourist interest transfer, the recommendation technology is improved and modeled based on time characteristics.

#### A. Construction of a Recommendation Model Based on Tourist Check-in Area Segmentation

In recent years, the rural tourism industry has flourished, with a large number of tourists flocking to rural tourist attractions, promoting the development of the rural economy. To meet the personalized tourism needs of tourists, accurate recommendation of tourist attractions is crucial for improving the quality of travel. A recommendation method based on the division of tourist check-in areas has been proposed [11]. This method mainly considers that tourists will sign in and clock in when visiting the scenic area, share on their respective social circles or social circles, and use Location-based Social Network (LBSN) data information to mine tourist behaviour data. Segmentation is carried out according to the check-in area, and multiple areas are delineated to achieve content recommendation based on the size of regional influence [12]. The proposed TSAS method can accurately capture the interests and preferences of tourists by conducting in-depth analysis of their check-in behaviour in different regions. Unlike traditional recommendation systems, TSAS comprehensively considers geographical location and time, enabling recommendation systems to provide more timely and regional recommendations based on the geographical location and different periods of tourists, enhancing the personalization of recommendations. The relatively small and sparse data for rural tourism attractions can be addressed by using clustering algorithms to model the check-in behaviour of tourists, and the efficiency of recommendation systems in utilizing limited data is improved. Fig. 1 shows the framework of the entire tourism recommendation system.

Fig. 1 shows the framework of the entire tourism recommendation system, which implements content recommendation by mining feature data of tourists and rural tourism attractions and ranking them based on personalized feature influence. The preferred method is to obtain check-in information for rural tourism areas from LBSN data and segment the area based on the dimensions of the check-in area to obtain multiple small grid areas [13]. Fig. 2 is a schematic diagram of segmenting rural scenic spots.

According to Fig. 2, each small grid area contains the check-in information of tourists, which gathers various check-in interest points, and the characteristics of interest points between different grids are not the same. The length and width of the entire rectangular area are defined as  $a$  and  $b$ . Two independent matrices need to be constructed after

dividing the rectangular area into multiple grids, namely the tourist activity area matrix  $X$  and the interest point area influence matrix  $Y$ . The matrix for a tourist  $u$  in the activity area is denoted as  $x_u$ . For some tourists who have checked in the grid, the probability of tourists appearing in the area will be greater than 0 [14]. The influence vector of a certain interest point  $l$  in the region matrix is set to  $y_l$ . The regional influence of scenic spots is mainly influenced by two factors: distance from surrounding locations and points of interest. The influence of interest point  $i$  on the network region  $l$  is represented by Eq. (1).

$$w_{ii} = p(i) \cdot \frac{1}{\sigma} K\left(\frac{d(i,l)}{\sigma}\right) \quad (1)$$

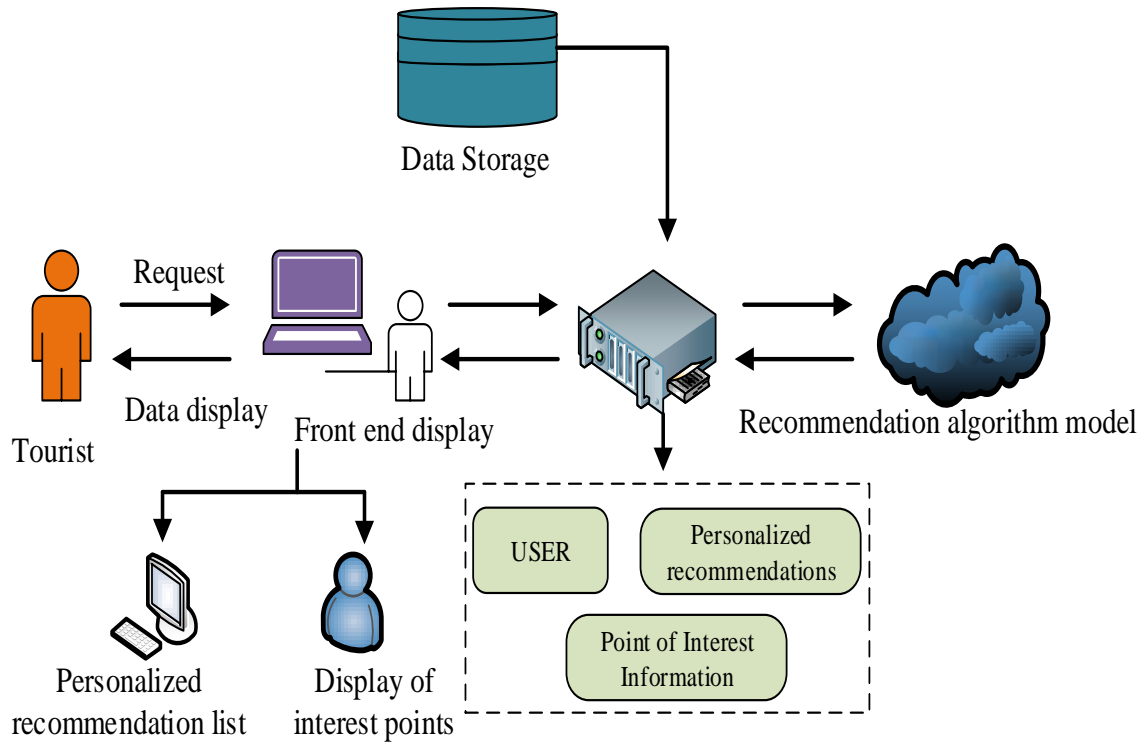


Fig. 1. Tourism recommendation system framework

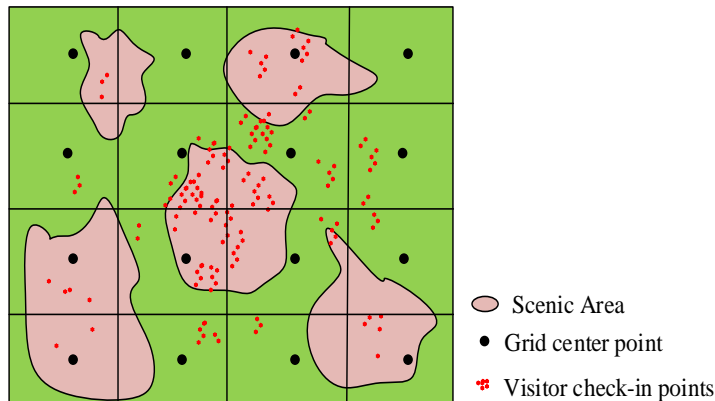


Fig. 2. Schematic diagram of rural tourist attractions segmentation.

In Eq. (1),  $d(i,l)$  is the distance from  $i$  to the center of the grid.  $p(i)$  represents the popularity of interest points.  $\sigma$  means the standard deviation.  $K(\cdot)$  is a normal distribution. The number of tourist check-in is used as the popularity of the region, and the check-in data are normalized using Eq. (2).

$$p(i) = 1 + \lg(1 + r_{ui} \times 10) \quad (2)$$

In Eq. (2),  $r_{ui}$  represents the cross-factor between geographic location and prevalence. Next, it is necessary to explore the relationship between the location and popularity of the region. The farther away the rural scenic spots are, the gradually decreasing influence can be considered. If they are too far away, the influence will be ignored. Taking the influence of interest points as a key consideration, the influence matrix  $\gamma$  of interest point regions is taken as the objective function, and a matrix decomposition model is used to solve it. The new interest point score is represented by Eq. (3).

$$\hat{r}_{ul} = p_u \cdot q_l + x_u \cdot y_l \quad (3)$$

In Eq. (3),  $q_l$  represents the matrix of interest points.  $p_u$  is the implicit vector of tourists. In practical recommendations, tourists are easily influenced by social circle factors, and similar preferences between tourists and friends can easily lead to the final target selection. Therefore, it is necessary to calculate the similarity between them [15], which is represented by Eq. (4).

$$s(u,v) = \theta \times \hat{F}_{uv} + (1-\theta) \times \frac{|F_u \cap F_v|}{|F_u \cup F_v|} \quad (4)$$

In Eq. (4),  $\theta$  is the adjustment parameter.  $\hat{F}_{uv}$  means the friend relationship judgment.  $F_u$  represents a collection of tourist friends.  $F_v$  represents the collection of friends of user  $v$ . The objective loss function of the TSAS model is obtained by integrating the regional division, tourist social

factors, tourist activity factors, and interest point influencing factors into the traditional matrix factorization model, which is represented by Eq. (5).

$$\lim_{P,Q,X} \sum_{(u,L) \in (U,L)} (r_{ul} - \hat{r}_{ul})^2 + \frac{\lambda_1}{2} (\|p_u\|^2 + \|q_l\|^2) + \frac{\lambda_2}{2} \|x_u\|^2 + \frac{\lambda_3}{2} \sum_{v \in F_u} s(u,v) \times \|p_u - p_v\|^2 \quad (5)$$

In Eq. (5),  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the weights controlled by three factors.  $p_v$  is the popularity of the user's area.  $r_{ul}$  represents the actual point of interest score.  $Q$  represents an implicit vector of interest points.  $p$  represents the implicit vector of tourists. In order to improve the training effect of the objective function, the gradient descent method is used to optimize the parameters of the objective function. The gradient of  $p_v$  is represented by Eq. (6).

$$\frac{\partial \Theta}{\partial p_v} = -2(r_{ul} - \hat{r}_{ul}) \cdot s(u,v) \cdot q_l - \lambda_3 \cdot s(u,v) \cdot (p_u - p_v) \quad (6)$$

In Eq. (6),  $\frac{\partial \Theta}{\partial p_v}$  represents the gradient of  $p_v$ . The gradient of  $q_l$  is represented by Eq. (7).

$$\frac{\partial \Theta}{\partial q_l} = -2(r_{ul} - \hat{r}_{ul}) \cdot (p_u + \sum_{v \in F_u} s(u,v) p_v) + \lambda_2 q_l \quad (7)$$

In Eq. (7),  $\frac{\partial \Theta}{\partial q_l}$  represents the gradient of  $q_l$ . The gradient of  $x_u$  is represented by Eq. (8).

$$\frac{\partial \Theta}{\partial x_u} = -2(r_{ul} - \hat{r}_{ul}) \cdot y_l + \lambda_2 x_u \quad (8)$$

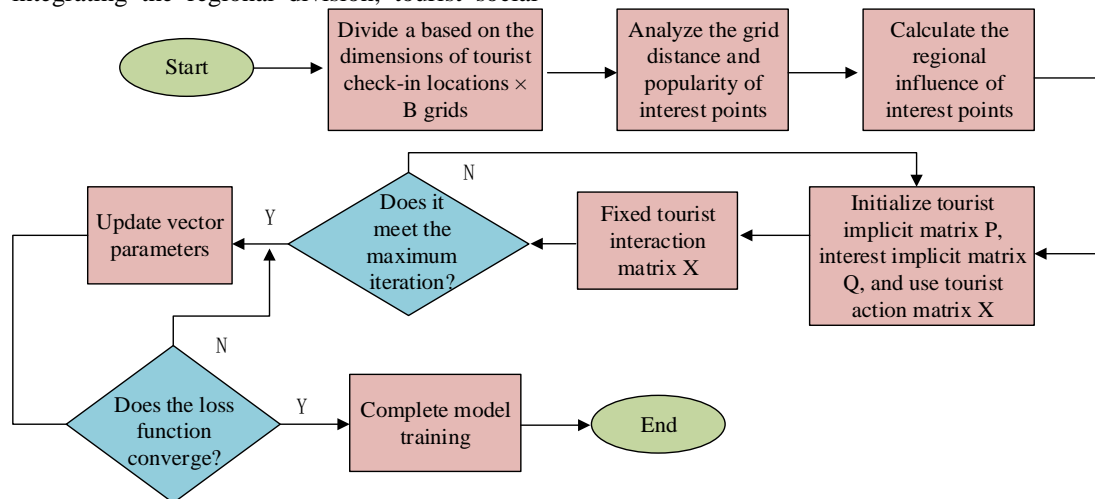


Fig. 3. Recommended process for rural tourist attractions.

By using Eq. (6) to Eq. (8), the gradient of each parameter can be obtained. TSAS continuously optimizes the parameters in each iteration until the model converges or reaches the maximum number of iterations, completing the model training. Fig. 3 shows the entire recommendation model.

**B. Construction of a Recommendation Model Based on Tourist Check-in Area and Time Factors**

In the construction of traditional interest point recommendation system models, it is impossible to avoid data sparsity and implicit feedback problems in the system. To avoid data sparsity, further mining can usually be done on regional geographic location, topic categories, time series, and tourist social information. However, there are hidden user behaviour patterns in tourist check-in data, and effectively extracting contextual information hidden in tourist check-in data is the key to improving model recommendation effectiveness [16]. Therefore, the TSAS recommendation model is improved by introducing a greedy clustering algorithm to search for tourist check-in center locations and divide them into different regions based on check-in points, analyzing the impact of different regions on tourist check-in interests. Meanwhile, the sequence of tourist interest points during a certain period is analyzed. By analyzing the time to reflect the transfer characteristics of tourist interest points during a certain period, the impact of time factor on tourist check-in can be obtained [17]. Greedy clustering method is used to partition and confirm the regions to find the center of each region in the sparse tourist check-in data. Fig. 1 shows its schematic diagram.

According to Fig. 4, the greedy clustering method is used to sort the check-in times of interest points. The region with the most check-in times is selected as the center, and the selected region center is scanned again, with the region less

than the distance  $d$  as the center point, and placed in the region set. If the current check-in reaches the set threshold ratio, the area will be divided, and the check-in times in the center of the high area will decrease towards the surrounding areas. The division of tourist check-in areas is made more reasonable by using the above methods [18]. The set of tourist check-in centers is defined as  $C_u$ , and the probability of tourists arriving at a given point of interest  $l$  is expressed using a central Gaussian model, represented by Eq. (9).

$$P(l | C_u) = \sum_{c_u=1}^{C_u} \frac{1}{dst(l, c_u)} \cdot \frac{f_{c_u}}{\sum_{i \in C_u} f_i} \quad (9)$$

In Eq. (9),  $\frac{1}{dst(l, c_u)}$  represents the distance from the point of interest to the center of the region.  $f_{c_u}$  represents the check-in frequency in different regional centers. The check-in of tourists is inversely proportional to their distance, with the closer they are, the more they check in. Therefore, the closer tourists are to the center of  $l$ , the higher the probability of check-in. In addition to analyzing the impact of check-in centers, it is also necessary to consider the influence of time factors on tourist interests. Therefore, the time proximity method is adopted to divide tourist check-in into implicit tourist vectors and implicit interest point vectors, and the product of the two factors is used to fit the rating prediction matrix [19]. If the probability of tourists checking in at  $l$  is defined as  $p(F_{ul})$ , then Eq. (10) can be obtained.

$$P(F_{ul}) \propto P^T Q \quad (10)$$

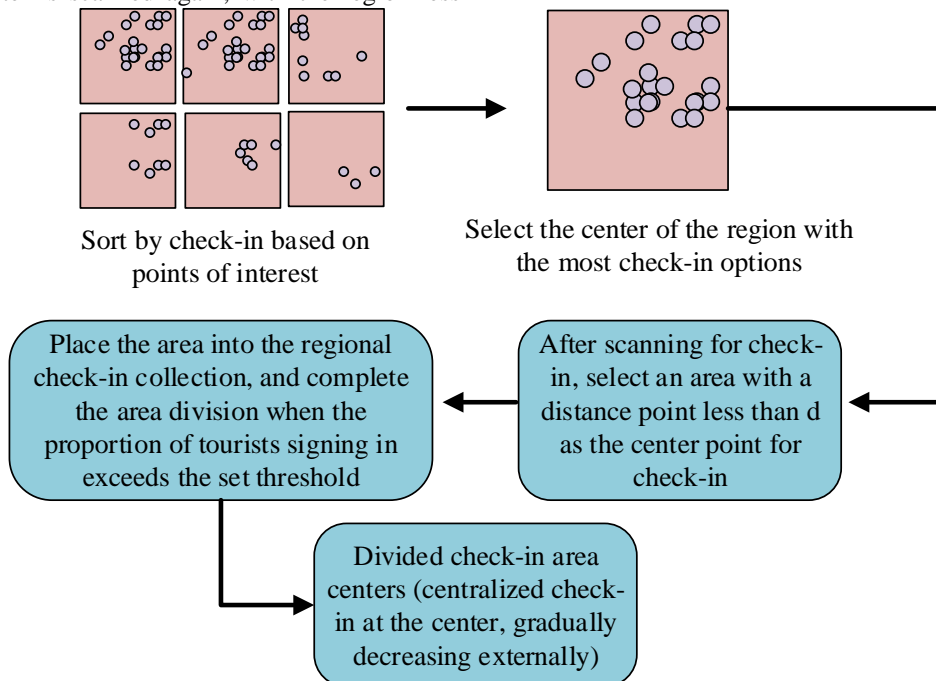


Fig. 4. Division of tourist check-in centers.

In Eq. (10),  $P^T$  represents the implicit matrix of tourists.  $\alpha$  is the fitting symbol. The check-in points of tourists have sequential characteristics in time, indicating that the check-in is influenced by time factors. For example, the check-in area at noon is concentrated in the restaurant area, and the check-in area in the morning or afternoon is concentrated in specific scenic areas. The  $k$  interest points of tourists who check-in within a certain period of time are analyzed to effectively analyze the check-in patterns of tourists during a certain period of time. These points are regarded as the current check-in records.  $N_k^T(l)$  is recorded as time neighbors. The implicit vectors of time neighbors are accumulated as the implicit vectors of tourist check-in interest points, represented by Eq. (11).

$$l^T = \frac{1}{k} \cdot \sum_{l' \in N_k^T(l)} l_r \quad (11)$$

Based on the above research, the transfer pattern of tourist interests during a certain period of time in Eq. (12) can be obtained.

$$P(F_{ul}) \propto (p_i \cdot q_j + p_i \cdot \frac{1}{k} \sum_{l' \in N_k^T(l)} l_r) \quad (12)$$

In Eq. (12),  $P(F_{ul})$  represents the probability of tourist check-in for fusion time transfer.  $q_j$  represents the implicit interest point  $j$ .  $p_i$  represents the implicit vector of user  $i$ . By using Eq. 12), the transfer pattern of tourists' interest at a

certain time can be obtained. In order to better improve the recommendation effect of the model, a probability matrix decomposition model is used to obtain the objective optimization function, represented by Eq. (13).

$$\min_{U,L} \sum_{i=1}^U \sum_{j=1}^L I_{ij} (h(F_{ij}) - h(p_i^T q_j))^2 + \lambda_1 \|p_i\|_F^2 + \lambda_2 \|q_j\|_F^2 \quad (13)$$

In Eq. (13),  $U$  represents the implicit matrix of tourists.  $L$  means the implicit matrix of interest points.  $I_{ij}$  is the attendance record of user  $i$  at the point of interest  $j$ .  $h(\cdot)$  refers to the logistic function.  $F_{ij}$  represents the attendance status of user  $i$  for interest point  $j$ .  $F$  is a set of quantities. The time-transfer characteristics of tourists are integrated with the check-in modeling features to obtain the probability value of the final tourist  $u$  at interest point  $l$ , represented by Eq. (14).

$$P_{ul} = P(F_{ul}) \cdot P(l | C_u) \quad (14)$$

Recall (R) and Precision (P) are introduced to effectively evaluate the practical application effect of the recommendation model, represented by Eq. (15).

$$\begin{cases} recall @ k = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{|S_u(k) \cap V_u|}{|V_u|} \\ presion @ k = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{|S_u(k) \cap V_u|}{k} \end{cases} \quad (15)$$

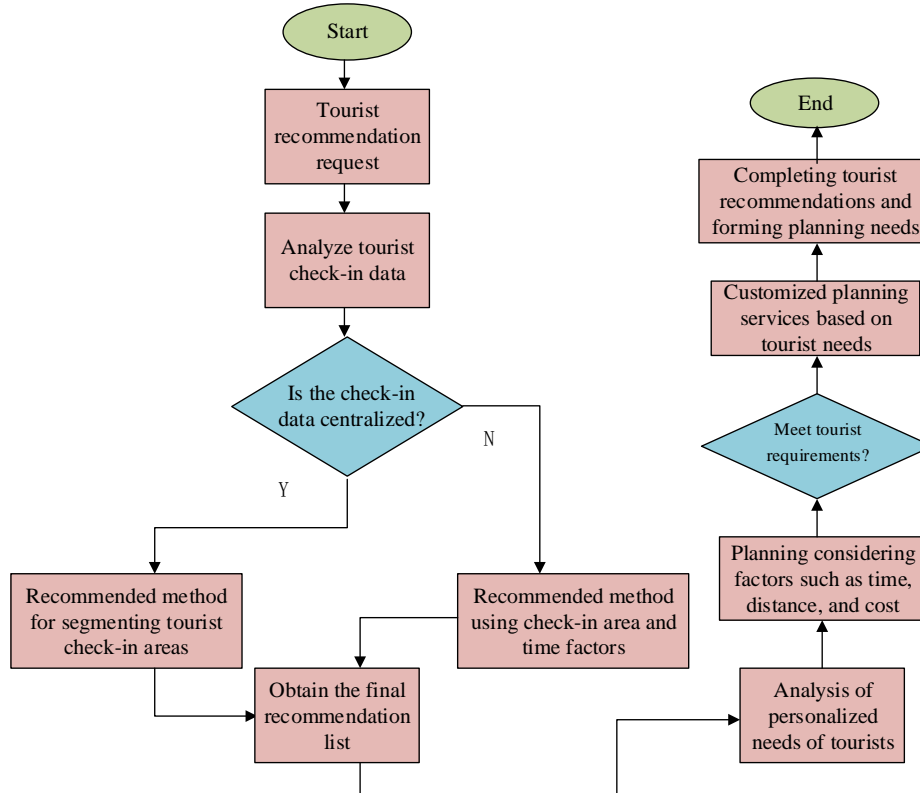


Fig. 5. Construction process of rural tourism attraction recommendation system.

In Eq. (15),  $k$  represents the number of recommended points of interest.  $S_u(k)$  indicates the interest points recommended by the top  $k$  tourists.  $V_u$  is a collection of interest points that tourists truly visit. High R and P indicate high accuracy of the recommendation. Fig. 5 shows the entire model construction process. In addition, the constructed recommendation system will fully take into account the distance and time factors of tourists, which provides more suitable rural tourism attractions for tourists. In a recommendation system, the main factor options to consider will be set, including time distance, cost, and comprehensive experience factors. Tourists prioritize experience and will overlook factors such as time, distance, and cost. Their overall planning focuses more on experience and functionality. In terms of time distance, more attention is paid to the travel time and distance, while also taking into account the experience [20]. Cost is mainly considered based on cost-effectiveness, taking into account factors such as distance and time experience, to meet the quality of tourist experience as much as possible while reducing the cost of the visitor. Fig. 5 shows the construction process of the entire rural tourism attraction recommendation system.

#### IV. SIMULATION TESTING OF ALGORITHMS

This section consists of two parts: model performance and practical scenario application analysis. The performance analysis part mainly tests the performance of the model on a universal dataset. In the actual scenario, specific rural tourism data are selected for training to test the application effect of different models in rural tourism attraction recommendation.

##### A. Performance Analysis of Rural Tourism Recommendation Models

Experimental tests were conducted on the WINDOWS 10 64 bit platform to test the performance of the proposed rural tourism recommendation model, with a running memory of 64GB, an Intel i9 16 core processor, and a graphics card NVIDIA RTX4080. Simulation experiments were performed on the Matlab platform for analysis. The Gowalla and Yelp datasets were selected for the experiment. Gowalla has 32510 points of interest and 18737 users. Yelp has 30887 users and 18995 points of interest. Singular Value Decomposition (SVD) and Probability Matrix Factorization (PMF) models were introduced as recommended testing benchmarks. In actual testing,  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are important weight parameters that affect the training of the proposed model. Therefore, it is necessary to select appropriate regularization parameters for testing. Root Mean Squared Error (RMSE) was used to reflect

the results in Fig. 6.

In Fig. 6,  $\lambda_1$  is mainly responsible for weighting the implicit vectors of tourists and interest points. In Gowalla, when  $\lambda_1$  was 0.5, RMSE was the lowest. In Yelp, when  $\lambda_1$  was 0.3, RMSE was the lowest. Overall, the analysis shows that Yelp is relatively sparse, and the model performs best when the dataset is sparse with a  $\lambda_1$  of 0.3.  $\lambda_2$  mainly affects the weights in the tourist activity matrix. Through experimental analysis, in Gowalla, the best model performance was achieved when  $\lambda_2$  was 0.3. When the Yelp was sparse, the best model performance was achieved when  $\lambda_2$  was 5.  $\lambda_3$  is a parameter that controls the social weight of tourists. In Gowalla and Yelp, the best performance was achieved when  $\lambda_3$  was 0.3 and 0.6, respectively. Therefore, in subsequent experiments, effective weights are set based on the sparsity of the test samples to ensure the testing performance of the model. Meanwhile,  $\theta$  represents the similarity adjustment parameter, which also has a direct impact on model testing in Fig. 7.

In Fig. 7, regardless of whether the dataset was sparse or not,  $\theta$  had no significant impact on the performance of the model. When  $\theta$  was 0.5, the model had the best testing performance. Therefore, based on the above experimental results, appropriate parameter values were selected for comparison. Fig. 8 shows the comparison results of recommendation accuracy between different models.

According to the results in Fig. 8, in Gowalla, the proposed model achieved the earliest convergence and had the highest recommendation accuracy of 0.965, while the PMF and SVD recommendation models were 0.912 and 0.946, respectively. Meanwhile, in Yelp, the recommended accuracy of the proposed model, SVD, and PMF was 95.65, 0.832, and 0.795, respectively. When the dataset is sparse, the recommendation performance of PMF and SVD significantly decreases, while the proposed model still has excellent recommendation performance. Fig. 9 compares the errors of two models.

According to Fig. 9, in Gowalla, the RMSE loss of PMF, SVD, and the proposed model towards convergence was 0.425, 0.335, and 0.120, respectively. In Yelp, when PMF, SVD, and the proposed model tended to converge, the RMSE loss was 0.865, 0.432, and 0.132, respectively. The proposed model has lower overall RMSE loss and better performance.

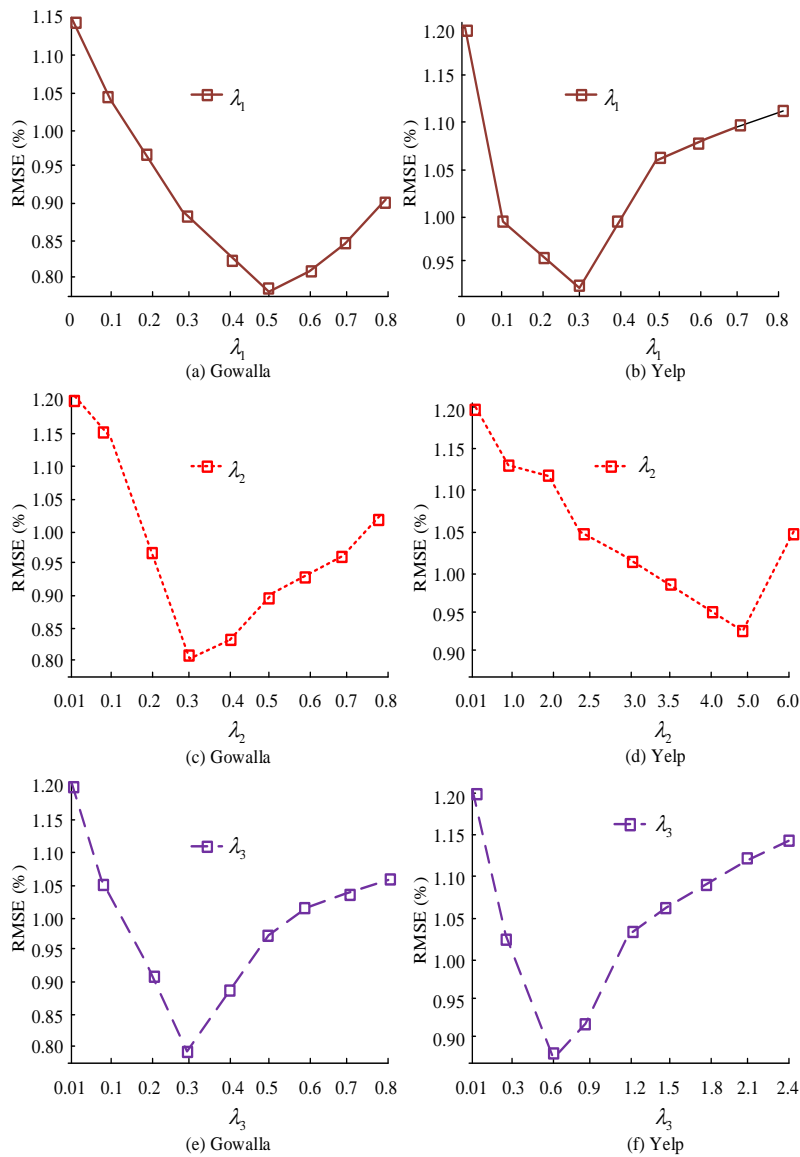


Fig. 6. Comparison of different weight parameters.

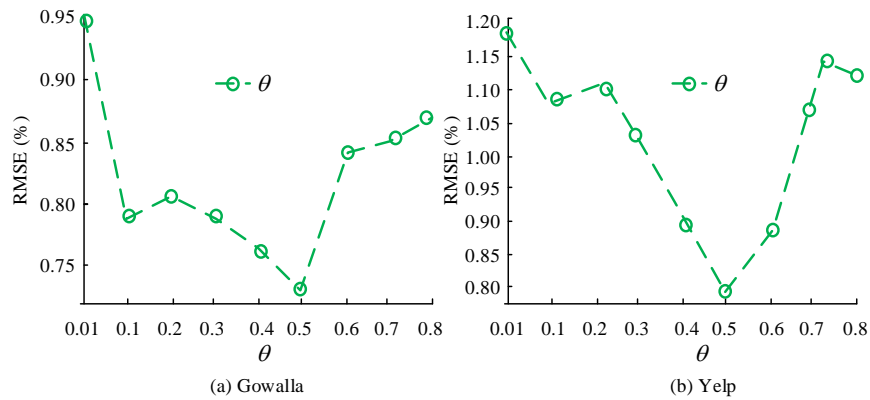


Fig. 7. Comparison of different similarity adjustment parameters.



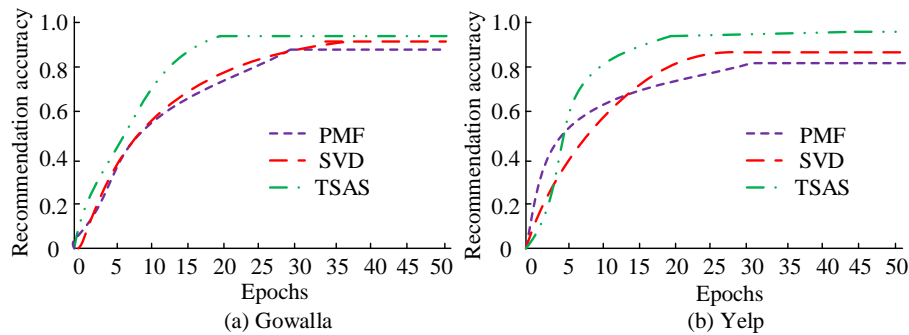


Fig. 8. Comparison of recommendation accuracy among different recommendation models.

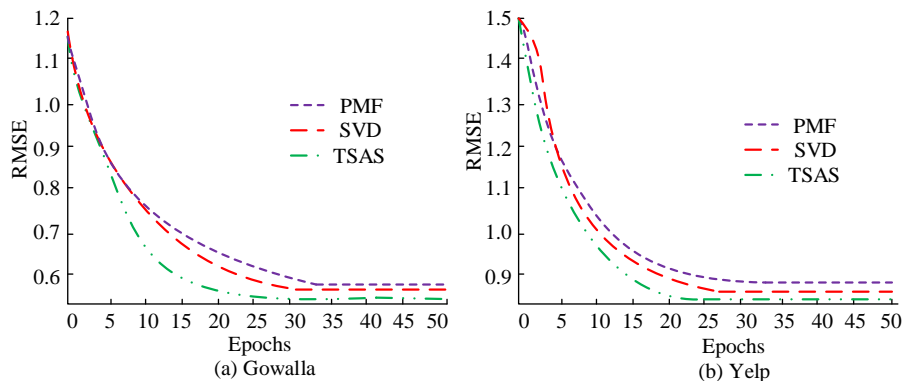


Fig. 9. Comparison of error performance of different recommendation models.

**B. Analysis of Practical Application Scenarios of Tourism Recommendation Models**

Crawler technology was used to crawl Ctrip tourist comment information, including 215654 rural tourism check-in score data, catering data, etc. Baidu Map platform was used to search for the longitude and latitude coordinates of rural tourist attractions, and the 8km range of scenic spots were classified into the same section. Finally, 289456 distance section data were collected, and the final regional feature data of rural scenic spots were obtained through sorting. Table I shows the specific parameters.

TABLE I. COLLECTION OF INFORMATION FOR RURAL TOURISM DESTINATIONS

Types of	Training set	Test set	Section data
File size	2.26M	2.41M	5.07M
Number of tourists	23156	29651	32546
Number of attractions	16581	19635	19465
Number of elements	144488	71165	289456
Minimum score/minimum height range	1	1	0
Maximum score/maximum pitch range	5	5	425
The proportion of the dataset	67.00%	33.00%	-

In Fig. 10, SVD and SVD models are still used as benchmark models, and the recommendation performance of the models in actual rural scenic spots is compared.

According to Fig. 10, when the number of recommendations was 5, all three models had the best recommendation performance. The accuracy values of PMF, SVD, and the proposed model were 0.098, 0.111, and 0.138, respectively, when the recommended quantity was 5. Simultaneously comparing the recall rates of different models, when the number of recommendations was 5, the recall rate of PMF, SVD, and the proposed model was 0.048, 0.051, and 0.069, respectively. The proposed model has better accuracy and recall performance than other recommendation models. Finally, Fig. 11 compares the itinerary planning effects of three models in rural tourism scenarios.

Fig. 11 shows the effect of travel itinerary planning for different models, which include three recommendation modes: time distance, cost, and comprehensive experience. Travel arrangements are planned according to the needs of the tourists. As tourists spent more time in rural scenic areas, the planning time of different models varied significantly. Among them, the overall planning time of the PMF model was the longest, with the highest planning time reaching 11200ms after the tourist travel time reached 330 minutes. The longest planning time for SVD was 6212ms. The best performing model is the proposed one. Although the planning time of the proposed model increased after the tourist's play time reached 330 minutes, the planning efficiency was still the highest compared with the other two models, and the longest planning time of the model was 1956 ms. Therefore, the proposed model has excellent rural tourism recommendation performance.

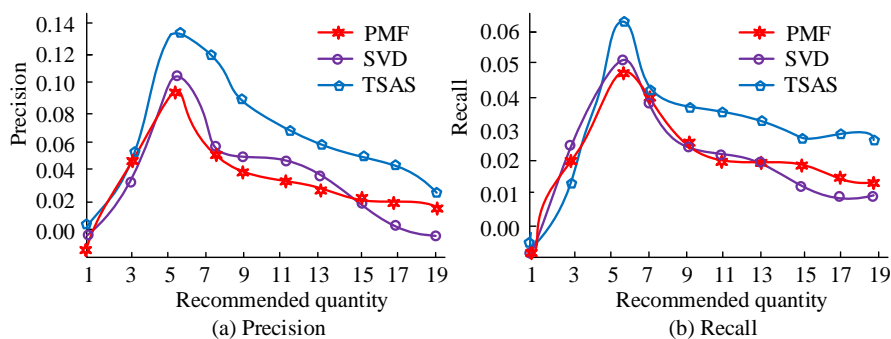


Fig. 10. Comparison of accuracy and recall performance of different models.

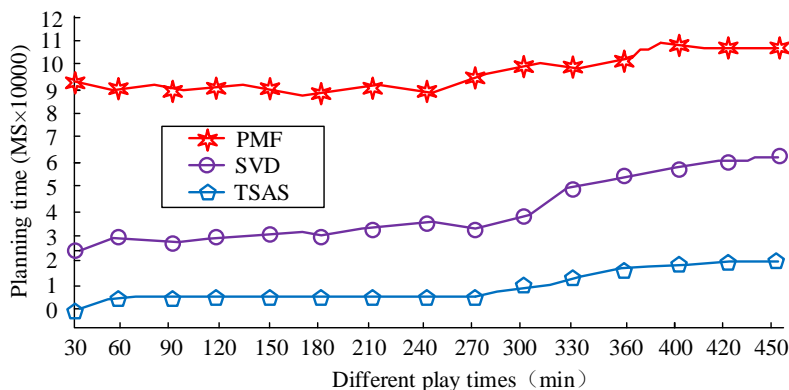


Fig. 11. Comparison of tourism itinerary planning efficiency among different models.

## V. CONCLUSION

Rural tourism has attracted a large number of tourists due to its unique culture and characteristics, but the recommendation of rural scenic spots has always faced difficulties and couldn't meet practical needs. A region segmentation-based recommendation technique is proposed for this purpose. Firstly, tourist check-in and geographical location are considered, and the check-in situation is used to reflect the process of segmenting regions, thereby achieving content recommendation. While in practical content recommendation, data sparsity and visitor interest transfer issues need to be considered as well. Therefore, the modeling is based on tourist check-in areas and time factors to capture the temporal changes of tourists. Finally, different itinerary planning schemes are matched according to the needs of tourists to achieve recommendations and itinerary planning for rural scenic spots. In the experimental analysis of model performance, the proposed model, SVD, and PMF models achieve recommendation accuracy of 0.956, 0.832, and 0.795 in Yelp, respectively. Meanwhile, the RMSE loss is 0.865, 0.432, and 0.132, respectively, when PMF, SVD and the proposed model tend to converge. In practical scenario application analysis, the optimal recall rate of the proposed model is 0.069, and the PMF and SVD are 0.048 and 0.051, respectively. Comparing the travel planning efficiency of different models, the highest time consumption of the proposed model is 1956ms, while PMF and SVD are 11200ms and 6212ms, respectively. Therefore, the proposed model has excellent recommendation and itinerary planning effects in rural tourism attraction recommendation. There are still

shortcomings in this study. The proposed method relies on tourist check-in data for recommendation. Although it introduces time-sensitive nearest weight, in some cases, the recommendation system may still not be able to fully capture the instantaneous interest changes of users. In the future, research technology also needs to consider regional meteorological factors, holidays and other factors, fully considering the impact of these factors on tourist recommendations, to optimize the practical application effect of recommendation technology.

## ACKNOWLEDGMENT

The research is supported by Research on Ecological Erosion and Cultural Ecological Protection of Traditional Villages in Guangdong Water Township: Key Research Platforms and Projects of General Universities in Guangdong Province - Special Projects in Key Fields (Science and Technology Services for Rural Revitalization) (2021ZDZX4069).

## REFERENCES

- [1] Wei G, Wu Q, Zhou M. A hybrid probabilistic multiobjective evolutionary algorithm for commercial recommendation systems. *IEEE Transactions on Computational Social Systems*, 2021, 8(3): 589-598.
- [2] Cao B, Zhang Y, Zhao J, Liu X. Recommendation based on large-scale many-objective optimization for the intelligent internet of things system. *IEEE Internet of Things Journal*, 2021, 9(16): 15030-15038.
- [3] Cao B, Zhao J, Lv Z, Yang P. Diversified personalized recommendation optimization based on mobile data. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(4): 2133-2139.

- [4] Cui Z, Xu X, Fei X U E, Cai X, Cao Y. Personalized recommendation system based on collaborative filtering for IoT scenarios. *IEEE Transactions on Services Computing*, 2020, 13(4): 685-695.
- [5] Zhou X, Li Y, Liang W. CNN-RNN based intelligent recommendation for online medical pre-diagnosis support. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020, 18(3): 912-921.
- [6] Cho J, Ahmed S, Hilbert M, et al. Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *Journal of Broadcasting & Electronic Media*, 2020, 64(2): 150-172.
- [7] Nitu P, Coelho J, Madiraju P. Improvising personalized travel recommendation system with recency effects. *Big Data Mining and Analytics*, 2021, 4(3): 139-154.
- [8] Giglio S, Bertacchini F, Bilotta E, et al. Machine learning and points of interest: typical tourist Italian cities. *Current Issues in Tourism*, 2020, 23(13): 1646-1658.
- [9] Huang F, Xu J, Weng J. Multi-task travel route planning with a flexible deep learning framework. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(7): 3907-3918.
- [10] Wang S, Gong M, Wu Y, Zhang M. Multi-objective optimization for location-based and preferences-aware recommendation. *Information Sciences*, 2020, 513(6): 614-626.
- [11] Al Farani K, Nafis F, Aghoutane B, et al. Hybrid recommender system for tourism based on big data and AI: A conceptual framework. *Big Data Mining and Analytics*, 2021, 4(1): 47-55.
- [12] Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, 2023, 31(6): 3360-3379.
- [13] Fan W, Ma Y, Li Q, Cai G. A graph neural network framework for social recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(5): 2033-2047.
- [14] Dhelim S, Ning H, Aung N, Huang R. Personality-aware product recommendation system based on user interests mining and metapath discovery. *IEEE Transactions on Computational Social Systems*, 2020, 8(1): 86-98.
- [15] Huang Z, Xu X, Zhu H, Zhou MC. An efficient group recommendation model with multiattention-based neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(11): 4461-4474.
- [16] Grossman G, Kim S, Rexer J M. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciences*, 2020, 117(39): 24144-24153.
- [17] Liang S. Research on Route Planning of Red Tourist Attractions in Guangzhou Based on Ant Colony Algorithm. *Automation and Machine Learning*, 2023, 4(1): 8-16.
- [18] Zhou X, Li Y, Liang W. CNN-RNN based intelligent recommendation for online medical pre-diagnosis support. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020, 18(3): 912-921.
- [19] Liu K, Sun Y, Yang D. The Administrative Center or Economic Center: Which Dominates the Regional Green Development Pattern? A Case Study of Shandong Peninsula Urban Agglomeration, China. *Green and Low-Carbon Economy*, 2023, 1(3), 110-120.
- [20] Ly A, El-Sayegh Z. Tire wear and pollutants: An overview of research. *Archives of Advanced Engineering Science*, 2023, 1(1): 2-10.