Optimized Unsupervised Semantic Trajectory Mining for Personalized Tourism Recommendations

Panpan Guo
School of Tourism Management, Zhengzhou Tourism College, Zhengzhou, 451464, China
Email: guogpp057@outlook.com

Keywords: unsupervised, multiple semantics, trajectory mining, intelligent tourism, personalized recommendations, multi granularity

Received: August 28, 2024

With the advent of the big data era, tourists receive a massive amount of travel information every day, and traditional personalized travel recommendation methods are difficult to provide personalized travel recommendation services. Therefore, the study first introduced a multi granularity recommendation framework based on attention mechanism, and then designed an unsupervised multi-implicit semantic trajectory mining model. The multi granularity recommendation framework was used for optimization to obtain more comprehensive user preferences. Finally, based on this, an intelligent personalized tourism recommendation method was established. The study used a dataset of user travel trajectory data collected and processed from various popular travel portals for subsequent testing, and evaluated the performance of different recommendation models from three aspects: computational efficiency, overall utility, and resource utilization. Accuracy, average training time F1 value, and resource utilization were selected to evaluate the performance of different recommendation models. The research results showed that compared with other algorithms, the recall and accuracy growth rates based on the optimized unsupervised multiple cryptic semantic trajectory mining model were the fastest, with corresponding recall and accuracy rates of 86.57% and 97.62%, respectively. Furthermore, in comparison with the most prevalent personalized tourism recommendation model, which was based on an enhanced frog leaping algorithm, the proposed model, which was based on an optimized unsupervised multiple implicit semantic trajectory mining model, demonstrated an average increase of 10.3% and 8.9% in Hit Ratio and normalized cumulative loss gain, respectively. In a comparison of the performance of the research method with that of mainstream algorithms, the research method was found to perform the best, demonstrating excellent results in terms of computational efficiency, F1 value, and resource utilization. The values for these metrics were 11.24 s, 97%, and 91%, respectively. Finally, in practical applications, two tourism planning schemes for visiting Guilin were generated based on user needs, and users were very satisfied with the generated results. In summary, the method proposed in the study has good performance and can effectively improve user satisfaction and service quality in the tourism industry.

Povzetek: Raziskava uvaja optimiziran model UMISTM za rudarjenje trajektorij in personalizirana priporočila v turizmu.

1 Introduction

In recent years, the tourism industry has been developing rapidly, and tourism services are also moving toward intelligence and informatization. However, there are still many problems in the development process, such as poor quality of intelligent recommendation services, information asymmetry, and so on. In addition, the personalized needs of tourism services are becoming more and more important. How to meet the personalized needs of different tourists in the tourism process has become an important challenge [1-3]. The Unsupervised Multiple Implicit Semantic Trajectory Mining (UMISTM) model does not require a large amount of labeled data for training, thus better mining user travel trajectories and interest preferences. At the same time, it can also mine user's cryptic semantic trajectories in different dimensions, thereby more comprehensively grasping the personalized needs of users [4-5]. However, this model requires a large amount of training data for training, as well as complex optimization and adjustment of model parameters, which will consume a lot of computational resources and time costs [6-7]. In response to the above issues, the study introduces a Multi Granularity Recommendation (MGR) framework based on Attention Mechanism (AM) to optimize the UMISTM model and establish an intelligent personalized tourism recommendation method. The research aims to improve the effectiveness of personalized tourism recommendations and user satisfaction, provide valuable insights for the tourism industry, and also provide new ideas for the development of tourist attractions. There are two main innovative points in the research. The first point is to propose an optimized UMISTM model to mine user travel trajectories, and apply it to intelligent personalized tourism recommendation methods. The second point is to design an AM-based MGR framework to improve the UMISTM model and use it to extract user preferences. The research structure is mainly divided into four parts. The first part is a review of relevant research results. The second part is to construct the UMISTM model and optimize its design for intelligent personalized tourism recommendation methods. The third part is to verify the performance and application effectiveness of the proposed research method. The last part is a summary of the research.

2 Related works

With the continuous improvement of the national economic level and the continuous increase in per capita disposable income of residents, people have more funds for tourism consumption, which has promoted the development of the tourism industry. However, designing and recommending more intelligent tourism trajectories for different user needs can further improve the quality of tourism services and user satisfaction. Lin et al. developed a two-stage optimization model to address the importance of recommendation diversity. In the first stage, the optimization model considered topic diversity, while in the second stage, a model that minimized misclassification costs was proposed to balance the diversity and accuracy of recommendations. The research method was shown to significantly improve the accuracy and diversity of recommendations in real-world travel data [8]. Nitu et al. proposed a personalized travel recommendation system based on social media activities and proximity effects. The results showed that the model was superior to the current personalized interest point model, with an overall accuracy of 75.23% [9]. Kumar et al. introduced a model based on explicit sentiment and star ratings to predict the best tourist destination. The model extracted user-selected topics through latent Dirichlet topic allocation modeling, and then introduced a morphological linear neural network model to generate sentiment scores for text content. The experimental results showed that the accuracy, recall and F1 score of the research method were 86%, 87% and 89%, respectively [10]. Fudholi et al. designed a deep learning based mobile tourism recommendation system to help tourists discover more different tourist destinations. The experimental results showed that the system achieved an average accuracy of over 85% and an average absolute percentage error of 5% [11].

The UMISTM model is an unsupervised learning method that can be used for text mining and can automatically learn the semantic and structural information of text, but its training process is more complex and time-consuming. Li et al. proposed a graph based encoding source code semantic method, which was mainly used for limited interpretive and implicit representation learning. The results showed that the proposed method outperformed text-based methods without increasing complexity [12]. Duan et al. believed that the behavior of users calling services belonged to implicit semantics. By analyzing user calls, mining their implicit semantic representations, and modeling users' hidden preferences, the research results showed that the proposed hybrid intelligent service recommendation method based on implicit semantics and explicit scoring had better recommendation accuracy and recall than existing methods [13]. Zhang et al. proposed a sparse user check-in location prediction method based on implicit meaning to predict the sparse online check-in records and decision-making context of location-based online social network users. The research results showed that this method significantly improved the prediction accuracy [14]. Due to the influence of not only network structure but also supervised and unsupervised learning on the learning of specific feature encoding models, Yao et al. designed a graph comparison accelerated encoder using unsupervised learning, which maintained the final performance at an advanced level [15]. To better demonstrate the above research findings, a summary is conducted as shown in Table 1.

Table 1: Summary of research findings

Author	Method	Index	Key Results
Lin et al. [8]	Developed a two-stage optimization model	Accuracy and diversity	/
Nitu et al. [9]	Personalized travel recommendation system based on social media activities and proximity effects	Overall accuracy	75.23%
Shah et al. [10]	Introducing a model based on explicit sentiment rating and star rating to predict the best tourist destination	Accuracy, recall, and F1 score	86%, 87%, and 89%
Kumar et al. [11]	Mobile tourism recommendation system based on deep learning	Accuracy and Mean Absolute Percentage Error	85% and 5%
Li et al. [12]	Graph based encoding source code semantic method	/	/
Duan et al. [13]	Propose a sparse user check-in location prediction method based on implicit semantics	Accuracy and recall rate	/
Zhang et al. [14]	Design a graph contrast acceleration encoder using unsupervised learning	Predictive accuracy	/
Yao et al. [15]	Sparse user check-in location prediction method based on implicit semantics	/	/

In summary, there is currently relatively little research on multiple implicit meanings, and personalized tourism recommendations are also facing great difficulties. Therefore, this study designs a UMISTM model for mining hidden information in tourism trajectories, and optimizes it using the MGR framework, ultimately obtaining an intelligent personalized tourism recommendation method based on the optimized UMISTM model.

3 Design of personalized recommendation method for intelligent tourism based on optimized UMISTM model

In recent years, with the improvement of people's living standards and the transformation of tourism concepts, more and more people have chosen to spend their leisure time on tourism activities. However, the rapid increase in tourist attractions has caused great difficulties for people to choose the attractions they are interested in. In response to the above issues, the study first constructs a UMISTM model for mining user travel trajectories, then improves it using the MGR framework, and finally constructs an intelligent personalized travel recommendation method.

3.1 User travel trajectory mining based on UMISTM model

User travel trajectory refers to the itinerary, duration of stay, and consumption situation of tourists during the tourism process. By obtaining user travel trajectory, it can better grasp user travel preferences and consumption habits, and provide reference for the development and marketing of the tourism industry [16-17]. However, it is difficult to extract valuable information from user travel trajectories. To solve the above problems, a UMISTM model is proposed, which can semantically represent various complex and diverse information in travel trajectories and has strong flexibility. Firstly, it is necessary to preprocess the data and study the collection of 200000 users' travel information from tourism portals such as Fliggy, mainly including scenic spot route information and time information. The collected data should be subjected to noise removal, missing value filling, and display of scenic spot attribute information through knowledge graphs. In the process of data cleaning, administrative region information and non-urban scenic spots are mainly screened out. Then the cleaned data is entity identical through the tourism domain knowledge graph. Finally, the route information in the tourist travelogue can be obtained. The specific process is shown in Figure. 1.

After data preprocessing, 14000 user travel trajectory data can be obtained, with 257 scenic spots in 2022. Through the above operations, a preliminary summary of user travel activities can be obtained, as shown in Figure. 2.

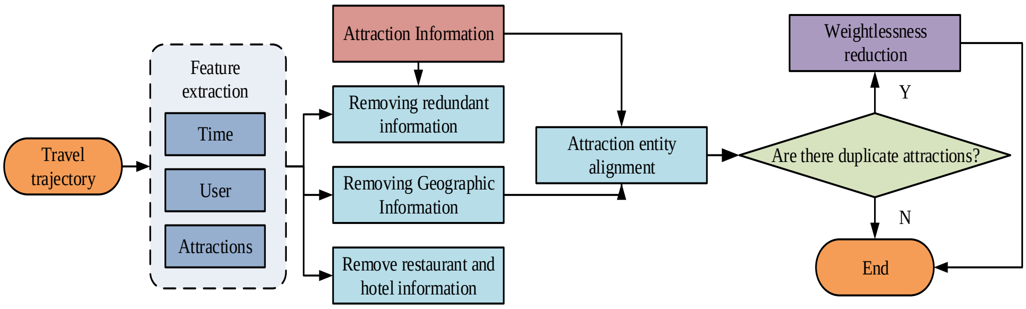


Figure 1: Data preprocessing process based on UMISTM model

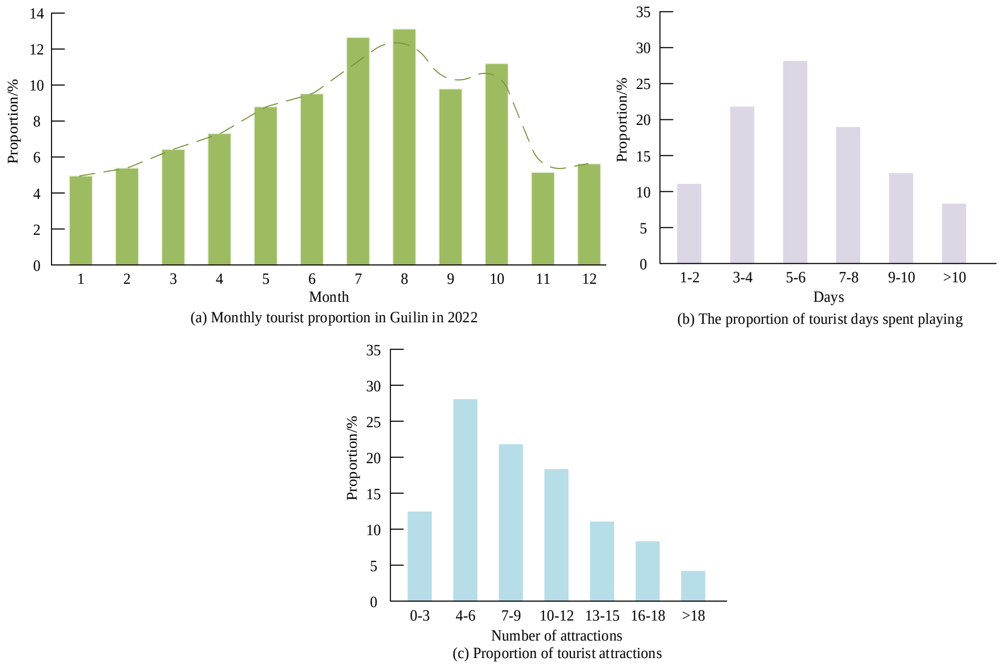


Figure 2: Preliminary summary of user travel behavior

In Figure. 2, the average number of tourist attractions in the area is 16, with over 80% of users staying in the area for 7 days and visiting multiple attractions in a relatively short period of time. The user travel trajectories obtained through research mainly include three types: time information, user information, and scenic spot information. These are integrated into the UMISTM model, and the above implicit semantic information is represented through the same dimensional space. Before constructing the UMISTM model, it is necessary to define the attraction records and travel trajectories. The former is when user $user$ travels at attraction a while in t, and the attraction record a_s can be obtained as $\{user, t, a\}$. The latter is the gameplay trajectory $PT:\{user, t_1, a_1\}, \cdots, \{user, t_i, a_i\}, \cdots, \{user, t_n, a_n\}$ where each a_s of $user$ travels continuously over time, where n represents the length of PT and satisfies the following condition $s_i < s_{i+1}, for\ i \le n-1$. The implicit meaning of user travel trajectory represents the factors that affect the travel route. Searching for the implicit meaning of user travel trajectory can better grasp the user's travel behavior. Therefore, the UMISTM model uses contextual information to obtain corresponding implicit semantic features. The specific structure is shown in Figure. 3.

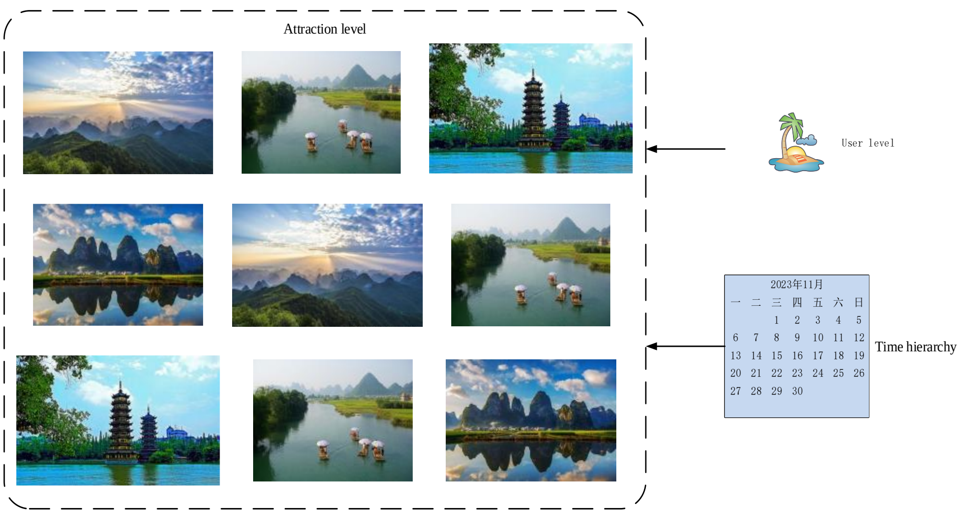


Figure 3: Structure diagram of UMISTM model

In Figure. 3, the model includes three levels of implicit meanings: time, user, and attraction. By using these implicit meanings, one can fully understand the user's tourism activities. In the modeling process of the UMISTM model, it is first necessary to represent the probability of tourism trajectory occurrence for $user$'s PT in month m through conditional probability p. Then, the multiple implicit semantic feature vectors are instantiated to obtain the objective function expression of the UMISTM model, as shown in equation (1).

$$\sum_{user\in U}\sum_{ut\in T}\sum_{i=1}^{N_t}\log p\left(a_i \middle| \underset{context}{user}, \underset{context}{t}, \underbrace{a_{i-K}:a_{i+K}}_{context}\right) \quad (1)$$

In equation (1), U, T, and N_t are the sets of $user$ and m, as well as the trajectory length, respectively. $a_{i-K}:a_{i+K}$ represents the position sequence of the model training window for a_i, and $context$ represents the context corresponding to different levels. Among them, p is usually established using the softmax function. However, in the study, the computational cost of using this function in the model is high, so negative sampling is chosen to train the model parameters. The probability of maximizing the location of all scenic spots is obtained by the given $context$ of PT and calculated as shown in equation (2).

$$\sum_{i=1}^{N_t}\log p\left(a_i \middle| f^{(i)}\right) \quad (2)$$

In equation (2), a_i represents the location of the target scenic spot, and $f^{(i)}$ represents $context$ with all implicit meanings. Among them, a_i is a positive sample, and attractions that are not in the context of the attraction are negative samples. For a given positive sample $\left(f^{(i)},a_i\right)$, the objective function is improved by using the random gradient ascent method to obtain the gradient results of the final objective function $\zeta\left(a_i,\psi\right)$ with respect to the different levels of $context$ vectors and $X_{f^{(i)}}$ in $f^{(i)}$. Therefore, the update expression V for all embedded vectors is shown in equation (3).

$$V_f := V + \eta \sum_{\psi\in\{a_i\}\cup NEG\{a_i\}} \frac{\partial \zeta\left(a_i,\psi\right)}{\partial X_{f^{(i)}}}, f \in context\left(a_i\right) \quad (3)$$

In equation (3), η represents the learning rate, and $NEG\left(a_i\right)$ represents a non empty negative sample subset related to a_i. By training through the above steps, the final UMISTM model can be obtained, which can be applied in tourism trajectory route recommendation in the future.

3.2 Optimization based on UMISTM model

Due to the unsupervised learning nature of the UMISTM model, it has high work efficiency but low accuracy [18-20]. Therefore, the MGR framework will be optimized and the learning effectiveness of the UMISTM model will be improved through diverse training data and multi granularity mining, thereby improving the personalized recommendation method for intelligent tourism. Firstly, it is necessary to define the following content: interaction event $user_i = \left(user,t,i,c\right)$ is when $user$ interacts with item i at time t, where c is the category to which i belongs. The interaction behavior sequence $S=\left\{\left(user,t_1,i_1,c_1\right),\cdots,\left(user,t_n,i_n,c_n\right)\right\},t_1<t_2<\cdots<t_n$ of $user$ concatenates all interaction points of $\left(user,t,c\right)$ in ascending order through timestamps. By forming a triplet $user$ with $user$, timestamp, and category among all interaction points, a coarse-grained interaction sequence $C=\left\{\left(user,t_1,c_1\right),\left(user,t_2,c_2\right),\cdots,\left(user,t_n,c_n\right)\right\}$ can be obtained. By combining $user$, timestamp, and i, a fine-grained interaction sequence $I=\left\{\left(user,t_1,i_1\right),\left(user,t_2,i_2\right),\cdots,\left(user,t_n,i_n\right)\right\}$ can be obtained by arranging them. User preferences are hidden in S, so it is very important to extract each $user$'s own preferences from the large amount of information collected. The study first learns and mines fine-grained preferences based on the I-sequence of each user, and then learns and mines coarse-grained preferences based on the C-sequence of each user. The specific schematic diagram of the MGR framework structure is shown in Figure. 4.

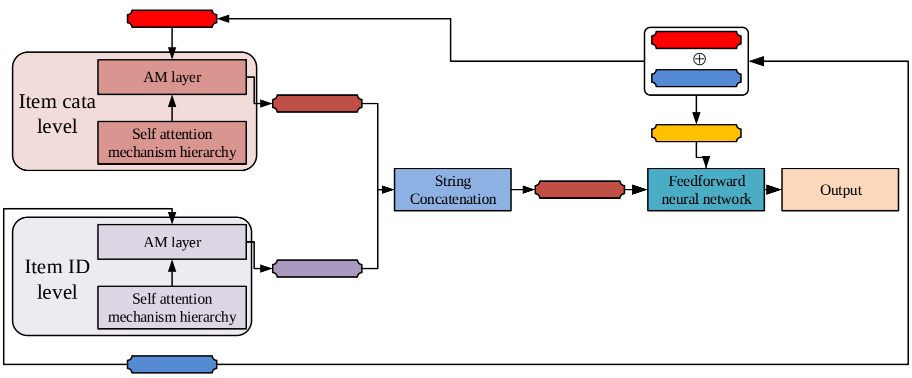


Figure 4: MGR frame structure schematic diagram

In Figure. 4, the MGR framework utilizes the same two parallel structures to mine user preferences of different granularities, and then uses the AM method to mine multi granularity preferences. Finally, static and dynamic user feature representations of different granularity levels are obtained. After concatenation, a complete preference representation of users after watching new travel trajectories can be obtained. The interaction points of the interaction behavior sequence are composed of four elements, and the information contributed by all of them is the project, the category to which the project belongs, and the time information. If there is more than one project in each category, the user's interaction behavior sequence can be divided into interaction sequences corresponding to the project and the category, which provides the user with different aspects of preference information. Therefore, an extensible two-level framework, namely MGRA, is designed to comprehensively analyze user preferences. The structure of the MGR framework specifically includes five levels: embedding layer, hidden layer, AM layer, attention layer, and application layer. The embedding layer includes embedding vector methods corresponding to c and i, and interaction time, in order to obtain the interaction behavior encoding of all $user$. The specific expression is shown in equation (4).

$$\begin{cases} user_i = fv(i) + lookup_i^t\left\{bucketize_i(t)\right\} \\ user_c = fv(c) + lookup_c^t\left\{bucketize_c(t)\right\} \end{cases} \tag{4}$$

In equation (4), $user_i$ and $user_c$ are the interaction behavior codes corresponding to i and c of $user$, $fv(i)$ and $fv(c)$ are the feature vectors corresponding to i and c after passing through the feedforward neural network, respectively. The above feature vectors are shareable, $lookup$ represents the lookup function, and $bucketize$ represents the automatic standardization of a certain time. The output result of this level is a vector list composed of all interaction i and c after combining time information, as shown in equation (5).

$$\begin{cases} I = \left\{user_{i1}, user_{i2}, \cdots, user_{in}\right\} \\ C = \left\{user_{c1}, user_{c2}, \cdots, user_{cn}\right\} \end{cases} \tag{5}$$

The above list is input to the hidden layer. Due to the significant differences between the two embedding vectors, it is not possible to perform concatenation operations to merge the information. Therefore, a linear projection method is used to uniformly project the two obtained vectors into the implicit semantic space of dimension K. The modified vector expressions for I and C can be obtained, as shown in equation (6).

$$\begin{cases} I' = cancat\left\{\chi_{\theta_1}(user_{i1}), \chi_{\theta_2}(user_{i2}), \cdots \chi_{\theta_n}(user_{in})\right\} \\ C' = cancat\left\{\chi_{\psi_1}(user_{c1}), \chi_{\psi_2}(user_{c2}), \cdots \chi_{\psi_n}(user_{cn})\right\} \end{cases} \tag{6}$$

In equation (6), $cancat$ represents the string concatenation function, and χ represents the projection function with parameter α and the relu activation function, which puts all interaction behavior and time information together into a m-dimensional vector space. To improve the expression of I and C sequences, G-fold projections are conducted on both sequences and the sequence information is described through D semantic spaces. The corresponding expressions are shown in equation (7).

$$\begin{cases} I_D = \chi_{\beta_D}(I) \\ C_D = \chi_{\varsigma_D}(C) \end{cases} \tag{7}$$

In equation (7), χ_{β_D} and χ_{ς_D} are both corresponding projection functions in $H-th$ space. In the AM layer, it is mainly used to capture the hidden intrinsic relationships in all semantic spaces. Due to the correlation between the occurrence of $user$'s interaction behavior and other interaction behaviors, the research conducts AM within previous interaction sequences to enhance the information of related interaction points. The attention dispersion matrices G_{I-D} and G_{I-C} corresponding to I and C in each subspace are calculated as shown in equation (8).

$$\begin{cases} G_{I-D} = soft\max\left(\tau(I_D, I)\right) \\ G_{C-D} = soft\max\left(\tau(C_D, C)\right) \end{cases} \tag{8}$$

In equation (8), $soft\max$ is the softmax function and τ is the calculation function of attention score, which is used to determine the impact of all interaction behaviors in $H-th$, as shown in equation (9).

$$\begin{cases} \tau(I_D, I) = I_D \cdot W_{I-k} \cdot I^T \\ \tau(C_D, C) = C_D \cdot W_{C-k} \cdot I^T \end{cases} \tag{9}$$

In equation (9), W represents the weight of the corresponding interaction behavior. In $H-th$, the expression for the attention vector is shown in equation (10).

$$\begin{cases} H_{I-D} = G_{I-D}\chi_{Q_{I-D}}(I) \\ H_{C-D} = G_{C-D}\chi_{Q_{C-D}}(C) \end{cases} \tag{10}$$

In equation (10), $\chi_{Q_{I-D}}(I)$ and $\chi_{Q_{C-D}}(C)$ are both projection functions with the activation function relu. Finally, after splicing, the recombination operation can be performed to obtain the static coarse-grained U_I and fine-grained preference U_C, as calculated in equation (11).

$$\begin{cases} U_I = \gamma_{e-I} \square cancat\{U_{I-1}, U_{I-2}, \cdots, U_{I-D}\} \\ U_C = \gamma_{e-C} \square cancat\{U_{C-1}, U_{C-2}, \cdots, U_{C-D}\} \end{cases} \quad (11)$$

In equation (11), γ_{e-I} and γ_{e-C} belong to the feedforward network of the embedding layer, and the static multi granularity preference can ultimately be obtained. In the attention layer, AM is introduced to capture the dynamic preference of $user$, and through the static multi granularity acquisition steps of the recommended items, the dynamic multi granularity preference is ultimately obtained. In the application layer, first set the sorting function rf to describe the dynamic preference q_t of $user$ and the correlation r between the recommended item b_{user}, as shown in equation (12).

$$r = h(b_{user}, q_t) \quad (12)$$

Due to the fact that the interaction probability between q_t and b_{user} belongs to a binary classification problem, the mainstream cross entropy loss function is selected, and the calculation is shown in equation (13).

$$LOSS = -\sum_{i,user} y_i \log r + (1 - y_i)\log(1 - r) \quad (13)$$

In equation (13), y_i is the predicted result. The aforementioned content illustrates that the incorporation of AM can augment the semantic data associated with pertinent interaction points, thus mitigating the detrimental effects of unrelated interaction points and facilitating the effective capture of user-specific preferences. In summary, an optimized UMISTM model can be obtained and applied to mining user travel preferences and tourism recommendations.

3.3 Intelligent personalized tourism recommendation method based on optimized UMISTM model

Based on the above content, this study constructs an intelligent personalized tourism recommendation method for cross domain preference acquisition, extracts tourism planning and preference information of new users in the city, and then generates personalized tourism routes. The specific structure of the intelligent tourism personalized recommendation method based on optimized UMISTM model is shown in Figure. 5.

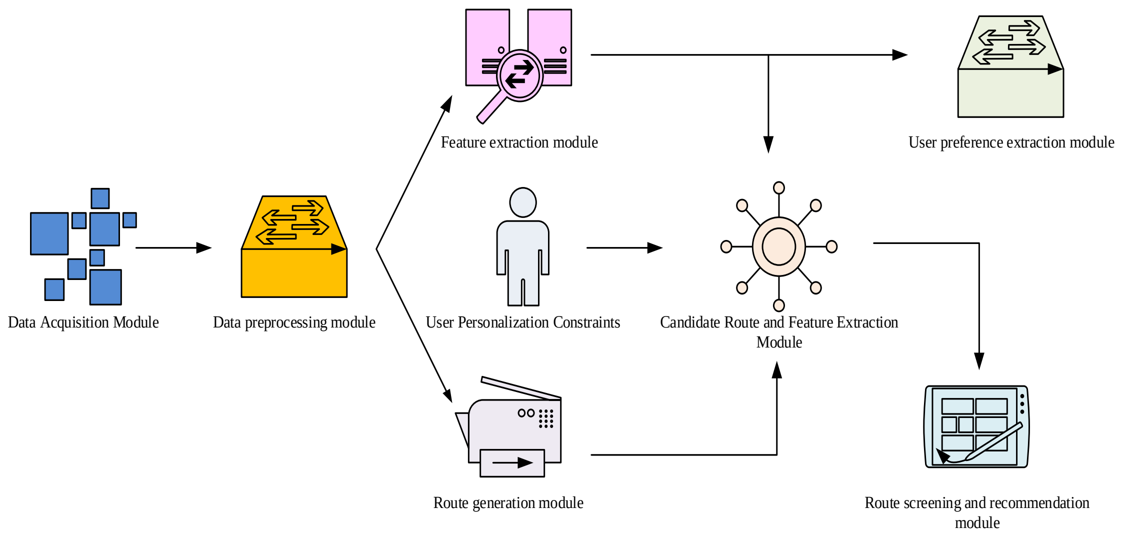


Figure 5: Process of personalized recommendation method for intelligent tourism based on optimized UMISTM model

In Figure. 5, the intelligent personalized tourism recommendation method consists of 7 modules, mainly divided into three parts: data collection and preprocessing, feature extraction and travel trajectory generation, and tourism recommendation. In the data collection and preprocessing part, the first step is to crawl the user's travel trajectory, including attractions, transportation, hotels, etc. Next, data cleaning operations are carried out to remove invalid and redundant information. Then, information such as time and travel expenses between different attractions is added, and various attributes of the attractions are encoded through a knowledge graph. Finally, the user's travel trajectory in other cities is queried to obtain corresponding personalized travel preferences, and all scenic spots are classified into fixed categories, including 25 categories such as humanities, karst caves, and history. Composite statistics are performed on the categories of scenic spots and the proportion of each category is calculated. In the feature extraction section, it mainly includes the extraction of scenic spot features and user travel preferences. In the part of travel trajectory generation and travel recommendation, personalized travel recommendation needs to meet both the user's constraints and personalized needs. The generated travel route feature representation vector is calculated as shown in equation (14).

$$b_a = \frac{b_{a1} + b_{a2} + \cdots + b_{aM}}{N} \tag{14}$$

In equation (14), N represents the feature representation vector of attraction a in the generated tourist route, and N represents the number of attractions in the tourist route. Because N and the user's personal preference b_p are known, the matching degree M between the two is calculated using equation (15).

$$M = b_a \cdot b_p^T \tag{15}$$

Through the above matching, the recommended route with the same number of candidate tourist routes and the highest matching value can be obtained.

4 Result analysis of personalized recommendation method for intelligent tourism based on optimized UMISTM model

To analyze the performance and application effect of the intelligent tourism personalized recommendation method based on optimized UMISTM model proposed in the study, the study first explored the performance of the optimized UMISTM model in the method, and then applied it to actual tourism recommendation to explore the application of the intelligent tourism personalized recommendation method proposed in the study.

4.1 Result analysis based on optimized UMISTM model

To evaluate the performance of the optimized UMISTM model, commonly used metrics such as recall, accuracy, training time, resource utilization, and F1 value were selected for evaluation. Among them, accuracy represented the overall prediction accuracy, recall was the probability of being predicted as a positive sample, F1 value reflected the comprehensive effect of the model, training was used to analyze the computational efficiency, and resource utilization measures the effective use of model resources in terms of their availability and capacity. A comprehensive analysis of the model was required in order to ascertain its accuracy, computational efficiency, and comprehensiveness. This will facilitate the development of efficient and accurate personalized tourism recommendation methods. In addition, to more scientifically validate the performance of the proposed model, the study selected the currently mainstream Personalized Travel Recommendation Based on Improved Frog Jump Algorithm (PTR-IFJ) and UMISTM models for comparative experiments. The dataset used in the study was 14000 user travel trajectory data collected and preprocessed from various popular tourism portal websites, which were divided into training and testing sets in a 7:3 ratio. The experimental parameter settings are shown in Table 2.

Table 2: Experimental parameter setting

Parameter	Corresponding parameters	Illustrate	Parameter value
thre		High frequency downsampling threshold	1E-5
η		Initial learning rate	0.05
threads		Number of program threads	9
nega		Negative sample count	25
iter		Iterations	300
size	Z	Vector dimension	200
wind	2c	Training window size	5

Experimental parameters were an extremely important part of model training, and in practical situations, it was necessary to improve the parameters based on the model, dataset, and training results. The training and experimentation of the model were conducted on the Hewlett Packard Z8 G4 workstation. The study first took the Four Seasons Autumn Begonia as an example, and its best viewing seasons were December and January, so its viewing had strong seasonal semantics. A study was conducted to compile and summarize the number of visitors to the Four Seasons Autumn Begonia in each month of 2022.

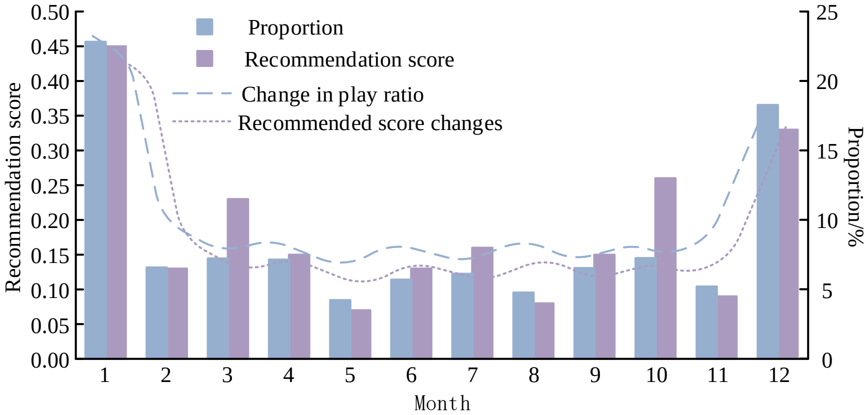
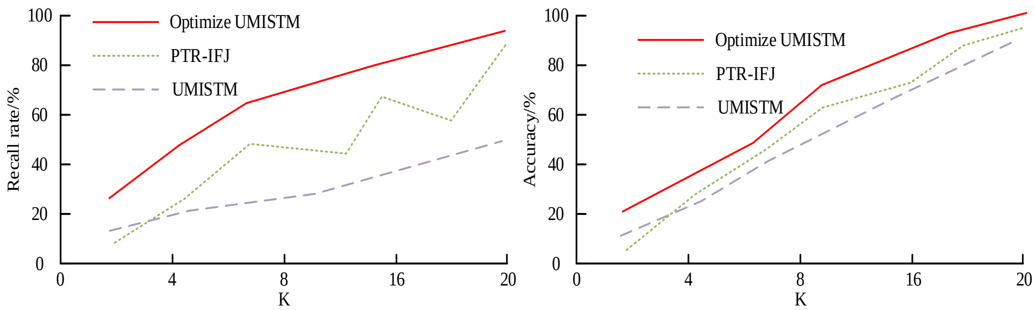


Figure 6: The Number of Visits and Recommended Results of Four Seasons Begonia in 2022 under the UMISTM Model

The number of visits and recommended results of autumn crabapple in each month of 2022 under the UMISTM model are shown in Figure. 6. Among them, the horizontal axis represented each month in mid-2022, the left vertical axis corresponds to the recommended ratings, and the right vertical axis corresponded to the share of tourists. In Figure. 6, the recommendation scores for users recommending the Four Seasons Autumn Begonia scenic spots were the highest in January and December, and the actual trend of gameplay was consistent with the changes in the UMISTM model recommendation scores. The above results indicated that the research method could fully preserve the seasonal semantics hidden in scenic spots. By studying the cosine similarity between users and attractions, a top-K candidate attraction list was generated, and then the recall and accuracy results of different recommendation models were evaluated.

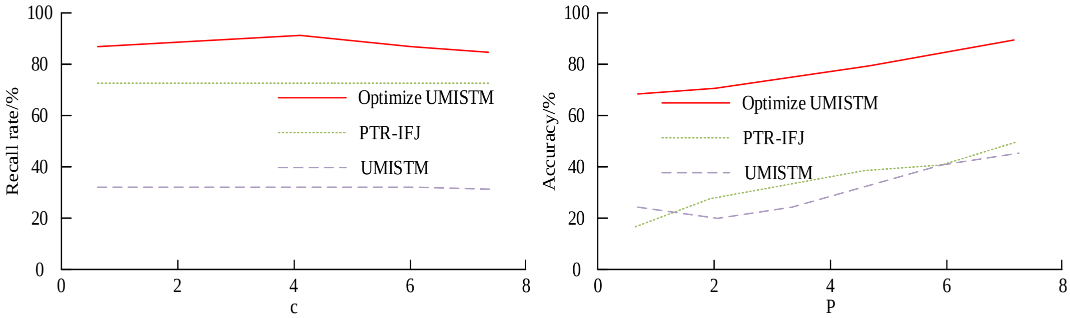


(a) The curve of recall rate with K for different models

(b) Curve of accuracy of different models as a function of K

Figure 7: Performance results of different recommendation models and the variation curve of K

The performance results of different recommendation models and the variation curve of K are shown in Figure. 7. Figure. 7(a) and 7(b) show the recall and accuracy results and the K value, respectively. The vertical axis corresponded to the recall and accuracy indicators, while the horizontal axis corresponded to the K value. As the length of the scenic spot recommendation list continued to expand, the optimized UMISTM model had the fastest growth rate in both recall and accuracy. When K was 20, the corresponding recall and accuracy were 86.57% and 97.62%, respectively. In the accuracy results, the UMISTM model gradually surpassed the PTR-IFJ model as the K value increased, ultimately reaching 95.16%. This was due to the fact that as the K-value increases, the recommended list corresponding to the recommendation model will expand, thereby increasing the probability of recommending accurate tourist attractions to users. Consequently, the results of the study will demonstrate an upward trajectory in terms of recall and accuracy, while the implicit semantic information embedded within the research method will exhibit greater richness and completeness. To explore the impact of the size of the model training window and the P-value on the model, a fixed K-value of 15 was set, where P was the number of historical scenic spots that the user watched.

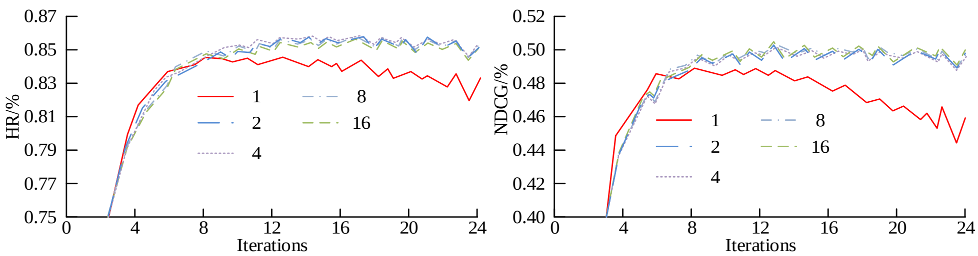


(a) The Change Curve of Recall Rate and Training Window for Different Models

(b) Curve of accuracy of different models with respect to P

Figure 8: Variation curves of different models under different influencing factors

The variation curves of different models under different influencing factors are shown in Figure. 8. Figure. 8 (a) and 8 (b) show the change curves of recall and accuracy under different influencing factors. As the training window size continued to expand, the optimized UMISTM model showed a trend of initially increasing at a larger rate and then slowly decreasing, while the curves of the other two recommended models were relatively stable. This indicated that when c was 4, the proposed model retained the most comprehensive implicit semantic information. When P was 24, the accuracy of the optimized UMISTM model, PTR-IFJ model, and UMISTM model were 97.25%, 96.33%, and 91.78%, respectively. To further evaluate the performance of the optimized UMISTM model, the study selected Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG) for evaluation.



(a) HR

(b) NDCG

Figure 9: The impact of the number of hidden semantic spaces on HR and NDCG of optimizing UMISTM models

The impact of the number of hidden semantic spaces on the HR and NDCG of the optimized UMISTM model is shown in Figure. 9. Figure. 9 (a) and 9 (b) show the results of HR and NDCG, respectively. As the number of iterations increased, the number of semantic spaces in different hidden layers showed an increase in both indicators. Among them, the performance of optimizing the UMISTM model was the worst when the number was 1, and the changes in HR and NDCG were relatively consistent under the other numbers. However, considering both indicators comprehensively, the model had the best performance when the number was 4, with the average values of the two indicators being 83.16% and 48.85%, respectively. To further analyze the performance of the proposed model, the TMDB 5000 Movie dataset was introduced for testing, which included 500000 comments from 975 users on 10000 movies, mainly including user, movie, and rating data.

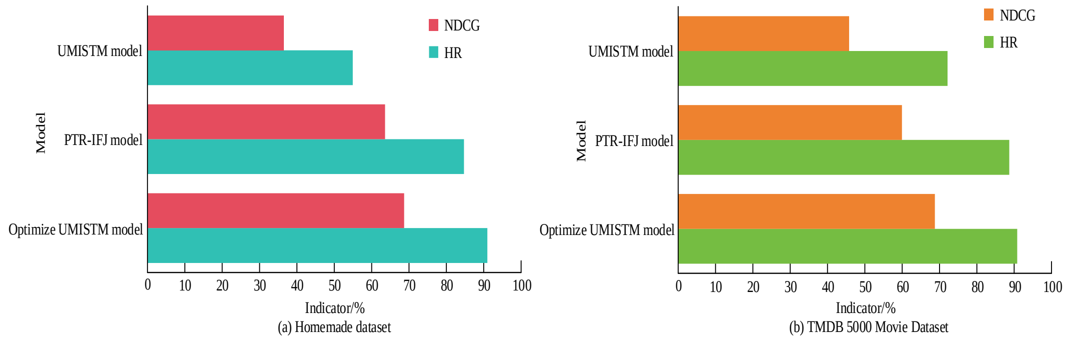


Figure 10: Performance results of different recommendation models on self-made datasets and TMDB 5000 Movie datasets

The performance results of different recommendation models on the self-made dataset and TMDB 5000 Movie dataset are shown in Figure. 10. Figure. 10 (a) and 10 (b) show the performance results of the self-made dataset and TMDB 5000 Movie dataset, respectively. The optimized UMISTM model achieved the best results on HR and NDCG on different datasets. Compared with the most mainstream PTR-IFJ model, the proposed model improved HR and NDCG by an average of 10.3% and 8.9%, which confirmed the effectiveness of the research method. In summary, the model proposed in the study had better performance and comprehensively captured user preferences and personalized preferences during travel. Finally, to further analyze the computational efficiency and resource utilization of the research method, the effectiveness of different recommendation models was tested. In addition to the comparative recommendation models previously discussed, the study also introduces the two-stage optimization model, which demonstrated the most optimal performance in the relevant work section. Furthermore, models based on explicit sentiment rating and star rating are presented for comparison. The results are presented in Table 3.

Table 3: Calculation efficiency and resource utilization results of different recommendation models

Recommendation models	Average training time/s	F1 value/%	Resource utilization rate/%
Optimize UMISTM	11.24	97	91
PTR-IFJ	18.36	91	87
UMISTM	13.65	92	89
Two stage optimizations	21.35	90	88
A model based on explicit emotional rating and star rating	26.83	88	85

In Table 3, the research method displayed excellent performance in terms of computational efficiency, overall performance, and resource utilization, with values of 11.24s, 97%, and 91%, respectively. However, the model based on explicit sentiment and star ratings performed poorly, with corresponding indicators of 26.83, 88%, and 85%, respectively. Among them, the PTR-IFJ recommendation model had excellent performance, with computational efficiency, comprehensive performance and resource utilization of 18.36s, 91% and 87%, respectively. Compared with them, the research method greatly improved in various aspects of performance. This was because the research method effectively and deeply explores various hidden semantic information in the route, and obtained more comprehensive user personalized preferences at different granularities in the feature extraction part. Lightweight processing methods were introduced, so the accuracy and computational efficiency of search was greatly improved. In addition, the PTR-IFJ recommendation model had low computational efficiency due to insufficient communication between local and global information during the search process. Compared with the best performing two-stage optimization models in related work, both made certain improvements in accuracy. Compared with the best performing two-stage optimization models in related work, both have shown certain improvements in accuracy. The two-stage optimization model improves the accuracy and diversity of tourism recommendations through different stages, but does not pay attention to computational efficiency, resulting in lower computational efficiency.

4.2 Application of intelligent tourism personalized recommendation method based on optimized UMISTM model

To test the practical application effect of the intelligent tourism personalized recommendation method based on the optimized UMISTM model, a study randomly selected a user who planned to play for 3 days in June, and generated two new routes through the intelligent tourism personalized recommendation method.

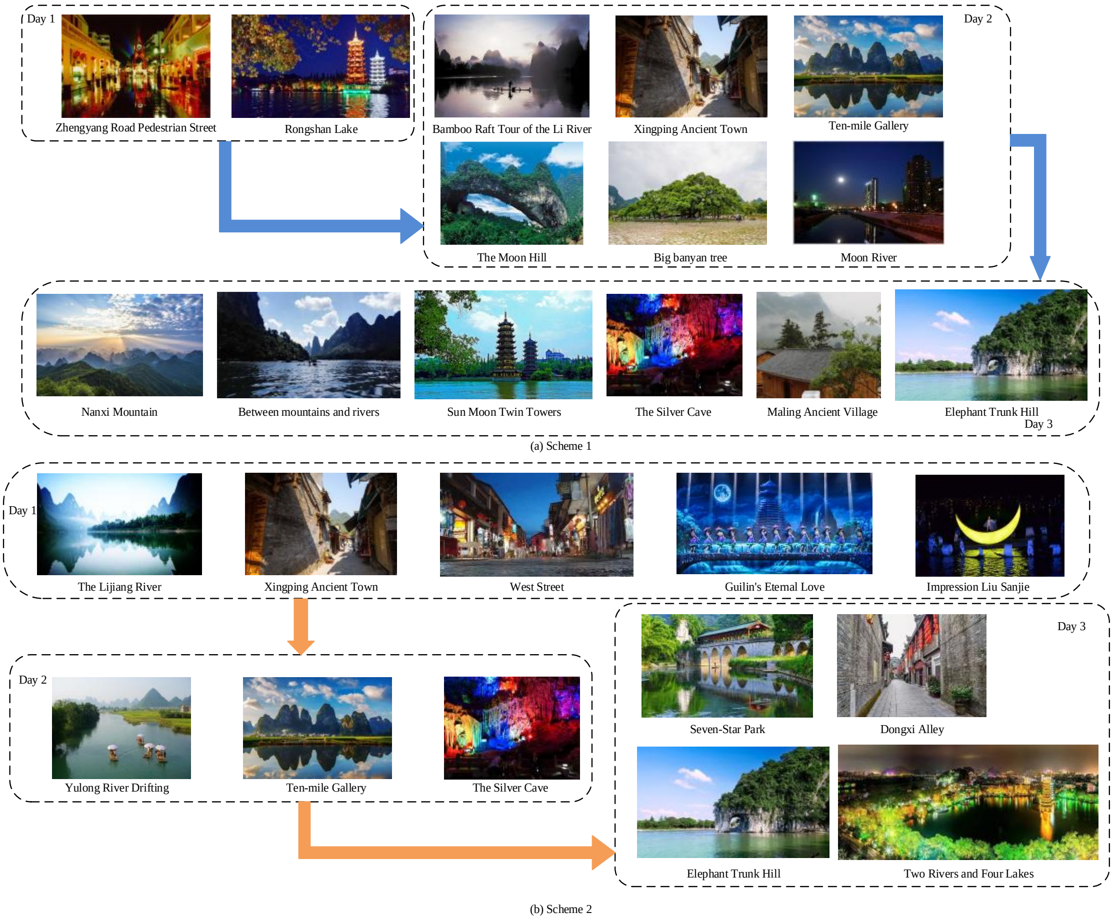


Figure 11: Recommendation scheme of intelligent tourism personalized recommendation method based on optimized UMISTM model in practical applications

The recommendation scheme of the intelligent tourism personalized recommendation method based on optimized UMISTM model in practical application is shown in Figure. 11. This method generated two completely different tourism routes based on user needs and personalized preferences. Scheme 1's 3-day journey arrangement included Zhengyang Road Pedestrian Street, Rongshan Lake, Bamboo Raft Tour of Lijiang River, Xingping Ancient Town, Shili Gallery, The Moon Hill, Big Banyan Tree, Moon River, The Silver Cave, Maling Ancient Village, Nanxi Mountain, Shanshui Jian, Sun Moon Twin Towers, and Elephant Trunk Mountain. Not only can tourists experience the local natural scenery, but also, they can also taste local specialties. The second plan was relatively simple, with itinerary arranged for Lijiang Scenic Area, Xingping Ancient Town, West Street, Guilin's Eternal Love, "Impression Liu Sanjie" performance, Yulong River Drifting, Ten Li Gallery, Yinziyan, Seven-Star Park, Dongxi Alley, Elephant Trunk Mountain, and Two Rivers and Four Lakes. This plan can also appreciate the local historical and cultural heritage. The user expressed satisfaction with both tourism planning solutions. By acquiring personalized preference and attraction vectors from users and inputting them into a feedforward neural network, it was possible to calculate the user's satisfaction with the corresponding attractions. This allowed attractions with high satisfaction ratings to be effectively recommended to users.

4.3 Discuss

In the context of the current vast tourism information network, the issue of how to provide tourist attractions that are easily accessible according to consumer needs represents a significant challenge that must be addressed in the process of China's tourism informatization development and industrial upgrading. Therefore, a UMISTM model was proposed to mine users' travel traces, and a more comprehensive user preference was obtained through the MGR framework, and finally an intelligent tourism personalized recommendation method was obtained.

Firstly, the performance of the optimized UMISTM model was analyzed. Among the change curves of recall rate, accuracy rate and K value of different models, the recall rate and accuracy rate of the optimized UMISTM model were the fastest growing rate. When K was 20, the corresponding recall rate and accuracy rate were 86.57% and 97.62%, respectively. The above results were generated because as the value of K increased, the list of recommendations corresponding to the recommendation model would increase, and the probability of recommending accurate attractions to users will increase. Therefore, the corresponding recall rate and accuracy results could show an increasing trend, and the cryptic information contained in the research method could be richer and more complete. The performance change curves of each influencing factor showed that with the continuous expansion of the size of the training window, the optimized UMISTM model displayed a trend of first increasing at a large growth rate and then decreasing slowly, which indicated that when c was 4, the proposed model retained the most comprehensive semantic information. The performance results of each recommended model in the self-made dataset and the TMDB 5000 Movie dataset were as follows: Compared with the current most mainstream PTR-IFJ model, the model proposed in this study had an average increase of 10.3% and 8.9% in HR and NDCG. It indicated that the research method could comprehensively capture users' preferences and personalized preferences during travel.

Then, the two-stage optimization model with the best performance in related works and the model based on explicit emotion score and star rating were compared and tested, and the analysis was carried out from the three perspectives of computational efficiency, F1 value and resource utilization. The research method showed the best results in each performance index, and the computational efficiency, F1 value and resource utilization ratio were 11.24s, 97% and 91%, respectively. Among them, the indicators of PTR-IFJ recommendation model were 18.36s, 91% and 87%, respectively. Compared with them, the performance of the research method greatly improved in all aspects, which was because the research method was effective and deeply excavated different hidden meaning information in the route, and obtained more comprehensive personalized preferences of users under different granularity in the feature extraction part. Lightweight processing was introduced. In addition, the model recommended by PTR-IFJ was not computationally efficient due to insufficient local and global information exchange in the search process. Compared to the two-stage optimization model, the accuracy of both models improved to some extent. The research method was processed from the data mining part, lightweight design and feature extraction part, and also effectively improved the computational efficiency and accuracy of its application. However, the two-stage optimization model improved the accuracy and diversity of travel recommendation through different stages, and it did not pay attention to the computational efficiency, so its computational efficiency was low.

Finally, two different travel routes are generated according to user needs and personalized preferences in practical applications. Since the designed system is based on the user's personalized preference vector and scenic spot vector, which are input into the feedforward neural network to obtain the user's satisfaction with the corresponding scenic spot, the users are satisfied with the generated system results.

In summary, the research method can fully meet the individual needs of users, is conducive to promoting the high-quality development of tourism industry, and is an important assistant to promoting the green and sustainable development of tourism.

5 Conclusion

To solve the problem of poor recommendation quality in traditional recommendation methods, a UMISTM model was proposed to mine user travel trajectories, and then the MGR framework was used to optimize it. Finally, an intelligent personalized travel recommendation method was designed. The experimental results showed that the recall and accuracy growth rates of the optimized UMISTM model were the fastest, with corresponding recall and accuracy rates of 86.57% and 97.62% at K=20, respectively. In the accuracy results, the UMISTM model gradually surpassed the PTR-IFJ model with the increase of K value, ultimately reaching 95.16%. The optimized UMISTM model achieved the best results on both HR and NDCG datasets. Compared with the current mainstream PTR-IFJ model, the proposed model showed an average increase of 10.3% and 8.9% in HR and NDCG. Finally, in the application results, the study provided two tourism planning schemes based on the personalized needs of users, and users believed that both schemes could meet their own requirements. In summary, the proposed method has good performance and can fully meet the personalized needs of users in actual tourism recommendations. However, there are still shortcomings in the research. The research only assists users in providing personalized travel routes. However, during the travel process, users not only need to visit scenic spots, but also need to provide real-time activity information such as restaurants, experience shops, and accommodation. Therefore, in future research, user preference information during travel can be further divided to provide better services to users.

Reference

[1] Hasan R, Koles B, Zaman M, Paul J. The potential of chatbots in travel and tourism services in the context of social distancing. International Journal of Technology Intelligence and Planning, 13(1):63-83, 2021. https://doi.org/10.1504/ijtip.2021.10041470

[2] Beldona S, Kher H V, Lin K. Gains-focused vs risk-averse orientations and their impact on location-based marketing services in tourism. Journal of Hospitality and Tourism Technology, 13(2):333-347, 2022. https://doi.org/10.1108/jhtt-07-2021-0209

[3] Shen L. Analysis of Immersive Virtual Reality Tourism Resources Based on Navier-Stokes Equations. Informatica, 48(6):131-140, 2024. https://doi.org/10.31449/inf.v48i6.5514

[4] Chefrour A, Drissi S. K-CAE: Image Classification Using Convolutional AutoEncoder Pre-Training and K-means Clustering. Informatica (Slovenia), 47, 2023. https://doi.org/10.31449/inf.v47i7.4499

[5] G Mehdi, H Hooman, Y Liu, S Peyman and R. Arif Data Mining Techniques for Web Mining: A Survey. Artificial Intelligence and Applications, 1(1):3-10, 2022. https://doi.org/10.47852/bonviewAIA2202290

[6] Jiang H, Wan C, Yang K, Ding Y, Xue S. Continuous missing data imputation with incomplete dataset by generative adversarial networks–based unsupervised

learning for long-term bridge health monitoring. Structural Health Monitoring, 21(3):1093-1109, 2022. https://doi.org/10.1177/14759217211021942
[7] Obert J, Loffredo T. Efficient Binary Static Code Data Flow Analysis Using Unsupervised Learning. International Journal of Semantic Computing, 16(4):569-583, 2022. https://doi.org/10.2172/1592974
[8] Lin Y, Huang C, Yao W, Shao Y. Personalised attraction recommendation for enhancing topic diversity and accuracy. Journal of Information Science, 49(2):302-318, 2023. https://doi.org/10.1177/0165551521999801
[9] Nitu P, Coelho J, Madiraju P. Improvising personalized travel recommendation system with recency effects. Big Data Mining and Analytics, 4(3):139-154, 2021. https://doi.org/10.26599/bdma.2020.9020026
[10] Shah C, Trupp A, Stephenson M L. Conceptualising local perceptions of research-related tourism in an indigenous village in Fiji. International Journal of Tourism Research, 25(4):416-428, 2023. https://doi.org/10.1002/jtr.2578
[11] Fudholi D H, Rani S, Arifin D M, Satyatama M R. Deep Learning-based Mobile Tourism Recommender System. Scientific Journal of Informatics, 8(1):111-118, 2021. https://doi.org/10.15294/sji.v8i1.29262
[12] Li L, Liu Y. Mapping Modern JVM Language Code to Analysis-Friendly Graphs: A Study with Kotlin. International journal of software engineering and knowledge engineering, 32(11/12):1667-1688, 2022. https://doi.org/10.1142/s0218194022500735
[13] Duan L, Gao T, Ni W, Wang W. A hybrid intelligent service recommendation by latent semantics and explicit ratings. International Journal of Intelligent Systems, 36(12):7867-7894, 2021. https://doi.org/10.1002/int.22612
[14] Zhang D, Zhang Y, Li Q, Wang D. Sparse User Check-in Venue Prediction by Exploring Latent Decision Contexts From Location-Based Social Networks. IEEE transactions on big data, 7(5):859-872, 2021. https://doi.org/10.1109/tbdata.2019.2957118
[15] Yao Z, Yu J, Ding J. Contrastive learning of graph encoder for accelerating pedestrian trajectory prediction training. IET image processing, 15(14):3645-3660, 2021. https://doi.org/10.1049/ipr2.12185
[16] Karami F, Malek M R. Trajectory similarity measurement: An enhanced maximal travel match method. Transactions in GIS, 25(3):1485-1503, 2021. https://doi.org/10.1111/tgis.12733
[17] Kim Y K, Choi W C. Effect of change of sand properties on travel distance of ricocheted debris. Defence Technology, 17(4):1486-1495, 2021. https://doi.org/10.1016/j.dt.2020.08.006
[18] Tian L. Analysis of Media Content Recommendation in The New Media Era Considering Scenario Clustering Algorithm. Informatica, 48(6):141-156, 2024. https://doi.org/10.31449/inf.v48i6.5375
[19] An Q, Maggioni M, Kevrekidis Y, Lu F. Unsupervised learning of observation functions in state space models by nonparametric moment methods. Foundations of Data Science, 5(3):340-365, 2023. https://doi.org/10.3934/fods.2023002
[20] Li L, Li X, Yang S, Ding S, Zheng X. Unsupervised-Learning-Based Continuous Depth and Motion Estimation with Monocular Endoscopy for Virtual Reality Minimally Invasive Surgery. IEEE transactions on industrial informatics, 17(6):3920-3928, 2021. https://doi.org/10.1109/tii.2020.3011067