[1] Siyang Wang

Deep Learning-based Visitor Behavior Analysis and Prediction for Rural Tourism Intelligent Platforms

Journal of Electrical Systems

Abstract: - With the continuous development of deep learning technology and the increasing maturity of rural tourism market, this paper obtains tourism user-generated content data through customized crawler technology, describes the data flow diagram of single-user crawling and the data flow diagram of database batch crawling module. A sentiment index covering multiple dimensions is constructed to mine the deep-seated features of tourist behavior. Fusing effective features in tourism data by using multiple topological maps, using graph convolution network to capture multiple spatial features of scenic spots and recurrent neural network to capture temporal features of traffic, to complete the analysis and prediction of tourists' behavior. Taking Jiangxi Wuyuan Huangling rural attraction market as an example for empirical analysis, the importance of historical flow and search volume under all time windows is as high as 111 and 117 respectively, proving that these two features have a significant impact on predicting the target variables. The model in this paper is highly fitted to the predicted value of actual passenger flow at 12 time points, especially in the 9th month, the predicted value is 402, which is 401 from the actual value, which is an important reference value for rural tourism management and marketing strategy.

Keywords: deep learning; rural tourism; crawling technology; multiple topological maps; tourist behavior analysis

1. Introduction

With the continuous improvement of national living standards, the concept of leisure tourism has been popularized [1]. Rural tourism has ushered in unprecedented changes and gradually become a hot topic in the current tourism market [2-3]. With its unique geographical location, rich folk culture connotation and relaxed and free lifestyle, rural tourism has attracted more and more tourists and gradually become a favored travel choice for the majority of tourists [4-5]. However, with the surge in the number of tourists, the rural tourism market is also facing new challenges, how to accurately grasp the characteristics and needs of tourists, optimize the allocation of resources, and enhance the tourist experience, which has become a key issue in the development of rural tourism at present.

The rise of deep learning technology has brought new opportunities for the development of the tourism industry By collecting and analyzing various types of behavioral data of tourists in the process of rural tourism, such as tour routes, length of stay, consumption preferences, evaluation feedback, etc., it can reveal tourists' preferences, needs and behavioral patterns, provide scientific decision-making support for tourism administrators, promote the development of rural tourism in the direction of personalization, and further enhance the rural tourism Intelligent level. There have been several writings in the field of intelligent tourism to provide data support, Li, D et al. proposed a data mining technique based on DA-HKRVM algorithm to predict the changes of tourist

[1] *College of Vocational and Technical,Guangxi Normal University,Guilin,541000,Guangxi China.Email:emiliedida666@126.com

flow in time and space distribution. By feeding the prediction results back to the attraction staff in real time, the scale of passenger flow distribution can be effectively controlled, and the balanced distribution of tourism resources can be realized, which further promotes the development of intelligent tourism [6]. Carrese, S et al. investigated the problem of network performance prediction using traffic data acquired by Bluetooth devices, adopting data-driven approach and testing different statistical models within the methodological framework to better predict path travel time [7]. Li, S et al. investigated a personalized recommendation algorithm for intelligent travel service robots in a big data environment considering food, accommodation, attractions and amusement routes planning and used the algorithm to develop travel service robots to provide personalized travel services and recommend optimal solutions to people [8]. Husain et al. supported decision making through a plain Bayesian algorithm system to determine the optimal and strategic location, which is applied to tourists' visit decision through three variables [9]. Hamid, R. A determines the type and name of places preferred by tourists by projecting the GPS location on Google maps and uses K nearest neighbor algorithm based on inverse distance to find the nearest location of the tourists [10]. Lee, G. H proposed mathematical expressions for the problem of clustering of touristic routes and two stages of sequential pattern clustering algorithm for similar or identical routes with examples. The first stage eliminates uncommon tourist route patterns from this matrix and the second stage uses sequence mining algorithms to determine the tourist routes [11]. Gao, Y used an advanced algorithm based on neural network integration and successfully constructed an effective model for predicting the tourism demand of museums. Relevant data such as the number of historical visitors to the museum, ticket sales data, holiday schedules, and information about special exhibitions were collected to enhance the prediction accuracy and stability [12].

In this context, a rural tourism intelligent platform based on a large deep learning model is constructed to provide in-depth analysis and accurate prediction of tourist behavior. In the data acquisition stage, the data flow graph of single-user and database batch crawling with custom crawler technology is used to provide a data basis for subsequent analysis of tourist behavior. By constructing a sense analysis system, the M-GCNGRU prediction model is proposed, which integrates graph convolutional network and gated recurrent unit, aiming to capture the spatio-temporal dependence of scenic features, and the model framework, scenic features modeling, spatio-temporal dependence modeling, and learning and optimization strategy are elaborated in detail. Finally, this paper takes the rural attraction market of Huangling in Wuyuan, Jiangxi Province as an example, with a view to providing strong technical support for the analysis and prediction of visitor behavior on the intelligent platform of rural tourism.

2. Rural Tourism Intelligent Platform Data Focus Crawler

MFWFetcher is a customized online tourism data crawler for tourists, crawling objects are user data and travelogue data of big data network, obtaining tourism user-generated content data through customized crawling technology, describing the data cargo period process of data flow diagram of single-user crawling data flow diagram and data flow diagram of database bulk crawling module, which provides a rich data base for subsequent analysis of tourists' behavior.

The data crawling module targets online travel information sharing websites for data crawling. The main function of the data crawling module is to crawl the webpage source code and parse out the structured data. For customized crawlers, it is crucial to formulate appropriate crawling rules and choose a suitable way to store the crawled data. Since a web crawler needs one or more URL addresses as the entry point for crawling, MFWFetcher's data crawling module has two different crawling methods, single-user crawling and database batch crawling, based on the difference of the provided URLs.

2.1 Single-user capture module

The data flow of the single-user crawling module is shown in Figure 1. This function is used for the first time or for crawling when the URL table of the user to be crawled is empty, when using it, the URL address of the user of the travel network is provided to the MFWFetcher, which first determines whether the user already exists in the Visitor Information Table, and gives up the crawling if he/she exists. If not, the crawler crawls the user's homepage and parses the user's basic information in the user's homepage and puts it into the visitor information table. Parses the URL address of the user's friend's homepage and puts it into the user URL table to be crawled. Parses the URL address of the user's travel notes contained in the user's home page and parses the travel notes.

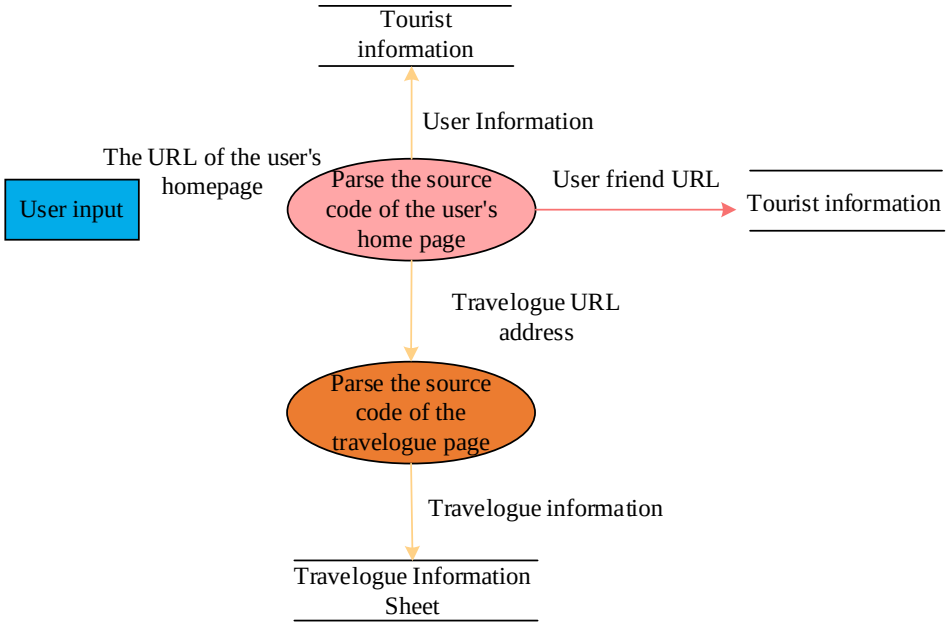

Figure 1 Single user crawling data flow

2.2 Database Batch Crawl Module

The database batch crawling module is shown in Figure 2, which is used in the scenario that there is a user's homepage URL address in the user URL table to be crawled. In batch crawling mode, MFWFetcher reads user URLs from the table of user URLs to be crawled, deletes the URLs in the table to be crawled after reading, and determines whether the user already exists in the table of visitor information, and if it exists, it selects the next user homepage URL to be crawled. If it does not exist, then crawl to work, the process is the same as the single-user crawling process. The batch crawling module operates in such a way that after a user has finished

crawling, it will select the next URL from the table of user homepage URLs to be crawled, and so on iteratively, the batch crawling function can be realized [13].

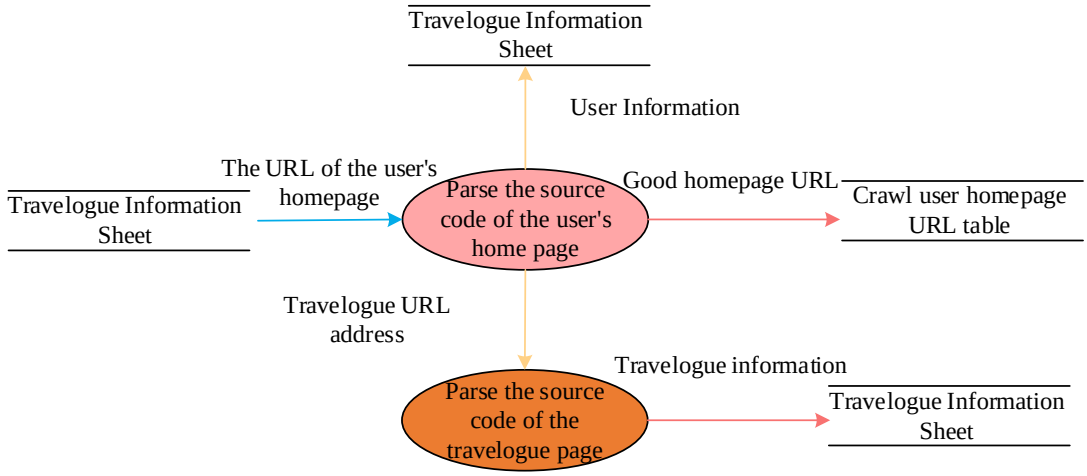

Figure 2 Database batch crawling module

3. Rural Tourism Intelligent Platform Visitor Characteristics Modeling

3.1 Tourism emotion construction

Social media review data is used to carry out sentiment mining and index construction, the sentiment index is designed to reflect the specific sentiment of travelers towards subdivided aspects of tourism activities, such as playing, shopping, dining, etc., and to categorize the sentiment polarity of the extracted aspect terms, positive or negative. For example, the review text for the view is very good, but the service attitude is very poor Li, the view and service for the corresponding aspect terms, very good and very poor, then for the two aspects corresponding to the emotional polarity, the combination of aspect terms and emotional polarity is called a feature-opinion pair.

The number and categories of aspect attributes are determined based on the LDA topic model, and all the output aspect terms are input into the model as a collection of text, and the optimal number of topics is determined by calculating the perplexity value. Low perplexity all indicate that the keywords of each theme are relatively independent and the theme characteristics are more distinct. At the same time, the representative keywords of each theme are output, and the theme name is determined by combining the keywords of each theme, and then all the aspect words are categorized according to the determined theme. The sentiment index for each dimension is constructed using the bullish index, which has been validated as the most stable method of calculating the sentiment tendency for reflecting the overall positive or negative sentiment tendency of the commenters in a given observation period [14]. The daily sentiment index was calculated using the following formula:

$$sentiment_{i,t} = \ln\left[\frac{1+N_{i,t}^{pos}}{1+N_{i,t}^{neg}}\right], i = 1,2,\cdots,8 \tag{1}$$

Where, $N_{i,t}^{pos}$ and $N_{i,t}^{neg}$ represent the number of positive and negative affective feature-opinion pairs for the

i rd aspect in the comments on day t, respectively. If the sentiment index is greater than 0, it means that the overall sentiment of the travelers is positive and vice versa the overall sentiment of the travelers is negative.

3.2 Multi-source heterogeneous data prediction

The characteristics based on the popularity of rural tourist attractions, the characteristics of rural tourist attractions land function scenic area, the location between rural tourist attractions, and the traffic connectivity of rural tourist attractions have an impact on the traffic flow of the region at a certain time in the future. The M-GCNGRU prediction model that integrates these characteristics to improve the performance of visitor behavior analysis and prediction, for all scenic spots traffic flow X, such that $X^{(t)} \in X$ denotes the traffic flow at the moment t. T denotes the input historical duration, and T denotes the future predicted duration [15]. The research problem in this paper can be described as, based on the historical traffic flow data of all scenic spots at $(t-T)$ to t time period, predict the future traffic flow data of each scenic spot at $(t+1)$ to $(t+T)$ time period, denoted as:

$$\left[X^{(t-T)},\cdots,X^{(t-1)},X^{(t)};G^{*}\right]\xrightarrow{h}\left[X^{(t+1)},\cdots,X^{(t+T)}\right] \tag{2}$$

where G^{*} is a collection of graph networks consisting of scenic attributes and relationships between scenic areas. Traffic flow prediction is realized by learning a function h which maps T-time-long historical traffic data to T'-time-long future traffic data [16]. Existing traffic flow prediction solutions are defective because they do not fully consider the characteristics of the target area as well as the spatio-temporal correlation, and this paper proposes a new M-GCNGRU traffic flow prediction model for predicting the traffic flow in scenic spots. Figure 3 shows the framework for analyzing and predicting tourist behavior in rural tourism scenic areas. In feature modeling, the traffic flow matrix and multi-feature map are constructed using multi-source heterogeneous information from vehicle trajectories, social platforms and maps. In spatio-temporal modeling, the model learns spatial correlation and temporal correlation by graph convolution network and cyclic recursive unit, respectively, based on the feature results.GCN learns spatial correlation between rural tourist attractions and combines them into a single map by weighted summation.GRU learns temporal correlation of traffic flow in rural tourist attractions and outputs analysis and prediction of tourists' behaviors of the attractions.

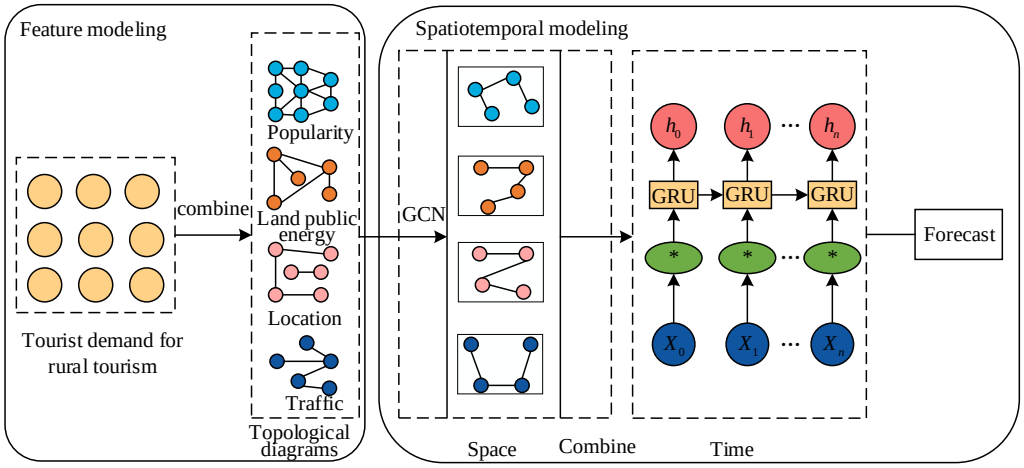

Figure 3 Tourist behavior analysis and prediction framework for rural tourism scenic spots

3.3 Characterization of rural tourist attractions

In this paper, the basic features of the attractions are selected in a data-driven manner, and the relationships between the attractions are constructed as a feature map G to represent their spatio-temporal correlations. Here, Figures G, $G=(V,E)$ are topological graphs consisting of a matrix of spatio-temporal traffic flows and a matrix of relationships between attractions. For each graph G, G is the set of attractions, corresponding to the nodes in the graph. E is the relationship between attractions, corresponding to edges in the graph. The traffic flow of all attractions V constitutes the spatio-temporal traffic flow matrix TF, and the relationships among all attractions constitute the adjacency matrix A. The adjacency matrix A of a multi-graph consists of four matrices $\{P,F,L,T\}$. The prevalence matrix P represents the attributes of the attractions, the functionality matrix F represents the land-use functions, and the distance matrix L and traffic accessibility matrix T represent the spatial relationships among the attractions.

3.3.1 Popularity of rural tourist attractions

More popular attractions generate more traffic in the surrounding area. Typically, the more visitor reviews, the more attention the attraction receives and the more popular it is [17]. Therefore, the number of reviews can indicate the popularity of an attraction. For two attractions V_i and V_j, C_{V_i} and C_{V_j} denote the number of their user reviews, respectively, the popularity similarity P_{V_i,V_j} of these two attractions can be expressed by the ratio of the number of fewer of them to the number of more of them, and thus the popularity similarity matrix P among all attractions is as follows:

$$P=\left[P_{V_i,V_j}=\begin{cases}\dfrac{C_{V_i}}{C_{V_j}}, if\ C_{V_i}\le C_{V_j}\\ \dfrac{C_{V_j}}{C_{V_i}}, if\ C_{V_i}>C_{V_j}\end{cases}\right] \tag{3}$$

3.3.2 Land functions of rural tourist attractions

Using the percentages of the different types of POIs, which indicate the land use function of the area in which the attraction is located, F will be used as the similarity of the land use function of the rural tourist attraction:

$$F=\left[\cos\left(F_{V_i},F_{V_j}\right)=\frac{\sum_{k=1}^{r}F_{v_i}^{k}\times F_{v_j}^{k}}{\sqrt{\sum_{k=1}^{r}\left(F_{v_i}^{k}\right)^2}\times\sqrt{\sum_{k=1}^{r}\left(F_{v_j}^{k}\right)^2}}\right] \tag{4}$$

where $F_{v_i}^{k}$ and $F_{v_j}^{k}$ represent the number of POIs of category k in the areas where attractions V_i and V_j are located, respectively, r is the total number of POI categories, and $\cos\left(F_{V_i},F_{V_j}\right)$ denotes the cosine similarity of the land-use functions in the areas where the two attractions are located.

3.3.3 Location between rural tourist attractions

The closer the predicted objects are geographically, the more similar the trend of traffic flow changes, and the distance between attractions V_i and V_j is calculated:

$$dist\left(V_i,V_j\right)=\frac{R\times\pi\times\arccos\theta}{180} \tag{5}$$

$$\theta=\sin\left(lat_{V_i}\right)\times\sin\left(lat_{V_j}\right)\times\cos\left(long_{V_i}-long_{V_j}\right)+\cos\left(lat_{V_i}\right)\times\cos\left(lat_{V_j}\right) \tag{6}$$

where $long_{V_i}$ and $long_{V_j}$ denote the longitude of the two attractions, lat_{V_i} and lat_{V_j} denote the latitude of the two attractions, and R is the length of the radius of the earth.

After calculating the distance between two attractions, the relative distance matrix of the attractions L [18]. The formula is as follows:

$$L=\left[1-\max\min\left(dist\left(V_i,V_1\right),\cdots,dist\left(V_i,V_j\right),\cdots,dist\left(V_i,V_n\right)\right)\right] \tag{7}$$

where $\max\min$ denotes the max-min normalization, for each element $L_{V_i,V_j}\in[0,1]$ in matrix L. The closer its value is to 1, the closer the two attractions are.

3.3.4 Transportation connectivity of rural tourist attractions

Inter-regional connectivity tends to have a mobility impact on inter-regional traffic changes, e.g., when congestion occurs in one section of a highway network, after some time the congested section is transmitted to other nearby sections. In this paper, the accessibility matrix T_{V_i,V_j} is defined to model the spatial impact between scenic areas. If V_i and V_i are directly connected through the public transportation network, the parameter of reachability $\left(V_i,V_j\right)$ is set to 1, and vice versa to 0. Traffic reachability matrix T :

$$T=\left[T_{V_i,V_j}=\begin{cases}1\\0\end{cases}\right] \tag{8}$$

4. Tourist flow forecast for rural tourist attractions

This section describes the M-GCNGRU model in four parts: model inputs, spatial dependency modeling, temporal dependency modeling, and model optimization.

4.1 Spatio-temporal dependence model inputs

The model inputs in this paper consist of a graph-structured adjacency matrix A and a scenic traffic flow matrix TF . The scenic traffic flow matrix TF consists of the average traffic volume of each scenic area at a given time. The traffic flow matrix TF is denoted as:

$$TF=\begin{array}{c} \\ t_1\\ t_2\\ \vdots\\ t_m\end{array}\begin{array}{c}\begin{array}{cccc}V_1 & V_2 & \cdots & V_n\end{array}\\ \begin{bmatrix}tf_{11} & tf_{12} & \cdots & tf_{1n}\\ tf_{21} & tf_{22} & \cdots & tf_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ tf_{m1} & tf_{m2} & \cdots & tf_{mn}\end{bmatrix}\end{array} \tag{9}$$

where TF denotes the time frame, columns denote each attraction, and elements in the matrix denote traffic flow values.

4.2 Modeling of spatial dependence

A GCN is employed to learn spatially related features between regions, capturing potential features from location proximity, transportation accessibility, popularity similarity, functional similarity, and other spatial correlations between attractions [19]. Given the adjacency matrix A and traffic flow matrix TF , the GCN constructs filters in the Fourier domain as the nodes of the graph, combined with stacking multiple convolutions to construct the GCN, which can be expressed as:

$$H^{l+1}=\sigma\left(D^{-1}\tilde{A}H^{l}\theta^{l}\right) \tag{10}$$

where D^{-1} is the inverse matrix of the degree matrix. $\tilde{A}=A+I$ is the adjacency matrix plus the unit matrix, i.e., the adjacency matrix with self-loops, which is used to represent the correlation between attractions, and the self-loop is used to indicate that each attraction has the maximum correlation with itself. H^{l} is the output of layer l, θ^{l} is the weight matrix of the layer, which is randomly initialized before model training and iteratively updated during training, and σ is the activation function of the nonlinear model [20]. In this paper, two layers of GCN are used to capture the spatial dependency, which can be expressed as:

$$f\left(TF,A\right)=\mathrm{Re}\,LU\left(D^{-1}\tilde{A}*\mathrm{Re}\,LU\left(D^{-1}\tilde{A}TFw_1\right)*w_2\right),w_1=R^{P\times Q},w_2=R^{Q\times T} \quad (11)$$

Where, w_1 is the weight matrix from the input layer to the hidden layer, P is the length of the input feature matrix, Q is the number of neurons in the current layer. w_2 is the weight matrix from the hidden layer to the output layer, and T is the predicted length of the scenic traffic flow.ReLU is a common activation function in deep neural networks, and $\mathrm{Re}\,LU\left(D^{-1}\tilde{A}TFw_1\right)$ denotes that the output of the first layer of GCN is used as the input of the second layer of GCN. The spatial dependency model is shown in Fig. 4, where the spatial feature matrix $L*T*P*F$ is multiplied as well as summed with the initialized $X\left(m,n\right)$ weight matrix $\lambda\left(m,n\right)$, to obtain the matrix H^{*}:

$$H^{*}=\lambda_1\times L^{*}+\lambda_2\times T^{*}+\lambda_3\times P^{*}+\lambda_4\times F^{*} \quad (12)$$

These four weight matrices $\lambda_1,\lambda_2,\lambda_3,\lambda_4$ are adaptively updated during model training.

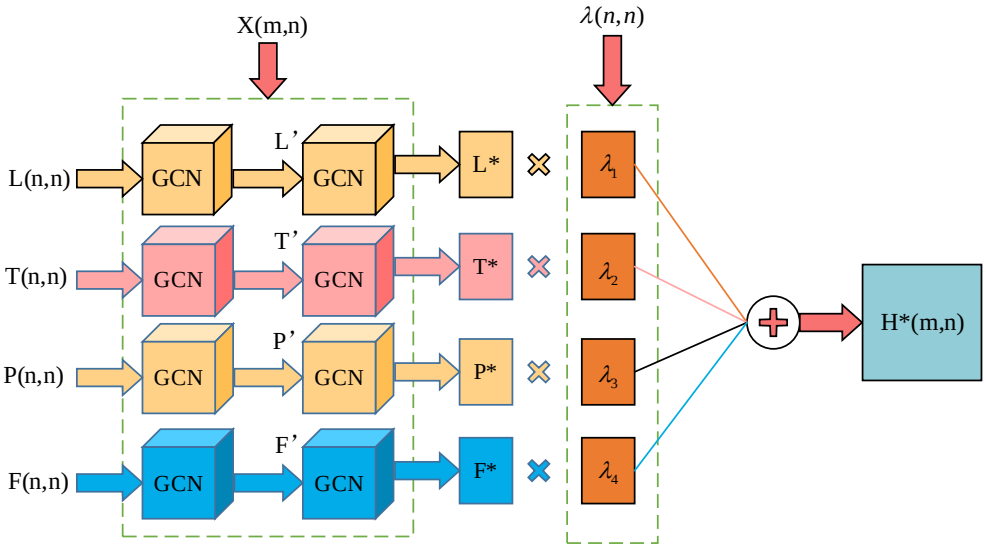

Figure 4 Spatial dependence model

4.3 Time-dependent modeling

Capturing temporal dependence is another key issue in the analysis and prediction of regional tourist behavior. The GRU with relatively few parameters is chosen to capture the temporal relationship of the scenic traffic flow sequence [21]. For the traffic flow matrix TF, TF_t representing the original traffic flow values of TF during the t time period, the corresponding output H_t can be calculated according to the following equation:

$$\begin{aligned} u_t &= \sigma\left(W_u * X_t + W_u * H_{t-1} + b_u\right) \\ r_t &= \sigma\left(W_r * X_t + W_r * H_{t-1} + b_r\right) \\ c_t &= \tanh\left(W_c * \left(r_t * H_{t-1}\right) + W_c * X_t + b_c\right) \\ H_t &= u_t * H_{t-1} + \left(1 - u_t\right) * c_t \end{aligned} \tag{13}$$

where u_t and r_t denote the update and reset gates, respectively, c_t denotes the state of the neuron, H_{t-1} denotes the output of the previous time period $t-1$, parameters W and b represent the weights and biases, $*$ denotes the convolution operation, and σ and $\tanh$ represent the activation function.

4.4 M-GCNGRU model learning optimization

The goal of model optimization is to minimize the error between the actual traffic flow and the predicted value during the training process. In this section, the loss function of the M-GCNGRU model is used to represent the real traffic flow and the predicted traffic flow using Y and respectively:

$$loss = \left\| Y_t - \hat{Y}_t \right\| + \mu L_{reg} \tag{14}$$

Where, $\left\| Y_t - \hat{Y}_t \right\|$ is the actual scenic tourist flow and the prediction error value, L_{reg} is the L regularization term, and μ is the hyperparameter. The activation function is shown as follows:

$$\begin{aligned} \sigma(x) &= \begin{cases} 0, if\ x \le 0 \\ x, if\ x > 0 \end{cases} \\ \tanh &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned} \tag{15}$$

5. Empirical Prediction of Tourist Behavior in Rural Tourism Intelligent Platform

In order to prove the practical application effect of this paper's method, the tourist user-generated content data on three websites, Ctrip.com, GoWhere.com, and MaHoneycomb.com, were acquired through customized crawler technology, and the data flow of single-user crawling data flow diagram and database batch crawling

module data flow diagram were utilized to obtain the tourists' data base of rural attractions of Huangling in Wuyuan, Jiangxi Province. The M-GCNGRU prediction model in this paper is used for data analysis to identify potential patterns and associations in the data by training a large amount of historical visitor data, including visitor behavior, feedback, and market research information. This includes the analysis of tourist behavior characteristics, tourism demand under different sentiment indices, and tourist attraction visitor volume prediction, and is compared and analyzed with the ConvLSTM model, the T-GCN model, and the ST-GCN model.

5.1 Characterization of Tourist Behavior

Table 1 shows the behavioral characteristics of tourists traveling to rural attractions, and from the point of view of tourism motives, relaxation and experiencing rural style are the main motives of tourists, occupying 32% and 35% of the historical data, respectively. The model prediction results of this paper are highly consistent with the actual survey data, confirming the accuracy of the model and showing that tourists' demand for leisure and experience in rural tourism continues to be strong. Knowledge growth and visiting relatives and friends, although not a high percentage, are still the motivation for some tourists to travel. In terms of information channels, the Internet became the main way for tourists to obtain information, accounting for 38%, reflecting the importance of the Internet in the dissemination of tourism information. Introductions from friends and relatives also accounted for a certain proportion of 15%, indicating that word-of-mouth still has a significant influence in the dissemination of tourism information. Travel agencies and outdoor advertising are also channels for tourists to obtain information, with a relatively low proportion, and the model predictions in this paper are also highly accurate with the original data.

In terms of resource preference, natural scenery and ancient villages and towns are the most popular among tourists, occupying 39% and 31% of the historical data proportion respectively, indicating that the natural landscape and historical and cultural heritage of Huangling in Wuyuan, Jiangxi Province are more attractive to tourists. Eco-agriculture and folk culture have their own specific audience groups although they account for a relatively low percentage. In terms of transportation choices, high-speed rail/motorized train and private car are the main ways for tourists to travel, occupying 25% and 22% of the proportion respectively. Travel agency chartered buses also occupy a large proportion of 38%, showing the important position of group travel in the rural tourism market. In terms of partnering, traveling with friends is the first choice of tourists, accounting for 33%, followed by traveling with family members and joining tour groups. The proportion of traveling alone is relatively low, accounting for only 11%. This shows that tourism activity is more of a social and family activity in people's mind. In terms of the number of travel days, short trips of 1-3 days and medium trips of 4-7 days were most favored by tourists, accounting for 39% and 38% respectively. This reflects that tourists are more inclined to choose the schedule with moderate time, which can be fully experienced and avoid excessive fatigue when planning their trips.The proportion of long-distance trips of 8-14 days and long-distance trips of more than 15 days is relatively low, but there is still a certain market demand. The model in this paper predicts the percentages of different brand positioning perceptions in multiple dimensions such as travel motivation,

information channel, resource preference, transportation, companionship mode and number of days of travel, which is highly consistent with the actual survey data, with an accuracy rate of up to 98%, and identifies the potential relationship between different factors and brand positioning perceptions.

Table 1 Behavioral characteristics of tourists visiting rural attractions

Behavioral characteristics	Options	History (%)	Forecast (%)	Options	History (%)	Forecast (%)
Travel motivation	Relax and relieve mental stress	32	31	Experience rural customs and atmosphere	35	36
	Visit relatives and friends and enhance relationships	15	16	Buy rural specialty products	5	6
	Visit relatives and friends and enhance relationships	8	8	Keep fit and improve physical fitness	5	5
Information channels	Introductions from relatives and friends	15	16	TV and radio advertising	12	11
	Internet	38	37	Outdoor advertising	11	11
	Travel agency	20	20	Travel books	4	5
Resource Preference	Natural scenery	39	39	Folk culture	5	5
	Ancient villages and towns	31	33	Specialties	4	4
	Ecological agriculture	22	20	Others	1	1
Means of transportation	Airplane	15	15	Private car	22	22
	High-speed rail/Electric train	25	26	Travel agency charter	38	37
Travel companions	Alone	11	11	Travel with family	27	27

	With friends	34	33	Join a tour group	28	29
Number of days for the trip	1-3 days (short trip)	39	39	8-14 days (long-distance travel)	18	18
	4-7 days (medium trip)	38	38	15 days or more (long-term travel)	5	5

5.2 Results of tourism demand under different combinations of characteristics

In order to explore the relative importance of features constructed based on data from different sources for prediction, Table 2 shows the M-GCNGRU model feature importance. The feature importance score reflects the contribution of the feature to the predictive power of the model, and the higher the score, the greater the impact of the feature on the predictive results of the model. Historical passenger volume and search volume show high importance under all time windows, proving that these two features have a significant impact on predicting the target variables. Especially under longer prediction time windows, such as h = 30, the importance scores of these two features remain high, indicating that they are also very important for long-term prediction. Affective features, including scenery affect, amusement affect, and service affect, also show some importance, especially in shorter prediction time windows, such as h=1 and h=3. This indicates that affect is informative for short-term prediction, but its influence diminishes over time. Other characteristics, such as average temperature and vacation, also have a slight influence. In particular, average air temperature has relatively high importance scores under certain time windows, suggesting that the harmful use of weather temperature is still an important factor in traveling for pleasure, and that the impact of these factors on scenic spot traffic may be more significant in a given time period.

Table 2 Feature importance of M-GCNGRU model

Features	h=1	h=3	h=7	h=15	h=30
Historical passenger flow	82	111	95	91	106
Search volume	80	96	67	95	117
Scenic sentiment	30	38	29	41	50
Play sentiment	19	22	10	21	32
Service sentiment	19	25	22	28	39
Transportation sentiment	13	19	14	13	19
Food sentiment	10	10	8	10	9
Ticket sentiment	10	15	9	10	12
Shopping sentiment	10	11	7	9	11
Accommodation sentiment	7	79	7	3	7
Average temperature	72	92	58	99	104
Holidays	13	11	14	15	15
Weather conditions	13	16	21	31	26

In order to study the differences in the prediction effect of different models, the prediction accuracy of the optimal model for each step length is shown in Fig. 5. It can be seen that the root-mean-square prediction errors of all methods show an increasing trend as the number of steps increases. Specifically, at h=1, the root mean square prediction error of the model in this paper is 0.122, which is the lowest among all methods. While the ConvLSTM model has the highest root mean square prediction error of 0.247. When the step size is increased to h = 30, the root mean square error prediction error of this paper's model rises to 0.182, which is still the lowest among all the methods, but the error increases by about 0.067 compared with that of h = 1. At the same time, the prediction errors of the ConvLSTM model, the T-GCN model, and the ST-GCN model the prediction errors of ConvLSTM model, T-GCN model and ST-GCN model also increase to 0.299, 0.254 and 0.324, respectively, with the most significant increase in the error of ST-GCN model, which reaches about 0.077. Although the RMSE prediction errors of all the methods increase with the increase of the number of hours, the model of the present paper has lower RMSE prediction errors compared with the other three methods at different prediction time spans, which shows a better prediction performance.

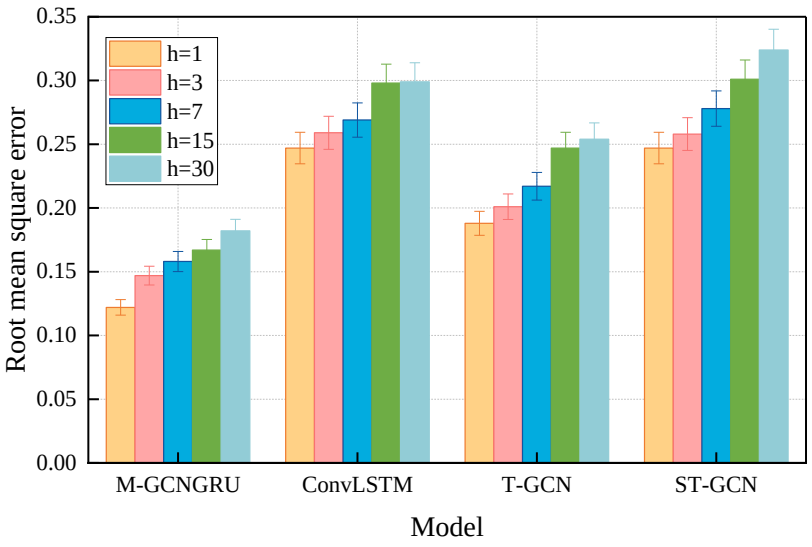

Figure 5 Prediction accuracy of the optimal model at each step size

5.3 Tourist Attraction Visitor Forecast

Figure 6 shows the results of predicting the number of visitors to rural attractions, and this paper's model is highly fitted to the actual flow of visitors, with only a little deviation, and the two maintain close consistency at most time points. In the 31st day time period, the actual flow of 5.82 million people, while the predicted value of this model is 5.81 million people, the deviation between the two is only 10,000 people. In the 56th day, the actual passenger flow is 9.74 million, while the predicted value of the model in this paper is 9.74 million. The fitting effect shows that the model and method adopted in this paper can accurately capture the trend of passenger flow and provide powerful decision support for actual tourism management and resource allocation.

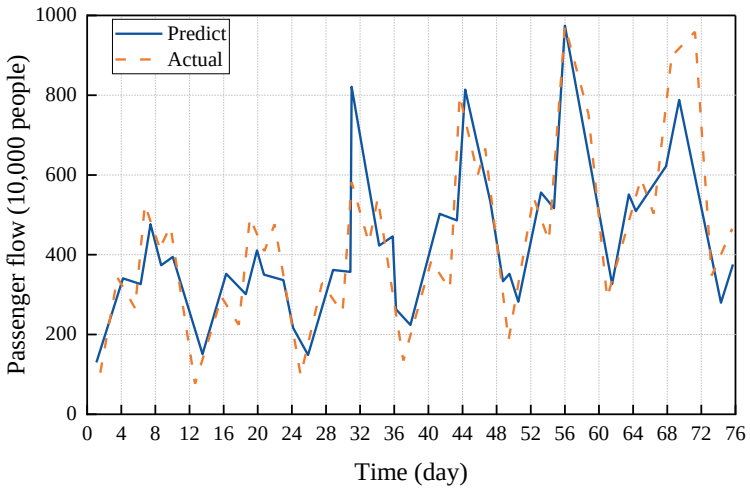

Figure 6 Forecast results of tourist volume of rural attractions

Table 3 shows the results of the 12-period forecasts from 2022.04 to 2023.03, and it can be seen that there are differences in the accuracy of the forecasts of each model in different months. The original value sequence shows the fluctuation of the actual passenger flow, from the lowest 302 to the highest 949, reflecting the trend of passenger flow over time. Comparing the other three prediction models, ConvLSTM model, T-GCN model, and ST-GCN model, the prediction results of this paper's model are the closest to the original values.The prediction results of the ConvLSTM model have a large deviation from the actual passenger flow in some months such as months 4, 5, and 6, with the absolute values of the deviation being 281, 257, and 123, respectively.The T-GCN model's The prediction results of the T-GCN model are also too far from the results of the actual values, such as the 8th and 12th months, the absolute value of the deviation is 214 and 128, respectively.The ST-GCN model predicts 722 in the 5th month, while the actual value is 749, and in the 8th month, its prediction is 899, which is a large difference from the actual value of 630. The predicted values of this paper's model at several points in time are comparable to the actual values, such as in months 4, 5, 11 and 12, the predicted values are 629, 740, 396 and 465, respectively, and all of them have a small deviation from the actual values. The forecast results are highly compatible with the actual passenger flow, with minor deviations only in a few months. Especially in September, the predicted value is 402, and the actual value is 401. In general, the prediction effect of the model in this paper is obvious to other models, which can more accurately reflect the actual changes of passenger flow.

Table 3 12-period forecast results from April 2022 to March 2023

Cycle		1	2	3	4	5	6	7	8	9	10	11	12
Original value	2022.04-2023.03	589	547	500	764	949	935	970	630	401	302	393	483
Predictive value	ConvLSTM	434	549	518	483	681	812	810	836	580	408	348	402
	T-GCN	471	642	593	618	752	744	672	617	472	350	388	455
	ST-GCN	508	557	610	623	722	796	705	899	381	297	282	373
	M-GCNGRU	584	530	540	629	740	895	927	689	402	298	396	465

6. Discussion

For intelligent travel in rural tourism, reasonable time and routes can be arranged through deep learning data, so that tourists receive better travel arrangements, avoid traffic jams, detours and other risks, reduce time costs and capital costs, and enhance the sense of travel experience. Through the integration of information on food, accommodation, transportation, tourism, entertainment and shopping, tourists can order food and choose accommodation through the platform in advance when they plan their trips, reducing the uncertainty of tourism. Through voice navigation, electronic tour guides and other detailed introduction of local customs and traditions, the tourists are more deeply integrated into the rural journey to meet the diversified needs of tourists.

7. Conclusion

This paper constructs a visitor demand and flow prediction model for multi-source heterogeneous data, with sentiment construction and M-GCNGRU prediction type as the core. The M-GCNGRU model framework integrates multiple aspects such as scenic spot feature modeling, spatio-temporal dependence modeling, and model learning optimization, which significantly enhances the accuracy and reliability of the prediction. Through the M-GCNGRU prediction model processing and analysis, key factors affecting tourists' multidimensional perceptions of travel motivation, information channels, resource preferences, transportation, companionship, and number of days of outing at Huangling rural attractions in Wuyuan, Jiangxi Province were identified. In the 31-day time period, the actual visitor flow was 5.82 million, while the predicted value of this paper's model was 5.81 million. In the 56th day, the actual passenger flow is 9.74 million, and the prediction value of this model is 9.74 million, which shows stronger stability in prediction. At the same time, the model of this paper is significantly better than other comparative algorithms in the measurement indexes of the prediction algorithm, which provides strong support for the intelligent management and decision-making of rural tourism.

REFERENCE

[1] Mukhopadhyay, K. . (2021). The global concept of sports tourism with emphasis on indian perspective. *International Journal for Modern Trends in Science and Technology*(2), 23-30.

[2] Dai, M. L., Fan, D. X., Wang, R., Ou, Y. H., & Ma, X. L. (2023). Does rural tourism revitalize the countryside? An exploration of the spatial reconstruction through the lens of cultural connotations of rurality. *Journal of Destination Marketing & Management*, *29*, 100801.

[3] Bustamante, A., Sebastia, L., & Onaindia, E. (2021). On the representativeness of openstreetmap for the evaluation of country tourism competitiveness. *ISPRS International Journal of Geo-Information*, *10*(5), 301.

[4] Fountain, J., Charters, S., & Cogan-Marie, L. (2021). The real Burgundy: negotiating wine tourism, relational place and the global countryside. *Tourism Geographies*, *23*(5-6), 1116-1136.

[5] Yang, J., Yang, R., Chen, M. H., Su, C. H. J., Zhi, Y., & Xi, J. (2021). Effects of rural revitalization on rural tourism. *Journal of Hospitality and Tourism Management*, *47*, 35-45.

[6] Li, D., Deng, L., & Cai, Z. (2020). Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms. *Personal and Ubiquitous Computing*, *24*(1), 87-101.

[7] Carrese, S., Cipriani, E., Crisalli, U., Gemma, A., & Mannini, L. (2021). Bluetooth traffic data for urban travel time forecast. *Transportation Research Procedia*, *52*, 236-243.

[8] Li, S. , & Lai, L. . (2022). Personalized recommendation algorithm for intelligent travel service robot based on big data.

[9] Husain, Zarlis, M. , Wahyudi, M. , Santoso, H. , Sihotang, H. T. , & Fitriani, N. . (2021). Smart tourims: decision support systems in the strategic location of inn/hotel for travelers using the naive bayes algorithm. Journal of Physics: Conference Series, 1830(1), 012012 (9pp).

[10] Hamid, R. A., & Croock, M. S. (2020). A developed GPS trajectories data management system for predicting tourists' POI. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *18*(1), 124-132.

[11] Lee, G. H., & Han, H. S. (2020). Clustering of tourist routes for individual tourists using sequential pattern mining. *The Journal of Supercomputing*, *76*(7), 5364-5381.

[12] Gao, Y. (2021). Forecast model of perceived demand of museum tourists based on neural network integration. *Neural Computing and Applications*, *33*, 625-635.

[13] Sinclair, M., Mayer, M., Woltering, M., & Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management*, *263*, 110418.

[14] Carlson, S. J., Levine, L. J., Lench, H. C., Flynn, E., Winks, K. M., & Winckler, B. E. (2023). Using emotion to guide decisions: the accuracy and perceived value of emotional intensity forecasts. *Motivation and Emotion*, *47*(4), 608-626.

[15] Shi, X. (2020). Tourism culture and demand forecasting based on BP neural network mining algorithms. *Personal and Ubiquitous Computing*, *24*(2), 299-308.

[16] Martínez, J. M. G., Martín, J. M. M., & Rey, M. D. S. O. (2020). An analysis of the changes in the seasonal patterns of tourist behavior during a process of economic recovery. *Technological Forecasting and Social Change*, *161*, 120280.

[17] Liu, Z., Wang, A., Weber, K., Chan, E. H., & Shi, W. (2022). Categorisation of cultural tourism attractions by tourist preference using location-based social network data: The case of Central, Hong Kong. *Tourism Management*, *90*, 104488.

[18] Tian, C., & Peng, J. (2020). An integrated picture fuzzy ANP-TODIM multi-criteria decision-making approach for tourism attraction recommendation. *Technological and Economic Development of Economy*, *26*(2), 331-354.

[19] Liao, Z., Zhang, L., & Liang, S. (2023). Spatio-temporal pattern evolution of China’s provincial tourism efficiency and development level based on DEA-MI model. *Scientific Reports*, *13*(1), 20227.

[20] Lan, Y., Wang, X., & Wang, Y. (2021). Spatio-temporal sequential memory model with mini-column neural network. *Frontiers in Neuroscience*, *15*, 650430.

[21] Martins, M. R., da Costa, R. A., & Moreira, A. C. (2022). Backpackers' space–time behavior in an urban destination: The impact of travel information sources. *International journal of tourism research*, *24*(3), 456-471.

ABOUT THE AUTHOR

Siyang Wang was born in Beiliu，Guangxi. China, in 1986. She obtained a master's degree from the University of Perpignan in France. She currently working at the college of Vocational and Technical, Guangxi Normal University,. Her main research direction is tourist behavior and co-creation of tourism value.

E-mail: emiliedida666@126.com