J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 7441--7458

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Digital Tour Route Planning for Historic Neighborhoods Driven by the Combination of BD and Intelligent Algorithms

Fangfei Bi[1,2,✉], Zhao Wang[1], Baogang Lin[2]

[1] *School of Urban Planning and Municipal Engineering, Xi'an Polytechnic University, Xi'an, 710048, Shaanxi, China*

[2] *School of Architecture, Xi'an University of Architecture and Technology, Xi'an, 710043, Shaanxi, China*

ABSTRACT

With the development of big data (BD) technology, tourism route planning of historical blocks relies on a large amount of real-time data. The existing research data sources are limited and difficult to integrate, which cannot meet the personalized needs of tourists. This paper combined BD and intelligent algorithms to realize personalized tourism route planning of historical blocks. By collecting tourists' behavioral data, scenic spot spatial data and real-time traffic information, the paper built tourist portraits and used the neural collaborative filtering algorithm to make personalized scenic spot recommendations. It used genetic algorithms (GAs) to optimize routes, taking into account factors such as tourists' interests, distances between scenic spots, and traffic conditions. With the help of the real-time data streaming platform Apache Kafka, the paper dynamically adjusted routes to deal with sudden traffic or crowded attractions, thereby improving the tourist experience. The experimental results analyze the consumption preferences and behavioral characteristics of different tourists. Tourist 1002 spent 500 yuan on shopping, and high-end shopping malls and food courts were recommended for him. Tourist ID 1005 preferred "snacks and coffee" in terms of dining, and showed no interest in souvenir consumption. This tourist preferred to stay in leisure places for a longer time rather than a compact travel route. The neural coordination filtering algorithm + GA performed well in terms of total travel time of 4.2 hours, total walking distance of 7.8 kilometers, and traffic congestion coefficient of 0.35, which was better than other algorithms, showing its significant advantages in digital tourism route planning in historical blocks. This method combines BD and intelligent algorithms to improve the tourist experience through personalized recommendations and route optimization, optimize the traffic management of scenic spots, flexibly respond to emergencies, promote the intelligent and refined management of historical district tourism, and provide innovative ideas for future tourism route planning.

Keywords: Big Data Technology, Tourism Route Planning, Personalized Recommendation, Genetic Algorithm, Tourist Behavior Analysis

✉ Corresponding author.
E-mail address: 20150610@xpu.edu.cn (F. Bi).

Received 03 March 2024; Revised 20 April 2024; Accepted 10 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-406

1. Introduction

With the rapid development of China's economy and the increasing standard of living of the general public, people in the pursuit of material life at the same time, the demand for spiritual life is also more and more important, making people's leisure time for reading, fitness, vacation travel and other needs more and more exuberant [1-3]. In the past, the way of traveling with the group in the choice of tourist attractions and attractions route planning can only provide a fixed, uniform tourist routes, can not provide tourists with a good independent choice of space [4-5]. And the transformation of consumer demand and the booming development of self-driving tours make people pay more and more attention to the sense of comprehensive tourism experience, such as smooth travel routes, rich attraction excursion activities, and the least cost of self-driving tours [6-7]. In addition, due to the more attractions and shorter time, planning an optimal route with the shortest distance is an indispensable good thing for both consumers and scenic spots [8-9]. A reasonable tourist route can spend the least time cost to link up various tourist attractions to form a complete tour route, so as to get the maximum viewing experience [10-12].

China's continuous promotion of cultural construction has triggered a wave of revival of historical and cultural neighborhoods. As the main expression of urban cultural heritage, historical and cultural neighborhoods focus on the regional culture and image characteristics of the city and carry the collective memories of each generation [13-15]. The atmosphere created by the authenticity feeling of historical and cultural neighborhoods allows foreign tourists to feel the cultural heritage of the city from the city's historical and cultural neighborhoods [16]. Therefore, in the context of the rapid development and wide application of cutting-edge technologies such as Internet technology and traffic detection technology, the use of digital technology to develop tourism planning routes can help tourists to obtain a more convenient and economical tourism experience.

A large number of scholars have combed through the current existence of geotourism route planning methods. Chen, B. H. et al. showed that automatic trip planning based on points of interest (POIs) has become mainstream, and based on this, they proposed a three-intelligence deep reinforcement learning method in combination with Bayesian model to further optimize the planning results by considering the connection between destination exploration, user preferences and attractions [17]. Zhou, X. et al. examined the application of a machine learning model for Naïve Bayesian interest data mining in tourism route planning, which was trained on a massive dataset of tourists' interests and needs, and was able to accurately and efficiently generate optimal tourism routes that maximize tourists' motivational benefits and travel experiences [18]. Dezfouli, M. B. et al. used information data such as cost, transportation, weather as well as time as parameters and input them into the constructed tourism planning model of user needs, combined with the selection of user's points of interest, to generate the optimal tourism visit paths using big data analysis [19]. Zhou, X. et al. used a decision tree model to optimally solve the constructed tourist interest field mapping model and geospatial information, which is an effective way based on massive tourism data in order to achieve tourism path planning [20]. Lim, K. H. et al. used POI popularity and user interest preferences etc. to form travel constraints and built a travel recommendation model and proposed a PersTour algorithm to solve it in order to automatically derive real travel sequences [21]. Zhou, X. et al. analyzed the feasibility and superiority of mining tourism route planning algorithms based on the precise interest data of attractions, and established attraction clustering models and tourist interest models respectively, so that the optimal tourism route planning algorithm based on interest labels generates attraction visit routes that meet the satisfaction level of tourists [22].

This study aims to optimize the tourism route planning of historical blocks and introduces a digital route planning method based on BD technology and intelligent algorithms. A variety of data sources, including smartphone apps, Wi-Fi tracking systems, sensors, and social media platforms, are used to comprehensively collect tourist behavior data. The NCF algorithm is used to analyze tourist historical behavior, generate personalized tourist portraits, and recommend attractions based on interests and consumption preferences. Combined with GAs to optimize the sequence of attractions, the distance between attractions and traffic conditions are integrated to design the optimal travel route and enhance the tourist experience. The real-time data streaming platform Apache Kafka is also introduced to deal with sudden traffic congestion or crowded attractions, ensuring dynamic adjustment and optimization of the route. Through these steps, BD technology and intelligent algorithms are closely linked to the theme of historical district tourism route planning, providing accurate customized travel routes, and real-time data effectively responding to traffic or crowded attractions.

2. Historical Block Tourism Route Planning Based on BD

2.1. User Behavior Data Collection and Analysis

In order to achieve personalized travel route recommendations for historical blocks, tourist behavior data is obtained through multiple channels such as smartphone APP, Wi-Fi tracking system, sensors and social media platforms. Smartphone apps record location, travel trajectory, and duration of stay, and Wi-Fi tracking locates paths and stop points through signal strength and MAC addresses (Media Access Control Address). Traffic sensors monitor the flow of tourists in real time and enrich behavioral information. Social media platforms such as Instagram and Weibo provide tourists with sharing content, supplementing the data on attractions and behaviors. It can integrate multi-source data, deeply analyze tourist behavior, generate portraits, and use them to recommend attractions and optimize route planning.

Data cleaning and preprocessing are important steps after data collection. The original data contains noise, outliers, and data from different sources have inconsistent formats, which need to be sorted and corrected. The Z-score and IQR (Interquartile Range) methods are used to remove outliers and avoid the impact of noise data on subsequent analysis. Aligning the timestamps of data from different devices and sensors ensures that the time points in the data set are unified, which facilitates subsequent analysis and processing. Normalizing values of different scales, such as GPS (Global Positioning System) coordinates and Wi-Fi signal strength, maps the data to the same scale to ensure that comparative analysis can be performed under the same standard. When dealing with missing data, KNN (K-Nearest Neighbor Interpolation) interpolation method or linear interpolation method is used to fill in the missing data to ensure the integrity of the data. Figure 1 is a diagram of tourist behavior data and processing flow.

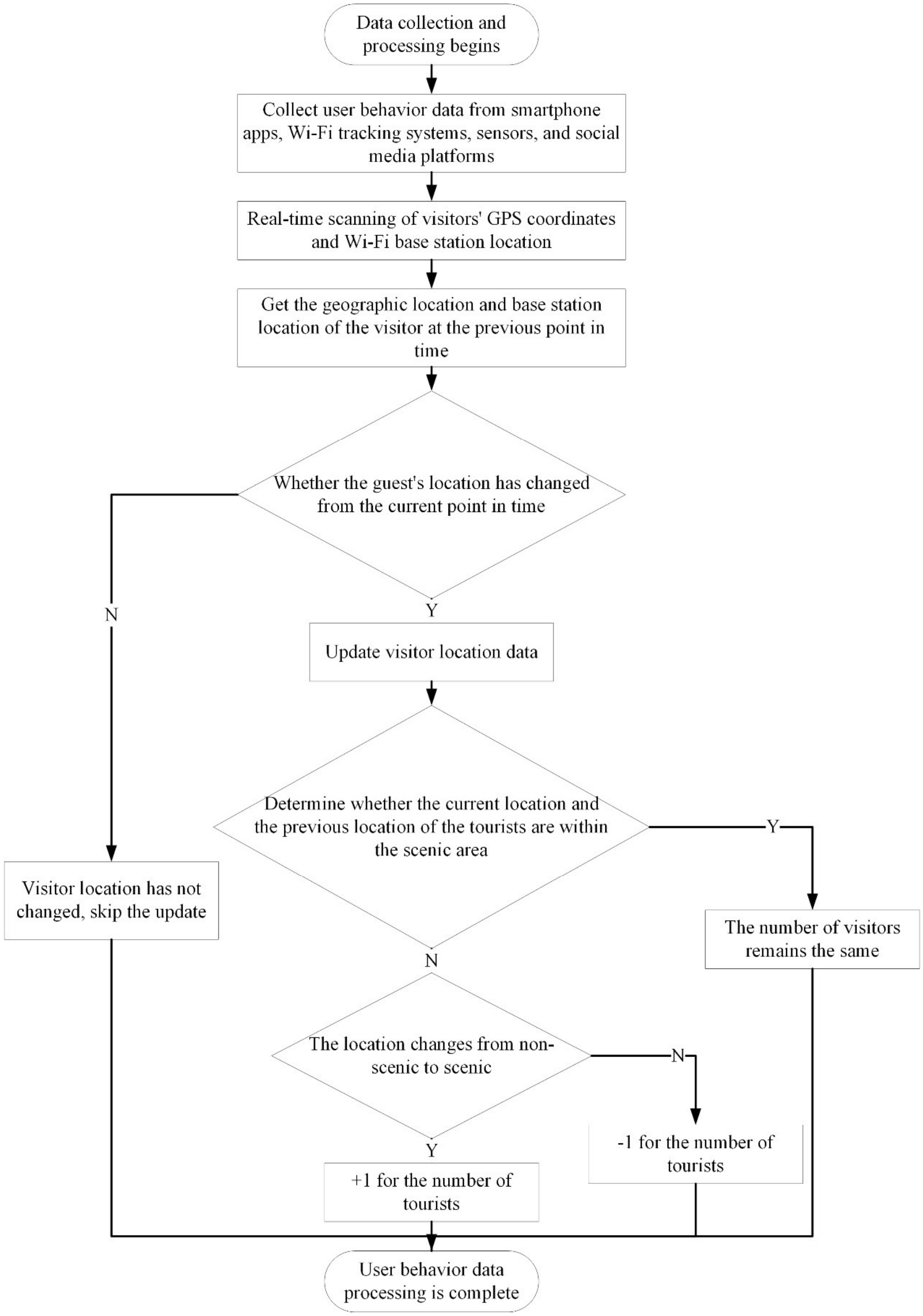

Fig. 1. Tourist behavior data and processing flow chart

Through the data processing flow in Figure 1, tourist behavior data can be analyzed comprehensively and efficiently, providing solid data support for personalized recommendation and optimization of historical district tourism routes. After data cleaning, behavioral feature extraction is performed. The tourists' trajectory data is analyzed to extract the tourists' interest characteristics, including indicators such as length of stay and frequency of scenic spot visits. The length of stay is an important criterion for judging tourists' interest in scenic spots. The longer the stay, the more attractive the scenic spot is to tourists. Analyzing the length of stay and frequency of visits of tourists at different scenic spots can identify tourists' points of interest and provide data support for personalized recommendations.

In order to further accurately identify tourists' interests and behavior patterns, a cluster analysis method is used. The K-means clustering algorithm and the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm are used to group tourists' behaviors. The K-means algorithm classifies tourists with similar interests into one category by calculating the Euclidean distance of tourists' interest characteristic, thereby identifying the typical behavior patterns of different tourist groups. K-means clustering algorithm formula:

$$J = \sum_{i=1}^{K} \sum_{x_j \in C_i} x_j - \mu_i^2 \tag{1}$$

J is the total error metric, K is the number of clusters, x_j is the data point of cluster C_i, and μ_i is the center of cluster C_i. The optimal K value is determined by the elbow rule to ensure the rationality of the clustering results. For data with uneven density or containing outliers, the DBSCAN algorithm can be used to better handle it and avoid interference from noise data.

Feature selection methods such as information gain and chi-square test are used to extract key features from tourist behavior data. Information gain can measure the impact of a certain feature on the classification of tourist interest categories, calculate the degree of reduction of uncertainty for each feature, and help screen out the most recognizable interest features. Information gain calculation formula:

$$InfoGain(T, A) = H(T) - \sum_{\upsilon \in Values(A)} \frac{|T_\upsilon|}{|T|} H(T_\upsilon) \tag{2}$$

T is a data set, A is a feature, $H(T)$ is the entropy of data set T, $H(T_\upsilon)$ is the subset entropy when feature A takes the value υ, and $\frac{|T_\upsilon|}{|T|}$ is the proportion of the subset in data set T_υ. The chi-square test is used to analyze the correlation between discrete features and further enhance the understanding of tourist behavior patterns. The results of feature extraction and selection can be directly used as input for user portraits in the NCF algorithm and GA to provide tourists with accurate personalized recommendations. Chi-square test formula:

$$x^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

O_i is the observed frequency, E_i is the expected frequency, and n is the number of categories. The chi-square statistic measures the difference between the observed frequency and the expected frequency. Through cluster analysis and feature selection, a personalized portrait of each tourist is constructed, which contains information on multiple dimensions such as their interest type, visit frequency, and length of stay. Figure 2 is a diagram of the tourist portrait construction model.

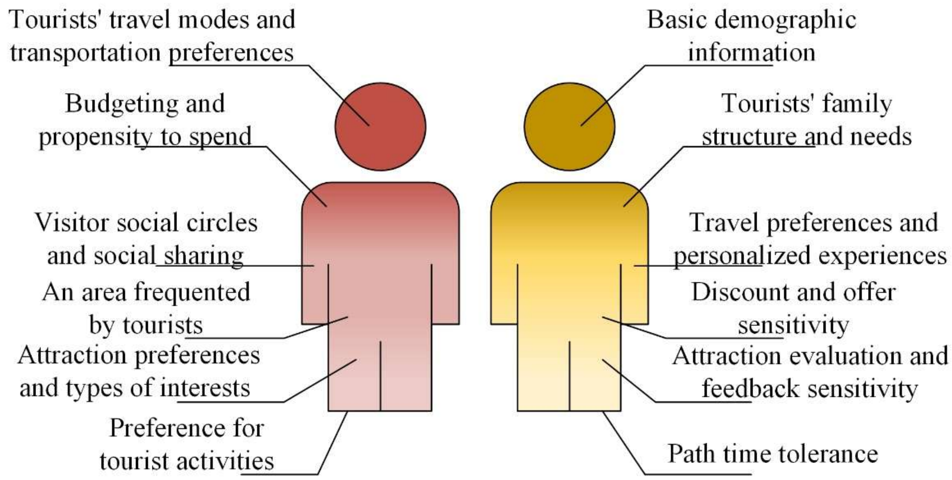

Fig. 2. Tourist portrait construction model diagram

Figure 2 shows the various features extracted from tourists' behavior data. Through cluster analysis and feature selection methods, a personalized tourist portrait is constructed. The portrait provides a solid data foundation for the subsequent personalized attraction recommendation and route planning based on the NCF algorithm. It can accurately customize the travel route that meets the needs of tourists according to their behavioral characteristics and preferences, and improve the overall experience of tourists.

2.2. Collection of Spatial Data of Historical Blocks

To realize the historical block tourism route planning based on BD, it is necessary to collect the spatial data of the blocks, build the spatial framework and dynamic variables required for route planning, and provide a basis for the construction of tourist behavior models and personalized route recommendations.

Spatial data collection relies on BD GIS (Geographic Information System) technology. The GIS platform obtains basic spatial data of historical blocks, including the geographic coordinates of scenic spots, opening hours, tourist flow data, block traffic information, etc. The geographic coordinates of scenic spots are accurately obtained through the GPS positioning system and used to locate the location of scenic spots in route planning. The opening hours of scenic spots and the flow of tourists are provided by scenic spot management systems and flow monitoring equipment, which play an important role in dynamically adjusting tourist routes and avoiding overcrowding of scenic spots. Block traffic information is obtained through public transportation API (Application Programming Interface) or real-time traffic monitoring system to provide support for traffic optimization in route planning. Table 1 is the basic spatial data of historical district attractions.

Table 1. Basic spatial data of historical district attractions

Attraction ID	R1	R2	R3
2001	(39.9075, 116.3972)	09:00 - 18:00	Bus,Subway
2002	(39.9060, 116.3956)	08:30 - 17:30	Bus,Private car
2003	(39.9082, 116.3985)	10:00 - 21:00	Bus,Subway
2004	(39.9102, 116.4010)	Open 24 hours	Bus,Private car
2005	(39.9090, 116.4005)	09:30 - 17:30	Bus,Subway
2006	(39.9048, 116.3920)	08:00 - 19:00	Bus,Private car
2007	(39.9048, 116.3920)	08:00 - 19:00	Bus,Private car
2008	(39.9075, 116.3972)	09:00 - 18:00	Bus,Subway
2009	(39.9102, 116.4010)	Open 24 hours	Bus,Private car
2010	(39.9090, 116.4005)	09:30 - 17:30	Bus,Subway

Table 1 is the basic spatial information of the attractions in the historical district. R1, R2, and R3 represent the geographic coordinates, opening hours, and available travel modes, respectively. The attractions with IDs from 2001 to 2010 are the History Museum, Ancient Architectural Sites, Shopping Centers, Parks, Art Museums, Observatories, Observation Towers, Cultural Plazas, Water Parks, and Modern Art Museums. The path is dynamically adjusted according to the actual situation. The collected spatial data needs to be standardized to provide data for the path planning algorithm. The conversion formulas are:

$$x' = \frac{(x-x_0)\cdot\cos(\theta)-(y-y_0)\cdot\sin(\theta)}{R} \tag{4}$$

$$y' = \frac{(x-x_0)\cdot\sin(\theta)+(y-y_0)\cdot\cos(\theta)}{R}\text{z} \tag{5}$$

x', y' are the coordinates in the target coordinate system, x, y are the coordinates in the original coordinate system, θ is the rotation angle, x_0, y_0 are the far points of the coordinate system, and R is the scaling factor. Data formats from different sources need to be converted to a unified format. The data is converted into standard formats such as GeoJSON or Shapefile, which can store the geometric information and attribute data of spatial data. Through format conversion, the uniformity of spatial data is ensured, providing more efficient and accurate data support for subsequent path planning operations. GIS tools are also used to visualize the data, showing information such as the distribution of scenic spots, transportation networks, and tourist flows in historical districts, providing researchers and developers with intuitive block spatial layouts and helping with route planning decisions.

To store and manage large amounts of spatial data, the spatial database PostGIS is used to store formatted data. PostGIS is a spatial extension based on the PostgreSQL database that can efficiently store and query spatial data. Spatial data query and calculation formula:

$$d = 2r\cdot\arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right)+\cos(\phi_1)\cdot\cos(\phi_2)\cdot\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{6}$$

d is the spherical distance between the bright spots, r is the radius of the earth, ϕ_1 and ϕ_2 are the latitudes of the two scenic spots, $\Delta\phi$ is the latitude difference, and $\Delta\lambda$ is the longitude difference. Spatial data in the spatial database is stored as a geometric data type and managed in combination with attribute data. This database structure provides strong support for subsequent spatial query and analysis. In the process of route planning, the spatial query function of PostGIS can be used to quickly find other attractions, transportation routes or rest areas within a certain range of the scenic spot, and recommend more suitable tour routes for tourists. PostGIS also supports real-time updating of spatial data, ensuring that the spatial data used in the system is always up to date, further improving the real-time and accuracy of route planning. In order to cope with the dynamic changes in traffic conditions and tourist flow in historical districts, the stream data processing platform Apache Kafka is used to collect and process real-time data. The system can obtain real-time data from sensors, social media or traffic monitoring platforms in a timely manner to ensure that the data used in the route planning process is the latest, so as to dynamically adjust the tourist's travel route to cope with sudden traffic congestion or overcrowding of attractions.

2.3. Personalized Attraction Recommendation Based on NCF Algorithm

In order to achieve personalized recommendations for attractions in historical districts, this study constructs a neural collaborative filtering model based on tourists' historical behavior data to capture the complex nonlinear relationship between tourists and attractions and optimize the recommendation effect. By analyzing tourists' historical behavior data, key features such as interest type, visit preference, and attraction rating are extracted. The behavior records of tourists and attractions are mapped into a user-item interaction matrix. The values in the matrix represent the tourists' preference for the attractions. To ensure the independence of model training, the data set is divided into training set, validation set and test set in a ratio of 8:1:1. On this basis, the tourist ID and attraction ID are One-Hot encoded and mapped to a low-dimensional feature space through an embedding layer to generate training data. The embedding layer vector dimensionality reduction formula is:

$$e_u = W_U \cdot OneHot(u), e_z = W_Z \cdot OneHot(z) \tag{7}$$

Among them, $e_u \in \mathrm{R}^d$ and $e_z \in \mathrm{R}^d$ are the embedding vectors of users and attractions, $W_U \in \mathrm{R}^{d\times|U|}$ and $W_Z \in \mathrm{R}^{d\times|Z|}$ are embedding matrices, and OneHot(u) and OneHot(z) are the One-Hot codes of users and attractions, respectively. When designing the model, the NCF model combines deep neural networks and collaborative filtering methods, including embedding layers, interaction layers, fully connected layers, and output layers. The embedding layer encodes the IDs of tourists and attractions into low-dimensional embedding vectors to capture the potential features of tourists and attractions. The interaction layer realizes the interaction between tourists and attraction features through vector concatenation operations and inputs the results into the fully connected layer. The fully connected layer is composed of multi-layer perceptrons, which learn nonlinear feature relationships through the ReLU activation function. The output layer generates a recommendation score through the Sigmoid function. The score range is [0, 1], which represents the probability of tourists' preference for attractions. The specific calculation formula of the multi-layer perceptron is:

$$\gamma^{(l)} = \sigma\left(M^{(l)}\gamma^{(l-1)} + b^{(l)}\right), \forall l \in \{1,2,...,L\} \tag{8}$$

$\gamma^{(l)}$ is the output of layer l, $M^{(l)}$ and $b^{(l)}$ are the weight matrix and bias of layer l respectively. The model training uses the cross entropy loss function, which can measure the difference between the predicted value and the actual value, thereby optimizing the prediction ability of the model. The formula of the cross entropy loss function is:

$$\zeta = -\frac{1}{N}\sum_{(m,n\in\xi)}\left(\rho_{mn}\cdot\left[\log\left(\hat{y}_{mn}\right)+\left(1-y_{mn}\right)\log\left(1-\hat{y}_{mn}\right)\right]\right) \tag{9}$$

ρ_{mn} is the sample weight, y_{mn} is the actual value, $\hat{y}_{mn}$ is the predicted value, and ξ is the training sample set. Figure 3 is the structure diagram of the NCF model.

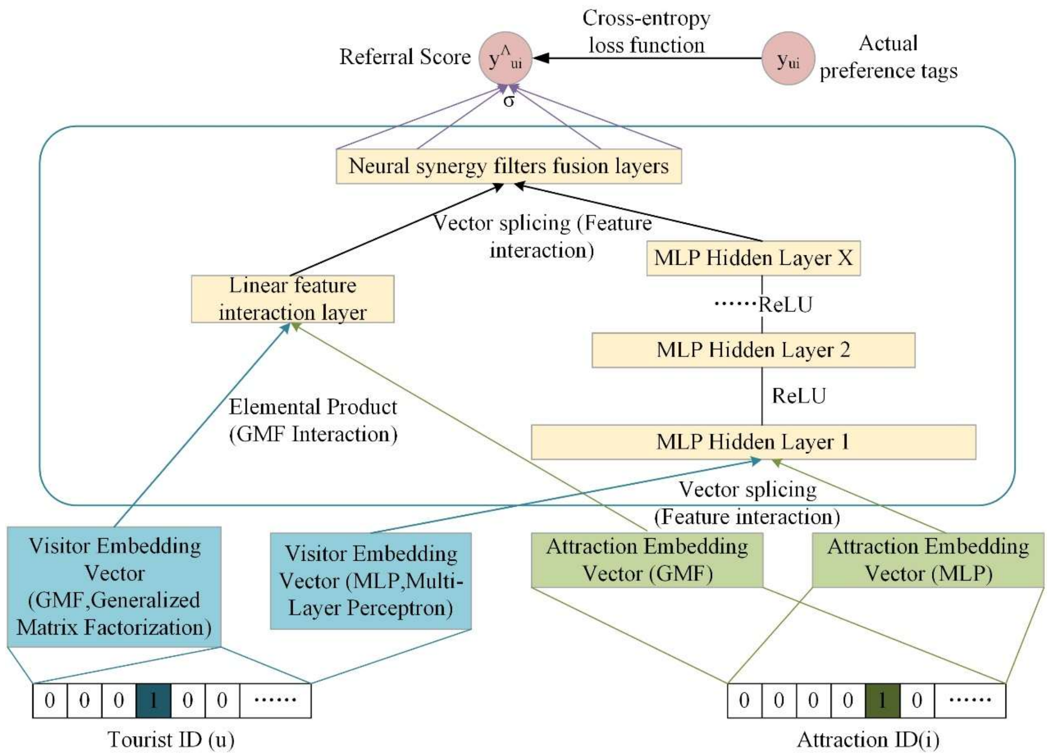

Fig. 3. NCF model structure diagram

Figure 3 shows the core architecture and data flow of the NCF model. This model effectively solves the limitations of traditional collaborative filtering models in sparse data and linear assumptions through the strong expressive power of deep neural networks, and provides technical support for achieving high-quality personalized recommendations. To achieve efficient training, the Adam optimization algorithm is used to update the model parameters. Adam's adaptive learning rate mechanism can accelerate convergence and avoid local optimality. During the training process, positive and negative interaction instances are randomly sampled, where positive samples are actual records of tourist-attraction interactions, and negative samples are generated through a negative sampling strategy to enhance the model's ability to distinguish. In actual training, a small batch gradient descent optimization model is used, with each batch containing a fixed number of positive and negative samples. To improve model performance, grid search is used to optimize hyperparameters, including embedding vector dimension, learning rate, and batch size. The combined effects of different parameters are evaluated on the validation set, and the optimal configuration is selected. After training, performance evaluation is performed on the test set, and indicators such as Precision@K and Recall@K are used to measure the recommendation effect of the model. Precision@K and Recall@K evaluate the accuracy and coverage of the top K attractions in the recommendation list. The calculation formulas are:

$$\mathrm{Pr}ecision@K = \frac{|\alpha_s \cap \beta_s|}{K} \tag{10}$$

$$\mathrm{Re}call@K = \frac{|\alpha_s \cap \beta_s|}{|\beta_s|} \tag{11}$$

α_s is the first K attractions in the recommendation list, and β_s is the set of attractions that the user is actually interested in. According to the trained model, the behavior data of tourists is input into the test set to predict the preference scores of tourists for all candidate attractions, and the scores are arranged in descending order to generate a personalized recommendation list. The NCF model can effectively overcome the limitations of the traditional collaborative filtering model in terms of data sparsity and linear assumptions, and provide accurate scenic spot recommendation support for the tourism route planning of historical blocks.

2.4. Path Optimization Combined with GA

The core goal of path optimization is to generate an optimal tourism route in the historical block according to the personalized needs of tourists. Combined with the sequence of scenic spots recommended by the NCF algorithm, the GA is used to further optimize the path. Combined with the sequence of attractions recommended by the NCF algorithm, the GA is used to further optimize the path. The entire path optimization process includes four steps: path initialization, fitness function design, genetic operation, and path selection output.

A preliminary travel path is generated based on the sequence of attractions recommended by the NCF algorithm. Assuming that tourists have been recommended a set of attractions, the path initialization constructs the initial path by arranging the attractions in the recommended order. After the initial path is generated, the fitness function needs to be introduced to evaluate the quality of the path. The fitness function is the core of the GA, which comprehensively considers factors such as tourists' interests, distances between attractions, traffic conditions, and tourist flow to evaluate the path. The goal of the fitness function is to minimize travel time, distance between attractions, traffic congestion, and maximize tourists' interest and attraction. Let path P be the order in which tourists visit attractions from the starting point to the end point. The fitness function F(P) can be calculated by the following formula:

$$F\left(P\right)=\omega_1\cdot T\left(P\right)+\omega_2\cdot D\left(P\right)+\omega_3\cdot C\left(P\right)-\omega_4\cdot I\left(P\right) \tag{12}$$

$T\left(P\right)$ is the total travel time of path P, $D\left(P\right)$ is the total distance between attractions, $C\left(P\right)$ is the traffic congestion in the path, $I\left(P\right)$ is the interest value of tourists in the attractions, and ω_1, ω_2, ω_3, ω_4 is the weight coefficient, which indicates the impact of different factors on the fitness of the path. The total travel time $T\left(P\right)$ of the path represents the sum of the travel time between attractions, and the traffic conditions need to be taken into account. The calculation formula of travel time $T\left(P\right)$ is:

$$T\left(P\right)=\sum_{i=1}^{n-1}\frac{r_{i,i+1}}{s_{i,i+1}} \tag{13}$$

$r_{i,i+1}$ is the distance between attractions i and attractions $i+1$, and $s_{i,i+1}$ is the traffic speed between the two attractions. The distance $D\left(P\right)$ between attractions in the path is an important factor in path optimization, which represents the total distance that tourists walk during the trip. The formula is:

$$D\left(P\right)=\sum_{i=1}^{n-1}r_{i,i+1} \tag{14}$$

The degree of traffic congestion $C\left(P\right)$ considers the impact of real-time traffic data and is calculated based on traffic flow $f_{i,i+1}$ and congestion coefficient α:

$$C\left(P\right)=\sum_{i=1}^{n-1}\alpha\cdot f_{i,i+1} \tag{15}$$

In path optimization, GAs continuously evolve and optimize paths through selection, crossover, and mutation operations. The selection probability $P_{select}\left(P\right)$ is determined based on the value of the fitness function:

$$P_{select}\left(P\right)=\frac{F\left(P\right)}{\sum_{i=1}^{N}F\left(P_i\right)} \tag{16}$$

N is the number of paths in the current population, and P_i is the i-th path. By performing crossover operations, partial genes of two parental pathways are exchanged to generate new offspring pathways. There are paths $P_1=\left(p_1,p_2,\ldots,p_n\right)$ and $P_2=\left(q_1,q_2,\ldots,q_n\right)$, and the crossover operation generates a new path P_3 by exchanging some genes in the path:

$$P_3=\left(p_1,p_2,\ldots,p_k,q_{k+1},\ldots,q_n\right) \tag{17}$$

K is the intersection point. Mutation operation is the small random adjustment of genes in a pathway to increase diversity and avoid premature maturation. The mutation operation is achieved by randomly swapping the positions of two scenic spots in the path, and the mutated path P_4 is:

$$P_4=\left(p_1,p_2,\ldots,p_i,p_j,\ldots,p_n\right) \tag{18}$$

i and j are two randomly selected scenic spot locations. Genetic operation iteratively optimizes the path, selecting the path with the highest fitness value as the optimal path output. This path can maximize the personalized needs of tourists, reflected in factors such as maximizing interest, minimizing time, and optimizing transportation, ensuring that tourists can efficiently and joyfully complete their travel itinerary. The selection of the optimal path is based on the following formula:

$$P_{optimal}=\arg\max PF\left(P\right) \tag{19}$$

2.5. Real Time Data Stream Integration and Dynamic Path Adjustment

The optimization of tourist routes in historic districts not only needs to consider the personalized needs of tourists, but also needs to respond promptly to sudden traffic congestion, scenic spot congestion and other factors. Therefore, real-time data collection and route adjustment are indispensable parts of the optimization process. Real time collection of traffic data, weather changes, and tourist attraction congestion can be achieved through sensors, mobile devices, and other means. The real-time stream processing of sensor data typically relies on the stream processing platform Apache Kafka, which is capable of collecting and processing real-time stream data from different data sources within the neighborhood. Real time traffic data $T_{real}\left(t\right)$ changes over time, representing the traffic conditions at a specific time t. By using traffic flow monitoring devices, the congestion level of each traffic route can be obtained, and the data can be used for path adjustment. Weather changes $W\left(t\right)$ can also affect tourists' travel decisions, and real-time monitoring of weather conditions is necessary to ensure that the impact of weather is taken into account in path planning. The weather changes can be represented by the following formula:

$$W\left(t\right)=f\left(T_{weather},P_{precipitation}\right) \tag{20}$$

$T_{weather}$ represents temperature, $P_{precipitation}$ represents precipitation. After real-time data collection, tourists' travel routes can be dynamically adjusted based on the latest data. When there is congestion at tourist attractions or severe traffic jams, adjust the route based on real-time traffic data to allow tourists to avoid high traffic areas. After the path adjustment is completed, the system further generates new path recommendations based on real-time data, dynamically updates the fitness function, and quickly optimizes the path.

3. Experimental Results

3.1. *Clustering of Tourist Behavior Data*

In depth research on tourist consumption behavior, with a focus on analyzing the consumption patterns of different tourist groups in the scenic area and exploring their behavioral characteristics. Figure 4 shows the consumption data of tourists at the scenic spot, and Table 2 shows their consumption preferences.

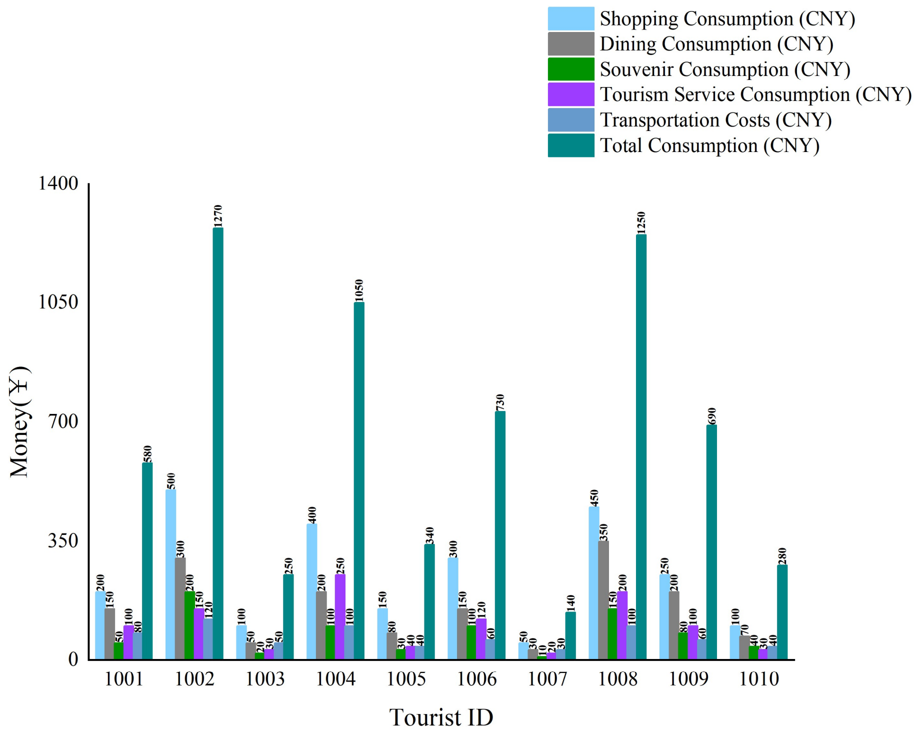

Fig. 4. Tourist Consumption Data

Table 2. Tourist Consumption Behavior Preferences

Tourist ID	R4	R5	R6	Tourist ID	R4	R5	R6
1001	Clothing, Accessories	Fast Food	1	1006	Souvenirs, Clothing	Fast Food, Full Meal	1
1002	Electronics, Accessories	Buffet, Full Meal	1	1007	Small Accessories, Souvenirs	Snacks	1
1003	Accessories, Souvenirs	Snacks	1	1008	Electronics, Accessories	Buffet, Full Meal	1
1004	Clothing, Souvenirs	Chinese Food, Full Meal	1	1009	Souvenirs, Accessories	Snacks, Coffee	0
1005	Small Accessories, Snacks	Snacks, Coffee	0	1010	Accessories, Souvenirs	Snacks, Fast Food	1

Figure 4 and Table 2 show the consumption behavior data of tourists. R4, R5, and R6 respectively represent product preferences, dining preferences, and whether tourists purchase souvenirs. 1 represents yes, and 0 represents no. The analysis shows that there are significant differences in the expenditure of tourists in different consumption categories. The total consumption of tourist with ID 1002 is 1270 yuan, including 500 yuan for shopping and 300 yuan for dining. The transportation and tourism service expenses are relatively low. There is a high demand for shopping and dining, with a preference for historical districts with concentrated shopping areas and scenic spots. The paper recommend historical museums, shopping centers, etc. The tourist with ID 1007 spent a total of 140 yuan and 50 yuan on shopping, showing a preference for visiting a small number of attractions in a short period of time to reduce travel and transportation burden.

From the perspective of preference for products and types of dining, there is a clear preference trend in the data. The tourist with ID 1006 tends to purchase "souvenirs and clothing" and prefers "fast food and meals". This consumer places more emphasis on places that offer quick meals and souvenir sales when selecting tourist attractions. Tourist ID 1005 tends to focus on "snacks and coffee" in terms of dining, and has not shown interest in souvenir consumption. Their data on whether they have purchased souvenirs is 0, indicating that they have not purchased any souvenirs. This indicates that the tourist is more inclined to stay in leisure places for a longer period of time rather than a compact travel route. By utilizing data and combining NCF algorithm for personalized attraction recommendation and GA for path optimization, it can develop diverse travel routes tailored to the consumption patterns and preferences of different tourists. For high spending tourist 1002, the paper recommend high-end shopping centers, food courts, and other attractions. For low spending tourist 1007, the paper recommend small souvenir shops and other low-cost small-scale attractions.

3.2. Traffic Flow Data of Scenic Spots

The traffic flow data of scenic spots is a key basis for the planning of tourist routes in historical districts. This paper collects data on tourist flow to scenic spots, frequency of transportation usage, and conducts in-depth analysis of tourist travel patterns and preferences. This paper evaluates the popularity of tourist attractions based on data, providing accurate information for path planning. Table 3 shows the traffic flow data of different scenic spots, where R7-R12 respectively represent the daily tourist volume, bus passenger volume, subway passenger volume, private car passenger volume, peak traffic hours, and monthly tourist volume.

Table 3. Traffic Flow Data of Scenic Spots

Attraction ID	R7	R8	R9	R10	R11	R12
2001	1200	150	50	30	10:00-12:00	36,000
2002	800	70	20	10	14:00-16:00	24,000
2003	3500	300	200	100	16:00-18:00	105,000
2004	2000	100	50	200	08:00-10:00	60,000
2005	1000	120	30	20	11:00-13:00	30,000
2006	400	50	20	30	18:00-20:00	12,000
2007	1500	200	150	100	09:00-11:00	45,000
2008	2500	180	120	90	13:00-15:00	75,000
2009	3000	250	100	80	11:00-	90,000

					13:00	
2010	600	70	40	20	16:00-18:00	18,000

Table 3 shows that the daily average number of visitors to the scenic spot with ID 2003 is 3500, making it the most popular attraction and demonstrating high commercial appeal. The daily average number of tourists for attraction ID 2006 is 400, which is located at the end, indicating that there are fewer tourists interested in astronomy. The data reveals the trend of tourist visits, providing a basis for the development of refined transportation and path planning for high traffic attractions.

The usage of transportation shows the travel preferences of tourists from different scenic spots. Popular attractions such as shopping centers and water parks attract a large number of tourists by bus and subway, with the bus and subway traffic in the shopping center reaching 300 and 200 people per hour, respectively. The bus and subway traffic of the water park are 250 people/hour and 100 people/hour respectively, indicating that these scenic spots mainly rely on public and subway transportation systems to carry passenger flow. Park data shows that private car traffic is high, reaching 200 people per hour. This indicates that tourists of this type of attraction tend to choose self driving travel, depending on the geographical location of the attraction, which is generally located on the edge of the city and requires a longer stay.

The analysis of peak traffic hours provides key information for path optimization. The peak traffic period of the History Museum is from 10:00 to 12:00, during which there is a large flow of tourists, which can cause congestion on public transportation. The peak period for ancient architectural sites is from 14:00 to 16:00, and the tourist flow is relatively scattered, which allows for more flexible path planning and avoids excessive concentration during peak hours. Through the application of BD technology and intelligent algorithms, it is possible to accurately grasp the tourist flow and transportation usage of scenic spots, alleviate traffic pressure on scenic spots, improve overall tourism management efficiency, and provide decision support for digital tourism path planning in historical districts.

3.3. Personalized Attraction Recommendation Matrix

To achieve personalized attraction recommendation, a user attraction interaction matrix is constructed by analyzing the interaction data between tourists and attractions. The matrix displays the level of interest of each tourist in different attractions, with a rating range of 0 to 5, indicating the level of preference of tourists for the attractions. A zero value in the matrix indicates that the tourist has not visited or rated the attraction. Figure 5 shows the personalized attraction recommendation matrix.

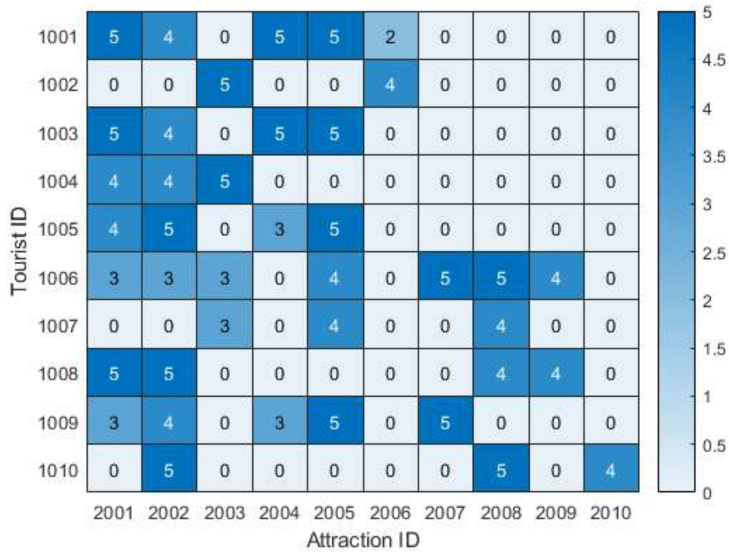

Fig. 5. Personalized Scenic Spot Recommendation Matrix

Figure 5 shows that tourist 1001 gave high ratings to attractions such as the 2001 History Museum, 2002 Ancient Architecture Ruins, 2004 Park, and 2005 Art Museum, with ratings of 5, 4, 5, and 5, respectively. This indicates that tourist 1001 has a greater interest in these attractions, while their interest in the Observatory scenic area is not high, with only a rating of 2. Tourist 1007 gave ratings of 3 points, 4 points, and 4 points respectively for the 2003 Shopping Center, 2005 Art Museum, and 2008 Cultural Square attractions, indicating that their interest in the scenic areas is relatively evenly distributed. Analyzing rating data can generate personalized recommendations for each tourist. Tourist 1006 gave high ratings to attractions such as the 2007 Observation Tower, 2008 Cultural Square, and 2009 Water Park, indicating that this tourist is more inclined to visit attractions that are entertaining and interactive. Based on interest preferences, this type of attraction should be recommended to tourists as a priority. When planning the route, the geographical distance and traffic conditions between attractions should be considered to optimize the tourist experience.

3.4. Comparison of Path Optimization Effects

In the research of digital tourism path planning in historical districts, path optimization is a key link to achieve efficient tourism experience. To evaluate the performance of different algorithms in path planning, several commonly used path optimization algorithms were compared, and their respective advantages and characteristics were analyzed. Figure 6 shows the path optimization effect.

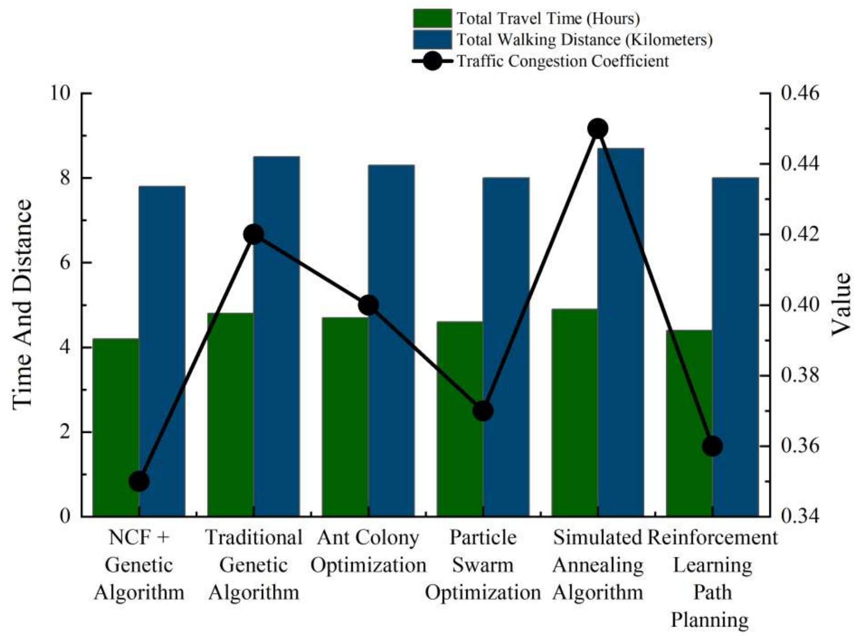

Fig. 6. Comparison of Path Optimization Effects

From the data in Figure 6, regarding total travel time, NCF+GA leads other algorithms with the shortest time of 4.2 hours, while traditional GA takes 4.8 hours and simulated annealing algorithm takes 4.9 hours. This indicates that the NCF algorithm can effectively reduce travel time and avoid unnecessary delays by combining historical data with personalized recommendations. In terms of total walking distance, NCF+GA also performed well, with a total walking distance of 7.8 kilometers, lower than the traditional GA's 8.5 kilometers and the ant colony algorithm's 8.3 kilometers. NCF+GA can provide tourists with more efficient routes and reduce unnecessary walking. In terms of traffic congestion coefficient, NCF+GA dominates with a minimum value of 0.35, which is much lower than the traditional GA's 0.42 and the simulated annealing algorithm's 0.45, demonstrating the advantage of this algorithm in avoiding traffic congestion areas.

4. Experimental Analysis

This study introduces a digital tourism path planning method for historical districts based on personalized recommendation and path optimization by combining BD technology and intelligent algorithms. By constructing a tourist attraction interaction matrix and utilizing the NCF algorithm, personalized attraction recommendations are made based on tourists' interests and consumption preferences. GA can be used to optimize the path, considering factors such as distance between scenic spots and traffic flow, to help tourists efficiently and comfortably explore the scenic area. It can be combined with a real-time data streaming platform to dynamically adjust routes to cope with sudden traffic or scenic congestion, enhancing the overall experience of tourists. This method not only enhances the personalized experience of tourists, but also provides effective decision support for traffic management and tourist flow scheduling in scenic spots.

There are shortcomings in the research, such as a large space for algorithm optimization. Insufficient representativeness of tourist behavior data can affect the accuracy of the model. The existing algorithms have limitations in handling complex tourist demands and attraction relationships. Research can improve the accuracy of recommendation and path planning by enriching data sources and introducing advanced algorithms. By combining VR (Virtual Reality), AR (Augmented Reality) and other technologies with intelligent hardware, the paper aimed to provide a more intelligent and personalized travel experience, and promote the development of path planning towards refinement and real-time.

5. Conclusions

This paper introduces a tourism path planning method suitable for historical districts, combining BD and intelligent algorithms. Tourist profiles can be constructed based on tourist behavior data, attraction information, and real-time traffic data, and personalized attraction recommendations can be achieved using the NCF algorithm. Path optimization adopts GA, which integrates tourist interests, scenic spot distance, and traffic conditions. This paper introduces the Apache Kafka platform to process real-time data, dynamically adjust paths, and solve traffic and attraction congestion problems. The experimental results show that this method accurately generates customized routes and effectively improves the tourist experience. In the future, the paper can expand data sources and optimize algorithms to improve accuracy and provide new directions for intelligent path planning in historic districts.

Funding

This work was supported by 2023 Shaanxi Province Art and Science Planning Project (2023HZ1754).

References

[1] Ecola, L., Zmud, J., Gu, K., Phleps, P., & Feige, I. (2016). Future travel demand in China: Scenarios for Year 2030. Transportation Research Record, 2581(1), 57-65.

[2] Xiao, L., Liao, J., Wu, S., Tian, Y., & Sun, J. (2024). Understanding the attitudes of travelers towards incentive-based travel demand management strategies in Suzhou, China. Travel Behaviour and Society, 35, 100752.

[3] Xu, Z., Lv, Z., Li, J., Sun, H., & Sheng, Z. (2022). A novel perspective on travel demand prediction considering natural environmental and socioeconomic factors. IEEE Intelligent Transportation Systems Magazine, 15(1), 136-159.

[4] Zheng, W., Li, M., Lin, Z., & Zhang, Y. (2022). Leveraging tourist trajectory data for effective destination planning and management: A new heuristic approach. Tourism Management, 89, 104437.

[5] Li, S., Luo, T., Wang, L., Xing, L., & Ren, T. (2022). Tourism route optimization based on improved knowledge ant colony algorithm. Complex & Intelligent Systems, 8(5), 3973-3988.

[6] Haris, E., & Gan, K. H. (2017). Mining graphs from travel blogs: a review in the context of tour planning. Information Technology & Tourism, 17, 429-453.

[7] Aoike, T., Ho, B., Hara, T., Ota, J., & Kurata, Y. (2019). Utilising crowd information of tourist spots in an interactive tour recommender system. In Information and Communication Technologies in Tourism 2019: Proceedings of the International Conference in Nicosia, Cyprus, January 30–February 1, 2019 (pp. 27-39). Springer International Publishing.

[8] Kotiloglu, S., Lappas, T., Pelechrinis, K., & Repoussis, P. P. (2017). Personalized multi-period tour recommendations. Tourism Management, 62, 76-88.

[9] Li, S., Komatsu, I., Yamada, K., Takagi, M., & Sasaki, J. (2018). Personal tour planning system (PTPS) for use in urban and rural areas. In New Trends in Intelligent Software Methodologies, Tools and Techniques (pp. 259-270). IOS Press.

[10] Hidaka, M., Matsuda, Y., Kawanaka, S., Nakamura, Y., Fujimoto, M., Arakawa, Y., & Yasumoto, K. (2017,

August). A system for collecting and curating sightseeing information toward satisfactory tour plan creation. In The Second International Workshop on Smart Sensing Systems (IWSSS'17) (Vol. 2018).

[11] Deng, Z., Weng, D., & Wu, Y. (2023). You are experienced: Interactive tour planning with crowdsourcing tour data from web. Journal of Visualization, 26(2), 385-401.

[12] García-Hernández, M., De la Calle-Vaquero, M., & Yubero, C. (2017). Cultural heritage and urban tourism: Historic city centres under pressure. Sustainability, 9(8), 1346.

[13] MIAO, C. H., XU, J. W., YU, Z. L., LI, L. J., & ZHANG, Y. (2025). Evaluation and enhancement of historic district regeneration effectiveness in the context of cultural-tourism integration: A case study of three historic districts in Kaifeng. JOURNAL OF NATURAL RESOURCES, 40(1), 164-180.

[14] Liang, F., Pan, Y., Gu, M., Liu, Y., & Lei, L. (2022). Research on the paths and strategies of the integrated development of culture and tourism industry in urban historical blocks. Frontiers in Public Health, 10, 1016801.

[15] Hsu, W. L., Chang, Y. J., Mou, L., Huang, J. W., & Liu, H. L. (2024). The Distribution and Accessibility of Elements of Tourism in Historic and Cultural Cities. Big Data and Cognitive Computing, 8(3), 29.

[16] Zhang, K., Deng, H., Wang, F., & Yuan, Y. (2023). Reproduction of consumer spaces and historic district touristification in Old Beijing City. Tourism Geographies, 25(2-3), 508-531.

[17] Chen, B. H., Han, J., Chen, S., Yin, J. L., & Chen, Z. (2022). Automatic itinerary planning using triple-agent deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, 23(10), 18864-18875.

[18] Zhou, X., Su, M., Liu, Z., Hu, Y., Sun, B., & Feng, G. (2020). Smart tour route planning algorithm based on naïve Bayes interest data mining machine learning. ISPRS International Journal of Geo-Information, 9(2), 112.

[19] Dezfouli, M. B., Shahraki, M. H. N., & Zamani, H. (2018, September). A novel tour planning model using big data. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) (pp. 1-6). IEEE.

[20] Zhou, X., Zhan, Y., Feng, G., Zhang, D., & Li, S. (2019). Individualized tour route plan algorithm based on tourist sight spatial interest field. ISPRS International Journal of Geo-Information, 8(4), 192.

[21] Lim, K. H., Chan, J., Leckie, C., & Karunasekera, S. (2018). Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. Knowledge and Information Systems, 54, 375-406.

[22] Zhou, X., Sun, B., Li, S., & Liu, S. (2020). Tour route planning algorithm based on precise interested tourist sight data mining. IEEE Access, 8, 153134-153168.