Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

Research on Domestic Tourism Route Planning Based on Dynamic Programming and Heuristic Search

Qing Liu, Xudong Zhan

School of Electronic and Information Engineering, Liaoning Technical University, Huludao, 125100, China

Abstract: *This study focuses on the planning of travel routes for foreign tourists in China, aiming to maximize the weighted sum of the number of tourist attractions and their ratings while minimizing the total cost. For this purpose, this paper selects "the 50 most desirable cities for foreign tourists" from 352 cities in China as the research object. The dynamic programming algorithm is adopted, and various constraint conditions such as starting the tour at a designated location and ending the journey within a specified time are considered to formulate travel routes. In addition, for tourists who prefer visiting attractions of specific types, this study also collects the latitude and longitude information of 352 cities in China, constructs a graph model, and designs a heuristic function and uses the A* algorithm for heuristic search. Eventually, an effectively implementable travel route is planned.*

Keywords: *Tourism Route Planning, Dynamic Programming Algorithm, Heuristic Search, Specific Types of Tourist Attraction*

1. Introduction

In recent years, domestic research on tourism route planning has made remarkable progress in different travel modes and optimization goals. In terms of high-speed rail travel optimization, Shan Xinghua et al.[1] conducted passenger flow distribution research targeting the travel preferences of different passenger groups. They constructed a three-dimensional travel network of "time-space-group" and established a network flow model based on the constraint of a certain number of train passengers to solve the passenger flow distribution state of each group in terms of travel routes. In terms of algorithm application, for the tourism route design of Jingdezhen ceramic culture scenic area, Tang Kezong et al.[2] proposed a Weight Disturbance Mechanism-based Simulated Annealing Genetic Algorithm (WDMAGA). This algorithm introduces the weight disturbance mechanism of the path length change function into the path evaluation criterion and incorporates the temperature annealing factor in the simulated annealing algorithm, so as to make the evaluation value of each path closer to the real environment. Regarding the NP characteristic of tourism route planning problems, Zhang Ruijiao et al.[3] proposed a genetic algorithm that can effectively maintain population diversity. This algorithm generates an initial population by means of the Jaccard coefficient, uses multiple mutation operators to generate multiple offspring, and selects the better individuals from the offspring and the parent generation to form a new population. In practical applications, tourism route planning problems often have the characteristics of large scale and high complexity, and efficient algorithms are needed to solve them. At the same time, accurately predicting tourists' behaviors and preferences is also the key to improving the quality of tourism route planning. At present, there are still some deficiencies in this aspect, and further in-depth exploration and improvement are needed. For this reason, this study focuses on the planning problem of foreign tourists' tourism routes in China, aiming to maximize the weighted sum of the number of tourist attractions visited and their ratings, while minimizing the total cost. By screening the "50 cities most yearned for by foreign tourists", the dynamic programming algorithm is adopted and multiple constraint conditions are considered to formulate tourism routes. At the same time, for tourists who prefer specific types of attractions, city latitude and longitude information are collected, a graph model is constructed, and effective tourism routes are planned through heuristic search, hoping to provide scientific and reasonable route planning schemes for foreign tourists traveling in China.

(Data Source: https://www.saikr.com/c/nd/22100,https://www.macrodatas.cn/article/1147471811)

ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

2. Based on the dynamic planning of tourism route planning research

2.1 Data pre-processing

(1) Handling Missing Values

Read the data from all CSV files and merge them into one file. Add a new column "City" to indicate the city where the tourist attraction is located. Through the inspection of the merged data, it is found that there are many missing data. For example, the city address, introduction, opening hours, etc. of some tourist attractions are missing. Even the rating is empty. To provide foreign tourists with a better travel experience, for the data with ratings of "--", 0, or empty, the corresponding tourist attraction information is directly deleted.

(2) Outlier Detection

By using box plots to detect the merged data, it is found that the distribution range of ratings is all from 0 to 5 (including 5), and there are no outliers.

2.2 Data analysis

To select the 50 cities that are most desirable for foreign tourists, this paper collects relevant data sets on city scale (household registered population, built-up area, gross regional product, per capita gross regional product, growth rate of gross regional product), environmental protection (park green area, green coverage rate of built-up areas, centralized treatment rate of sewage treatment plants, harmless treatment rate of domestic waste), cultural heritage (number of schools), transportation convenience (passenger volume of highway transportation), and medical care (number of hospitals). For missing values, rule filling is adopted, and the missing values are replaced vertically with the values above the missing values. Min-Max Scaling is used to standardize the processed data to eliminate the influence of different dimensions and magnitudes. And through the TOPSIS method, systematically evaluate and determine which cities are most desirable for foreign tourists. Combining multiple evaluation indicators and weights (as shown in Table 1), by calculating the distance between positive and negative ideal solutions and relative closeness, 50 cities that are most desirable for foreign tourists are obtained as shown in Table 1.

Table 1: Weight values of each index

Entropy weight method			
Item	Information entropy value e	Information Utility Value d	Weight (%)
Household registration (10,000 prople)	0.912	0.088	7.283
Number of schools (unit)	0.924	0.076	6.323
Number of hospitals (unit)	0.915	0.085	7.003
Passenger volume of highway transportation (10,000 people)	0.865	0.135	11.165
Growth rate of regional GDP (%)	0.992	0.008	0.671
Park green area	0.876	0.124	10.267
Centralized treatment rate of sewage treatment plants (%)	0.631	0.369	30.529
Harmless treatment rate of domestic waste (%)	0.998	0.002	0.187
Green coverage rate of built-up areas (%)	0.999	0.001	0.09
Per capita regional GDP (yuan)	0.955	0.045	3.703
Regional GDP(100 miooion yuan)	0.826	0.174	14.382
Built-up area (square kilometers)	0.898	0.102	8.396

The finally selected top 50 cities perform outstandingly in multiple aspects (such as city scale, environmental quality, economy, transportation, etc.) and have high attractiveness. As shown in Figure 1, which is part of the ranking data of these 50 cities. Among them, Beijing has the highest comprehensive score and ranks first. Daqing City has the lowest comprehensive score and ranks fiftieth.

ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

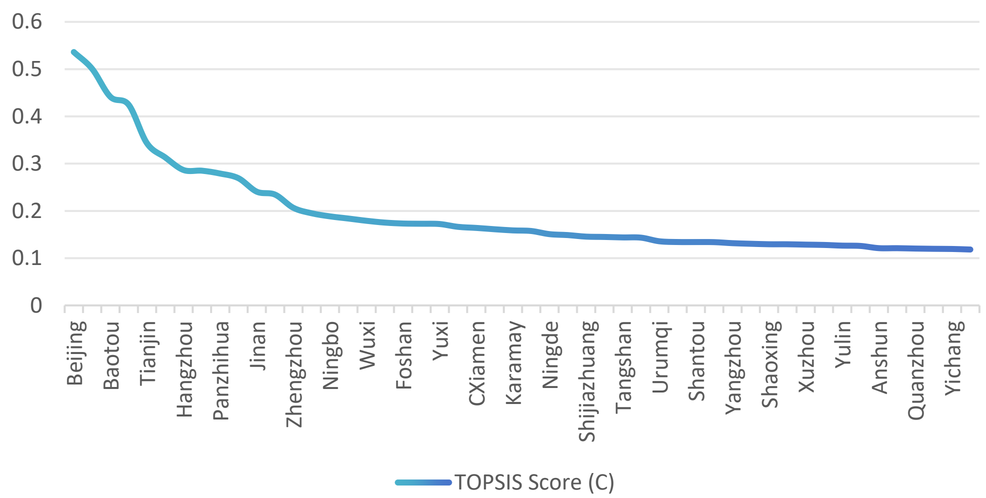

Figure 1: Top 50 cities section

2.3 Model construction

This paper conducts travel route planning for foreign tourists' six-day trip to China[4]. To provide better experience for tourists, this study considers tourists' travel modes and travel expenses as well as play expenses, aiming to minimize the total play cost and maximize the play experience.

(1) Objective Function

The average attraction score Score is a floating-point number, which represents the sum of the ratings of all visited attractions divided by the number of visited attractions. β is a positive floating-point number. As a weighting coefficient, it is used to balance the importance of the number of attractions and the attraction score in the objective function[5]. By adjusting the weighting coefficient β, a suitable balance point between the number of visited attractions and the attraction score can be found to maximize the overall objective function value. The smaller the total cost is, the better. However, here is to maximize the objective function. Therefore, a negative sign is added before the total cost to maximize the overall objective function value. The objective function is:

$$Maximize\begin{cases} Z_1 = \sum_{i=1}^{n} y_i + \beta \times score \\ Z_2 = -\sum_{i=1}^{n}\sum_{j=1}^{n} F_{ij} \times x_{ij} + \sum_{i=1}^{n} T_i \times x_i \end{cases} \tag{1}$$

$$T_{total1} \le t_{\max} \tag{2}$$

(2) Constraint condition

Each city can be visited at most once:

$$\sum_{j=1}^{n} X_{ij} \le 1, \forall i \tag{3}$$

The playing time cannot exceed 144 hours:

$$T_{total} = \sum_{i=1}^{n} A_i \times y_i + \sum_{i=1}^{n}\sum_{j=1}^{n} T_{ij} \times x_{ij} + s \times \left\lfloor \frac{T_{total}}{24} \right\rfloor \le t_{\max} \tag{4}$$

The departure city is f:

$$y_f = 1 \tag{5}$$

ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

If city j vsited after city i, $y_i = 1$ and $x_{ij} \le y_j$ hold for all i and j.

2.4 Based on dynamic programming solution model

(1) State definition

The dp$[i][t][k]$ is defined to represent the minimum cost of playing k cities in the first i cities in t time. Here i ranges from 0 to n (total number of cities), t from 0 to t_{max}, and k from 0 to n.

(2) State transition equation

For each city i consider whether to add it to the play path, if not:

$$dp[i][t][k] = dp[i-1][t][k] \tag{6}$$

If it is included, then we need to find a city j (j < i) such that the high-speed rail cost and travel cost from j to i are the minimum, and the time is allowed. That is:

$$dp[i][t][k] = \min_{j<i}\left(dp[j]\times\left[t - A_i - \frac{d_{ij}}{v} - 8\right][k-1] + T_i + c\left(d_{ij}\right)\right) \tag{7}$$

dp$[i][t][k]$ is the minimum cost for the first i cities, t time spent, and k cities visited. $dp[j]\left[t - A_i - \frac{d_{ij}}{v} - 8\right][k-1]$represents the minimum cost in the case of considering the first j cities, spending $t - A_i - \frac{d_{ij}}{v} - 8$ time, and visiting k-1 cities. Among them, A_i represents the travel time in city i (which may include the time for visiting scenic spots and other activities), $\frac{d_{ij}}{v}$ represents the high-speed rail travel time from city j to city i (distance divided by the average speed), 8 represents the time for sleeping every day, and T_i represents the entrance ticket of city i. $c(d_{ij})$ represents the high-speed rail cost from city j to city i, which is a function calculated based on the distance d_{ij}.

(3) Initialization

dp$[0][0][0]$= 0, which means you haven't started playing yet, time is 0, number of cities to play is 0, and cost is 0.

All other dp$[i][t][k]$ initializations are set to infinity (or a sufficiently large number) to indicate that a valid play path is not currently found.

(4) Results

The dp array is calculated in city order i, Time Order t, and Play City Order k. For each combination of i, t, and k, all possible j's need to be traversed to find the minimum cost.The final answer is min(dp$[i][t][k]$) , where $t \le t_{max}$, k values range from 0 to n. By traversing all possible t and k to find the minimum cost, and record the corresponding play path.

(5) Optimization

Because of the large state space, mnemonics search[6]was used in this article to avoid repeated computations. That is, before calculating the dp$[i][t][k]$check that the state has been calculated, and if so, return the results of the previous calculation directly. The cities are sorted by how close they are to city f (the city of departure). Thus, only j<i cities need to be considered when calculating dp$[i][t][k]$, thus reducing unnecessary calculations.

(6) Output

Through the dynamic planning table can find the optimal play path, as well as the corresponding total cost, total time and the number of cities to play. This paper starts with the state of the final answer and traces the play path backwards. That is, starting from dp$[i][t][k]$, find the state to get the minimum value of j, and then the city j into the play path, and continue to track the dp$[i][t{-}\ldots][k-1]$, until the dp$[0][0][0]$.

ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

2.5 Results Analysis

The Dynamic Programming Algorithm exhibits global optimality by traversing all possible state combinations and comparing costs under different choices to ensure finding the global optimal solution[7]. At each state transition, the algorithm considers all possible previous states and selects the one with the minimum cost. The time complexity of the algorithm mainly depends on the size of the state space and the computational complexity of state transitions. In this study, the state space is composed of three dimensions: the number of cities, time, and the number of visited cities. The space is extremely large, so the computational complexity of the algorithm is very high. The space complexity of the algorithm is mainly determined by the size of the dp array. The dp array is a three-dimensional array, and its size is proportional to the number of cities, the maximum playing time, and the number of visited cities. Since these parameters are very large, the algorithm also has a significant space requirement. Through memoized search, the algorithm can avoid repeated calculations of the same state, thereby improving efficiency. Taking Guangzhou as the departure city as an example, the following results are obtained through result solving:

Figure 2: Path planning

Starting from Guangzhou clockwise as shown in Figure 2, travel time: 9.59×10^1 hours, total consumption time: 1.44×10^2 hours, total consumption: 1.64×10^3 yuan, number of cities visited: 1.7×10^1 cities.

3. Heuristic search and play route planning for specific scenic spots based on graph model

3.1 Data processing

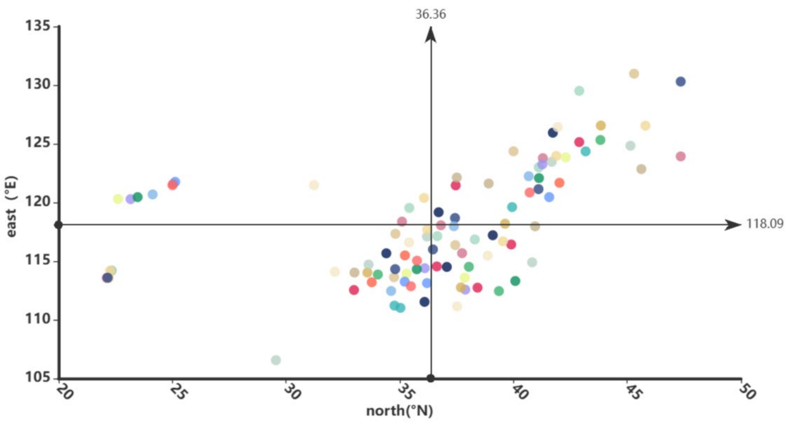

Figure 3: Latitude and longitude distribution of all cities

From a total of 35,200 tourist attractions in 352 cities, mountain scenery is screened. Considering that many city names and the districts where they are located contain the character "mountain", only tourist attractions with the character "mountain" as the last character are considered during screening.

ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

After screening, they are further screened according to the rating. Only one best mountain scenery is left for each city. This paper collects the latitudes and longitudes of these cities and uses the Haversine formula[8] to calculate the actual distance between two cities.

(1) Calculating the distance between cities the distribution of latitudes and longitudes of all cities in the collected data set is as shown in Figure 3:

The actual distance between two cities was calculated using the Haversine formula. The formula is:

$$d = 2r\sin^{-1}\sqrt{\sin^2\left(\frac{\triangle latitude}{2}\right) + \cos\left(latitude_1\right)\cdot\cos latitude_2\cdot\sin^2\left(\frac{\triangle longitude}{2}\right)} \tag{8}$$

Where R is the radius of the earth (about 6,371 km) and δlatitude and δlongitude are the difference between the latitude and longitude of the two cities, respectively.

(2) Calculate the cost and time of HSR

Train costs are calculated based on distance (d) , using the given discount rule. The formula is:

$$train\_\cos t = \begin{cases} d\times 0.46\left(d\le 500\right) \\ 500\times 0.46+\left(d-500\right)\times 0.46\times 0.9\left(500<d\le 1000\right) \\ 500\times 0.46+\left(1000-500\right)\times 0.46\times 0.9+\left(d-1000\right)\times 0.46\times 0.8\left(1000<d\le 1500\right) \end{cases} \tag{9}$$

By analogy, for every 500 km increase, the discount rate is reduced by 0.1.

Train time is calculated based on distance (d) and the average speed of the high-speed train (V, about 300 km/h) . The formula is:

$$\text{train}\_time = \frac{d}{v} \tag{10}$$

3.2 Build diagram model

(1) Nodes:

Each city is regarded as a node. The node contains information such as the city name and the ticket cost of the mountain with the highest rating.

(2) Edges:

There is an edge between each pair of cities. The weight of the edge is a combination of high-speed rail cost and time. In this paper, a weighted function is used to convert cost and time into a single weight as follows:

$$\text{weigh}t = \alpha\times\text{train}\_\cos t + \left(1-\alpha\right)\times train\_time \tag{11}$$

Where α is a weighting coefficient used to balance the importance of cost and time in path selection.

3.3 Heuristic search for solutions

(1) Design of Heuristic Function

Design a heuristic function to estimate the estimated cost of the lowest-cost path from the current node to the target node (or end point). Considering factors such as remaining time, the number of remaining mountains that can be visited, and expected transportation costs, the function is as follows:

$$\begin{aligned} \text{h}\left(\text{n}\right) &= \text{estimated\_remaining\_cost} = \beta\times\text{estimated\_remaining\_ticket\_cost} + \\ &\gamma\times\text{estimated\_remaining\_train\_cost} + \theta\times\text{estimated\_remaining\_time\_cost} \end{aligned} \tag{12}$$

Among them, β, γ, and θ are the weighting factors used to balance the importance of different cost factors.

(2) Search Algorithm

Use the A* algorithm[9] to find the optimal path. During the execution of the algorithm, maintain an open set (nodes to be visited) and a closed set (nodes that have been visited). In each iteration, the algorithm selects the node with the lowest f(n) value from the open set, where f(n) = g(n) + h(n). g(n) is the actual cost from the starting point to node n, and h(n) is the heuristic estimated cost from node n to the end point.

(3) Setting of Termination Conditions

The search process terminates when one of the following conditions is met: First, when the time limit of 144 hours is reached; second, at a certain node, a feasible path can no longer be found (that is, more mountains cannot be visited within the remaining time).

3.4 Analysis of results

In this paper, the Haversine formula is used to calculate the distance between cities because it takes into account the curvature of the earth and can provide more accurate geographical distances. A graph model is constructed with cities as nodes and transportation connections as edges. The weight of the edges considers travel costs and time, which is beneficial for comprehensively considering multiple factors when searching for the optimal path. The heuristic search algorithm can find the optimal or near-optimal path in a large graph[10], while considering travel costs and time constraints. By adjusting the weighting coefficient and heuristic function, the model can flexibly adapt to different tourism preferences and constraint conditions, improving practicability. The final result is: total time 143.13 hours, total cost 1536.40 yuan, 22 visited cities.

4. Conclusions

This research is dedicated to providing effective solutions for domestic travel route planning, especially for foreign tourists' travel in China. When screening the top 50 cities most desired by foreign tourists, multiple evaluation indicators and advanced techniques such as min-max normalization and the TOPSIS method with entropy weight determination are comprehensively utilized to provide a scientific and objective approach for determining cities with rich attractiveness and high quality. For travel route planning, this paper employs two unique methods. On the one hand, the dynamic programming algorithm is used to plan general travel routes, fully considering various constraint conditions such as travel expenses, time limits, and departure places. Through detailed formulas and optimization steps such as memory-based search and city sorting, efficient calculation is achieved and the optimal route is generated. The case study starting from Guangzhou successfully demonstrates the practical application effect and effectiveness of this method. On the other hand, for tourists who prefer specific types of attractions (mountain scenery), a heuristic search based on a graph model is implemented. This process involves data processing to select the best mountain scenery attractions in each city, using the Haversine formula to calculate the distance between cities, and constructing a graph model with cities as nodes and weighted edges representing travel costs and time. By designing a heuristic function and using the A* algorithm, considering various factors such as remaining time, the number of attractions, and transportation expenses, the optimal route is successfully searched. The results also fully prove the feasibility and practicability of this method.

In general, this research provides highly valuable insights and practical solutions for tourism planners, travel agencies, and foreign tourists. It not only greatly enhances our understanding of travel route planning but also provides tools and methods that can be applied in practical scenarios, contributing to improving the tourism experience and promoting sustainable tourism development.

References

[1] Shan Xinghua, Zhao Shuo. High-speed rail passenger flow distribution method considering travel preferences of passenger groups [J]. Railway Computer Application, 2024, 33(08): 12-18.

[2] Tang Kezong, Liu Kang, Liu Yujiao. Tourism route planning of Jingdezhen ceramic culture scenic area based on weight disturbance mechanism simulated annealing genetic algorithm[J]. Software Guide, 2023, 22(04): 96-102.

[3] Zhang Ruijiao, Chen Chongcheng, Huang Zhengrui, et al. Solving cultural tourism route planning problems with improved genetic algorithm[J]. Journal of Guizhou University (Natural Sciences Edition), 2022, 39(01): 57-64.

ISSN 2616-5775 Vol. 7, Issue 12: 16-23, DOI: 10.25236/AJCIS.2024.071203

[4] Gong Yuhuang. A study on artificial intelligence-based travel route recommendation system trust and willingness to accept [D]. Southwestern University of Finance and Economics, 2021.
[5] Song Y , Ding D , Wang R ,et al.Analysis and Design of the Shopping Trolley for Elderly People Based on Ergonomics[J].Springer, Singapore, 2023.DOI:10.1007/978-981-19-4786-5_31.
[6] Li Min, Zhu Nan, Li Minghui. A study on Chengdu inner-city travel route planning based on SA and GA algorithm [J]. Value engineering, 2022, 41(28): 38-40.
[7] Wong Chi Ho. Path planning and trajectory optimization based on improved A-star algorithm [D]. Jilin University, 2024.
[8] Wei Xin. Research on personalized travel route generation modeling based on user interest model [J]. Automation technology and applications, 2024, 43(03): 177-180.
[9] Liu Na. Method of route recommendation based on preference data [J]. Information Technology, 2023, (11): 148-152 + 157.
[10] Li Xia, Yin Chuandong, Yuan Yun. Comparative Analysis of travel route personalized recommendation algorithms [J]. Computer Technology and development, 26(09)2016:73-77.