Research on Optimization of Travel Plans for Foreign Tourists in China Based on Greedy Algorithm and Simulated Annealing Algorithm

Lingyu Ye *

College of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan, China, 030024

* Corresponding Author Email: yelingyu8658@foxmail.com

Abstract. This article leverages the Pandas library in Python to process and analyze extensive datasets, focusing on national tourist attractions. It reveals that 2563 attractions have achieved the highest rating of 5.0. Cities such as Sansha, Wujiaqu, Yuxi, Yiyang, Tianmen, Alar, Weifang, Yantai, Daxing'anling, and Xingtai top the list for the most fully rated scenic spots. A comprehensive evaluation system using the Analytic Hierarchy Process (AHP) is constructed. NLP technology builds a keyword library to match scenic area descriptions, evaluating environment, culture, transportation, climate, and cuisine. City line levels and intercity distances are integrated into an urban evaluation model. The study selects the top 50 cities attracting foreign tourists. A multi-objective 0-1 planning model for tourism planning is formulated to maximize tourist cities and scenic spot ratings. The greedy algorithm generates an initial solution, refined by the simulated annealing algorithm to seek the global optimum.

Keywords: Multi Objective Optimization, Greedy Algorithm, Simulated Annealing Algorithm, Tourism Planning Strategy.

1. Introduction

In recent years, the internet slang "city not city" has attracted widespread attention under the promotion of foreign internet celebrities. The implementation of China's visa free transit policy has attracted more and more foreign tourists to come to China. They share their travel experiences in China through online platforms, which not only promotes the growth of China's tourism industry but also presents a vivid and authentic image of China to the international community, with fruitful results [1].

Assuming that foreign tourists can stay in China for 144 hours after entering and can depart from any airport near a city. Considering that there are many tourist attractions in each city, the "Best Tourist Attraction in the City" principle has been established, which means that each tourist only selects the highest rated attraction in each city [2].

2. Research on Preprocessing Methods for Tourism Data

2.1. Structural Analysis of Tourism Data Deduplication and Outlier Handling Methods

This article first collects rating data on tourist attractions in various cities across the country from various data sources, such as tourism websites, government databases, third-party data providers, etc [3].Ensure the integrity and quality of data, and comply with relevant data protection regulations. Remove duplicate data: Use deduplication algorithms or database query techniques to remove duplicate records from the dataset, ensuring that each attraction is only counted once. Handling Outliers: Identify and handle outliers in rating data, such as extremely high or low scores, which may be caused by data entry errors or special circumstances. You can choose to remove outliers or replace them with reasonable values (such as mean, median, etc.). For missing rating data, appropriate interpolation methods should be used, such as using the historical average rating of the attraction, the average rating of similar attractions, or predicting and filling based on other relevant features. Merge and integrate data from different sources to form a complete dataset for subsequent analysis. According to the analysis requirements, necessary transformations are made to the data, such as converting categorical data (such as city names) into a format suitable for analysis (such as encoding), or standardizing rating data to eliminate the influence of different dimensions [4].

2.2. Validity

The calculation formula for KMO value is:

$$\mathrm{KMO} = \frac{\sum\sum r_{ij}^2}{\sum\sum r_{ij}^2 + \sum\sum u_{ij}^2} \tag{1}$$

Among them, r_{ij} is the simple correlation coefficient between variables? u_{ij} Is the partial correlation coefficient between variables?

Table 1. KMO Bartlett test.

KMO and Bartlett inspection		
KMO sampling suitability quantity		0.793
Bartlett sphericity test	Approximate chi square	3942.865
	freedom	153
	significance	0.000

According to Table 1, the KMO result mentioned above is 0.793. This value is greater than 0.7, according to Kaiser's criteria, indicating that the data is suitable for factor analysis. The closer the KMO value is to 1, the higher the commonality between variables, and the data is more suitable for factor analysis.

Composition load diagram, as shown in Figure 1:

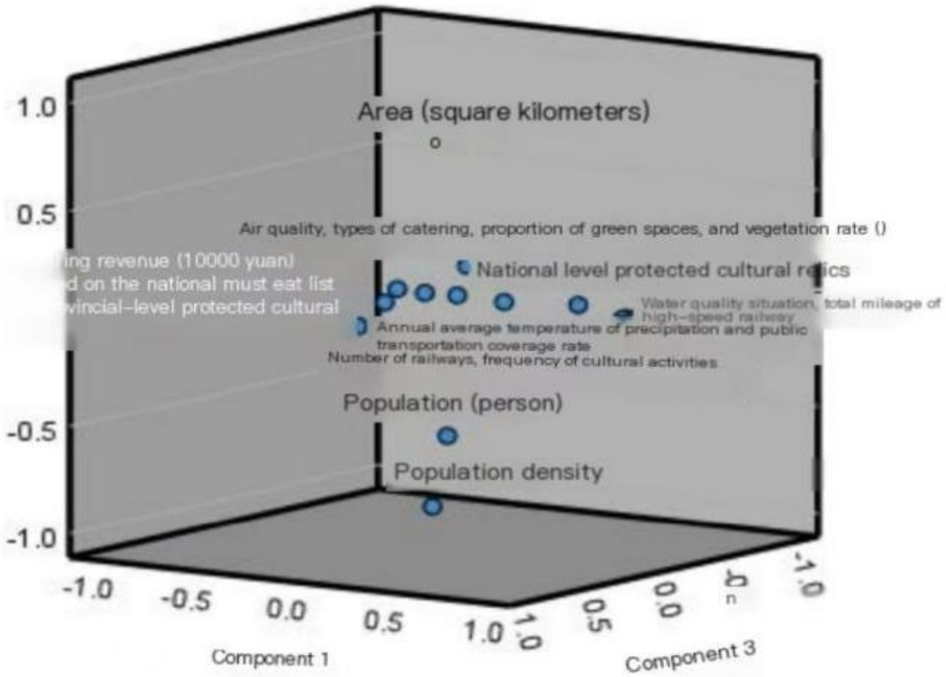

Figure 1. Component Load Diagram.

The component loading diagram shows the projected positions of each variable on three principal components (component 1, component 2, and component 3). Each point represents an original variable, and the position in the graph represents the magnitude and direction of the load on that variable in principal component space. Through this information, it can be seen which variables contribute the most to each principal component, thus understanding the actual meaning of the principal components.

Principal Component 1 represents the characteristics related to tourism consumption and passenger flow, reflecting the level of tourism economy in a city.

Principal Component 2 represents features related to urban area and cultural preservation, reflecting the cultural heritage and infrastructure coverage of a city.

Principal Component 3 represents the characteristics related to cultural activities and public services, reflecting the frequency of cultural activities and the quality of public services in a city.

Principal Component Analysis (PCA):

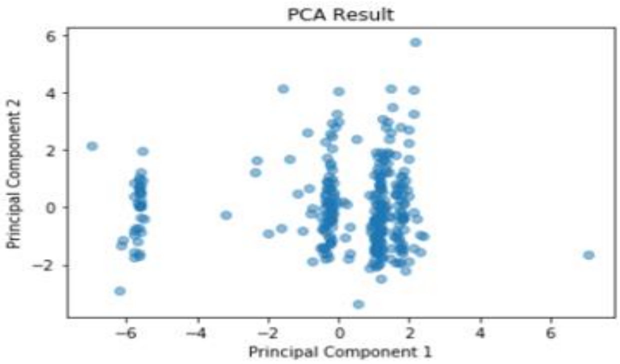

Figure 2. PCA test chart.

According to Figure 2, the distribution on the two main components (Principal Component 1 and Principal Component 2) can be seen. Each point represents the position of the original data in the reduced two-dimensional space. The purpose of PCA is to project the original data onto a new coordinate system through linear transformation, so that the variance of the data on these new axes is maximized. The distribution in the figure indicates a clear clustering or distribution pattern of the data on these two main components, which can help identify the structure or features of the data.

3. Establishment of Time Factor Model

3.1. objective function

Minimize total time:

$$T_{\text{total}} = \sum_{(i,j)\in P} \left(\frac{D_{ij}}{S} + T_{\text{station}_\text{h}\text{otel},i} + T_{\text{queue},i} + T_{\text{consult},i} + T_{\text{play},i}\right) \tag{2}$$

Constraint condition

(1) Total time not exceeding 88 hours:

$$\mathrm{T}_{\text{total}} \leq 88 \tag{3}$$

(2) Time constraints for each path segment:

$$\begin{gathered} T_{\text{travel},ij} = \frac{D_{ij}}{S} \\ T_{\text{total},ij} = T_{\text{travel},ij} + T_{\text{station}_\text{h}\text{otel},i} + T_{\text{qucuc},i} + T_{\text{consult},i} + T_{\text{play},i} \end{gathered} \tag{4}$$

(3) Path selection constraint:

$$\sum_{(i,j)\in P} D_{ij} \tag{5}$$

Firstly, variables related to travel time, path distance, etc. were defined. The definitions of these variables are the foundation of the model, ensuring that each time period and path distance are clearly represented. Established an objective function to minimize the total time. By adding all time components together, including high-speed rail travel time, high-speed rail station to hotel time, queue time, consultation time, and travel time, and the expression for the total time can be determined. By introducing constraints, ensure that travel time is within a reasonable range and meets actual travel needs. For example, the constraint of a total time not exceeding 88 hours ensures that the travel plan is feasible and time effective. Ensure the rationality and effectiveness of path selection. This part is very important to ensure that the chosen path is feasible in reality and meets travel needs.

3.2. Calculate the highest rating for each city

Sort the ratings of all attractions in each city, identify the highest rated attractions in each city, record and save these highest rating values for subsequent comparison and statistics.

3.3. The highest national rating has been determined

Compare the highest ratings of all cities to determine the highest rating nationwide, and count the number of highest rated attractions in each city: count the attractions in each city that have reached or exceeded a certain benchmark value (such as a certain percentage of the highest rating nationwide). This benchmark value can be set according to the actual situation to reflect the city's performance in high-quality scenic spots.

3.4. Sort and list the top 10 cities

Ranking the top ten cities based on the number of highest rated attractions in each city, these cities demonstrate the most outstanding performance in terms of high-quality attractions. The highest rating (BS) is 5.0. There are a total of 2563 scenic spots in China that have received the highest rating (BS). The top 10 cities with the most BS rated scenic spots are shown in Figure 3.

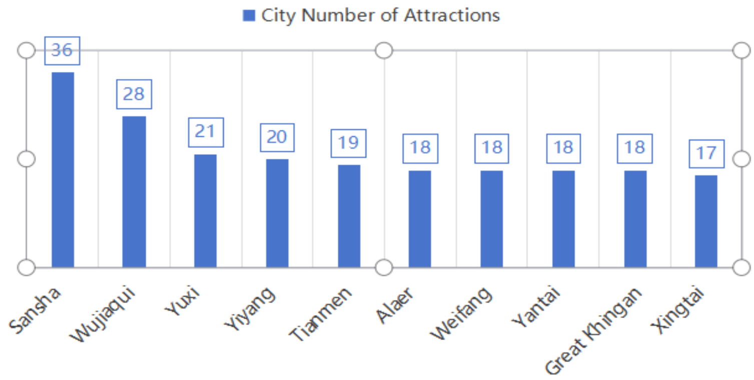

Figure 3. Distribution of the number of scenic spots in the highest rated cities.

4. Research on Multi dimensional Indicators of Accessibility in Integrated Road Networks

This article aims to integrate multidimensional indicators such as road network accessibility, vehicle travel accessibility, and people's perception of accessibility, and then use a comprehensive evaluation [5]. Index system, including city size, ecological environment, cultural heritage, transportation convenience, climate conditions, and food characteristics, to score 352 cities. Through this evaluation method, cities are comprehensively ranked and the top 50 cities are selected. The basis for choosing Analytic Hierarchy Process (AHP) is a large amount of subjective data based on personal feelings in written descriptions or ratings from netizens [6]. When dealing with subjective data, using objective models may not yield highly interpretable evaluation results. The Analytic Hierarchy Process (AHP), as an essentially subjective method, relies on calculating scores through subjectively determined weights in a comparison matrix. In addition, the paper "Research on Tourist Attraction Evaluation Model Based on AHP Method" written by Professor Yan Xuebo and published in the journal "Western Tourism" also highly praised this method. Therefore, it is appropriate to use Analytic Hierarchy Process (AHP) as the main solution model for this problem. At the same time, in order to enhance the effectiveness of AHP, we will also introduce other indicators to make its evaluation results more comprehensive and rich. The following section will demonstrate through examples how to use these indicators to apply the Analytic Hierarchy Process (AHP).

Firstly, the highest rated attractions in each city were selected from 352 cities. When encountering multiple attractions with the same highest rating in certain cities, further screening was conducted based on the richness of the attractions' connotations, and finally 352 cities' characteristic attractions were determined.

Then, according to the latest urban classification standards in China in 2024, the scale of these 352 cities was classified as follows:

When dealing with various evaluation indicators involved, this article adopts the technical path of natural language processing (NLP). I have determined the following keyword list by filtering key feature words in the description of tourist attractions and data scraping related web pages (such as those targeting food indicators). The following is an exemplary presentation that presents the format for setting keywords:

Set weights for six indicators, as shown in Table 2:

Table 2. Indicator setting weight.

City size	0.34
Environmental Protection	0.10
Cultural heritage	0.14
convenient transportation	0.12
climate	0.05
delicious food	0.25

5. Planning the Best Tourist Route Based on Greedy Algorithm

In order to meet the demand of foreign tourists entering Guangzhou and visiting the most cities within 144 hours, the Analytic Hierarchy Process (AHP) was first used to determine the list of the "50 most desirable cities for foreign tourists" based on factors such as city size, environment, cultural heritage, transportation convenience, climate, and cuisine [7]. Then, following the principle of "the best tourist attraction in the city", combined with greedy algorithm and simulated annealing algorithm, plan the best tourist route to ensure that tourists can experience the richest Chinese urban scenery in a limited time.

5.1. Model establishment

Determine the roadmap for 50 cities to calculate the optimal outcome as shown in Figure 4:

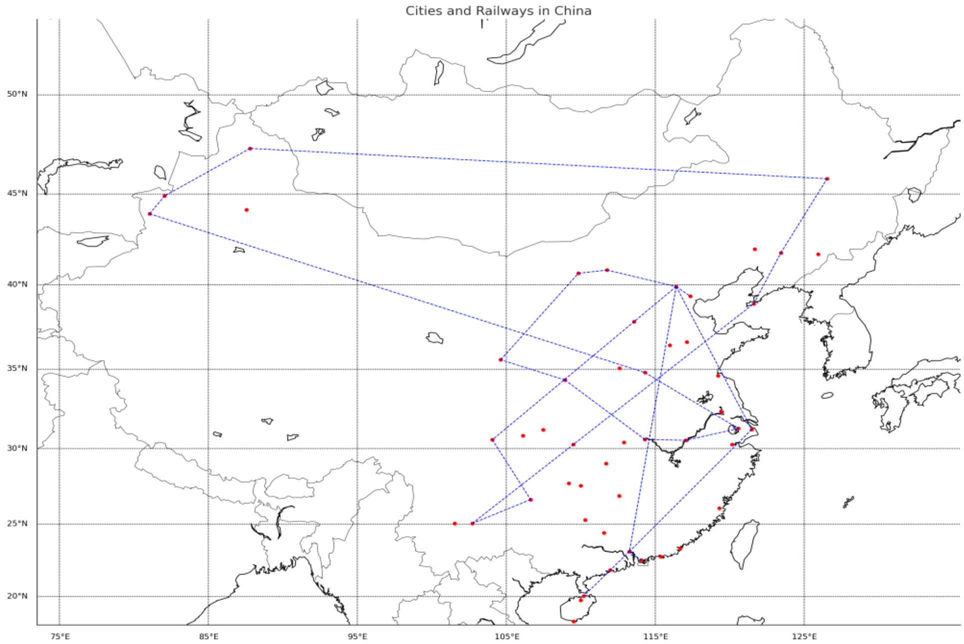

Figure 4. City Route Map.

The application of Greedy Algorithm in Traveling Salesman Problem (TSP) is to select the optimal local selection in the current state at each step, in order to achieve the global optimum [8].

5.2. Model solving

Optimized route: Guangzhou → Shenzhen → Haikou → Chengdu → Guiyang → Kunming → Dalian → Shenyang → Harbin → Beijing, total time: 77.5 hours, total cost: 2500 CNY See Figure 5 and Figure 6.

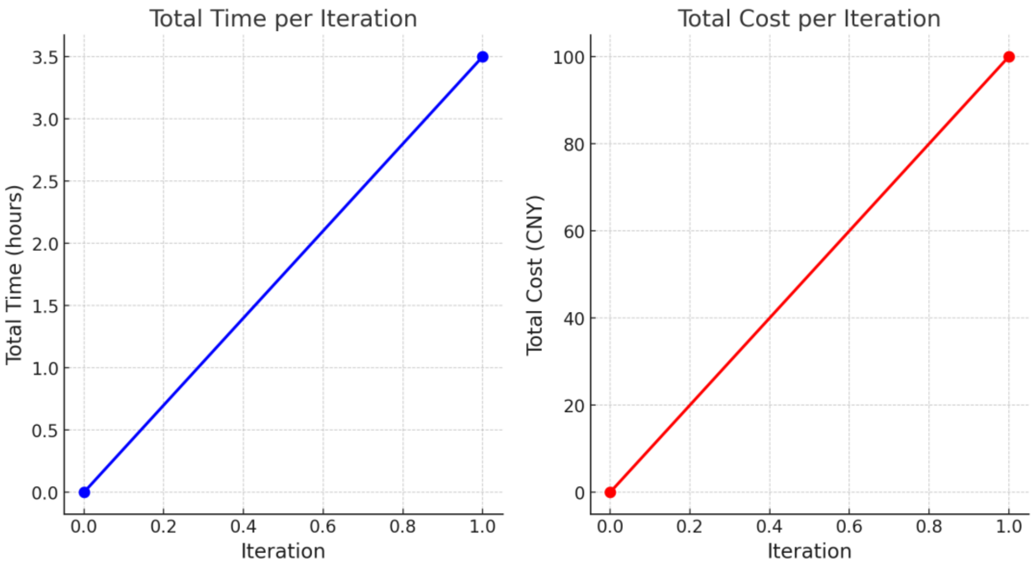

Figure 5. Iterative Process Chart.

Model visualization, as shown in Figure 6:

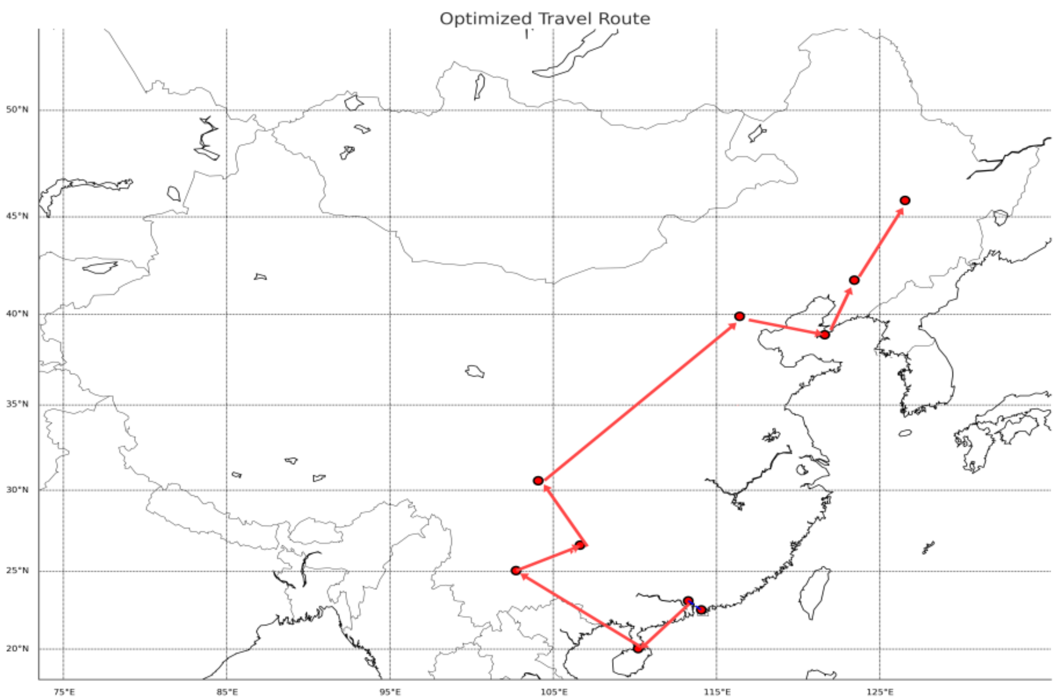

Figure 6. Optimal Route Map.

6. Research on Optimal Tourism Route Based on Greedy Algorithm

The goal of this article is to plan a tourism route that can visit as many cities as possible while minimizing the total cost of tickets and transportation, in order to achieve the optimal balance between time and economy. Departing from Guangzhou, travel to as many cities as possible within a time limit of no more than 88 hours (taking into account factors such as visa expiration, full schedule, and lifestyle habits, the travel time is increased to 144-7 * 8=88 hours, in order to provide foreign tourists with a better understanding of Chinese culture and tourism), while minimizing the total cost of tickets and transportation. The greedy algorithm will be used to achieve this goal [9] [10].

6.1. Model establishment

Determine the route locations of 50 cities as shown in Figure 7:

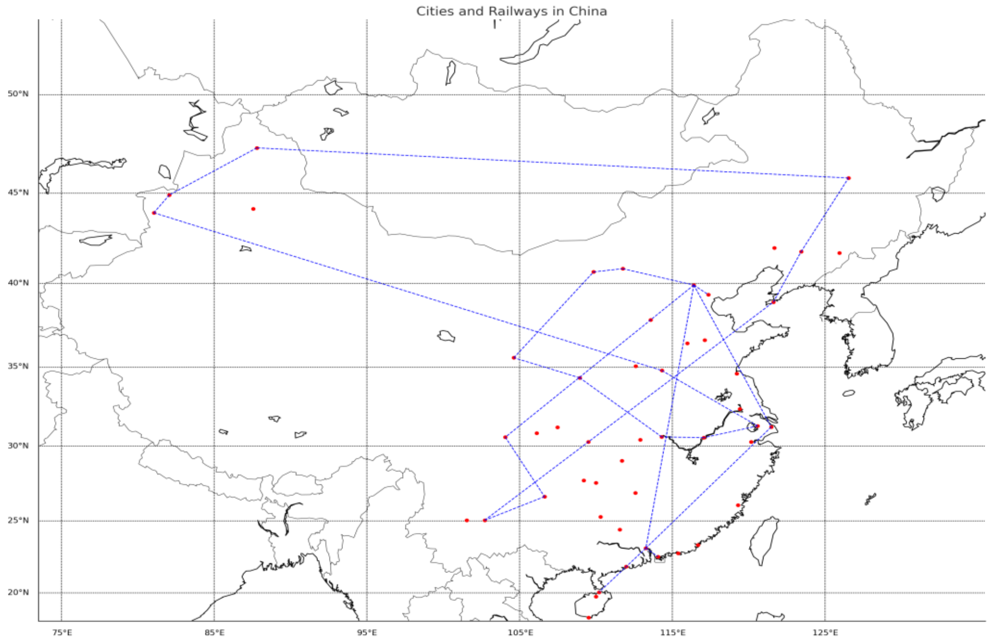

Figure 7. City Route Map.

6.2. 6.2 Model solving

Final result route: Guangzhou → Shenzhen → Haikou → Guiyang → Chengdu → Kunming → Dalian → Shenyang → Harbin → Beijing → Shanghai → Jinan, total time: 83.5 hours, total cost: 3303 CNY, visualized results as shown in Figure 8:

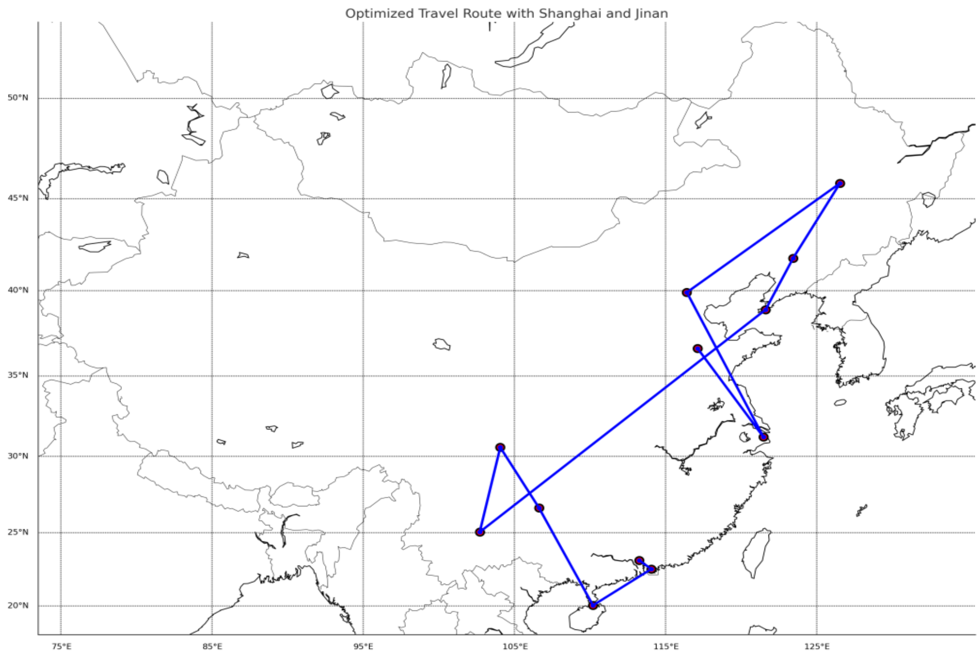

Figure 8. Best Route Map.

7. Conclusions

The relaxation of China's visa-free entry policy has effectively boosted the tourism industry and enhanced the country's international image. Utilizing efficient algorithms with Python and the Pandas library, we have identified top tourist attractions nationwide. Through the application of the Analytic Hierarchy Process (AHP) and a comprehensive evaluation system, we have selected the top 50 cities most appealing to foreign tourists. The multi-objective 0-1 planning model, combined with optimization algorithms, has led to the design of efficient and cost-effective travel itineraries. Moreover, tailored models for mountain scenery preferences highlight China's rich natural resources these findings demonstrate the significance of scientific route planning in enhancing tourist experiences and optimizing resource allocation.

References

[1] Xu Xiaobo, Zhao Lei, Liu Binyi, Wu Biju, Zhong Lina. Characteristics and Variations of Perception of China's Tourist City Images [J]. Geographical Research, 2015, 43(7): 1367-1379.

[2] Yan Xuebo, Zhu Duoman. Research on Tourist Attraction Evaluation Model Based on AHP Method [J]. Western Tourism, 2023, (06): 4-7

[3] Xie Shuangyu, Zhang Qi, Gong Jian, et al. Construction and Application of a Comprehensive Evaluation Model for Accessibility of Urban Tourist Attractions: A Case Study of the Main Urban Area of Wuhan City [J]. Economic Geography, 2019, 39 (03): 232-239.

[4] Zhang Shaopeng, Han Lin. Evaluation Study on Environmental Comfort Degree of Main Tourist Cities in China [J]. Geographical Science Research, 2020, 9(4): 215-222.

[5] Yang Xuanye, CaO Weiwei. Evaluation of Excessive Tourism in Typical Chinese Tourist Cities [J]. Land and Natural Resources Research, 2021, (2): 85-91.

[6] Wang Xinyue, Lu Xuejing, Zhu Wenliang. Analysis and Evaluation of Influencing Factors on the Development of Tourism Industry in Main Tourist Cities of China [J]. Economic Geography, 2020, 44(5): 198-209.

[7] Feng Dongfa, Zhang Tao, Li Ao, Gong Runa. Parameter estimation and application of heterogeneous spatial stochastic frontier model [J]. Statistical Research, 2023, 40 (01): 144-156

[8] Su Shufeng, Pan Kuanyou. Evaluation of Enterprise Technological Innovation Capability Based on AHP-Entropy Weight TOPSIS Model [J]. Logistics Engineering and Management. 2021, 43(11).

[9] Cai Muyang, Shi Jizhi, He Xiaorong. From Tourism to Leisure: Strategies for the Transformation of Chinese Tourist Cities [J]. Economic Geography, 2020, 44(11): 225-231.

[10] Su Haiyang, Chen Chaolong. Progress and Enlightenment in Urban Tourism Planning in China [J]. Journal of Guangzhou City Vocational College, 2019, (3): 28-33, 51.