

Article

Intelligent Geo-Tour Route Recommendation Algorithm Based on Feature Text Mining and Spatial Accessibility Model

Xiao Zhou ^{1,2,3} , Zheng Zhang ^{2,*}, Xinjian Liang ^{3,*}  and Mingzhan Su ² 

¹ Institute of Culture and Tourism, Leshan Vocational and Technical College, Leshan 614000, China; zhouxiao@infu.ac.cn

² Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China; smzh625@infu.ac.cn

³ Center for Southeast Asian Economic and Culture Studies, Chengdu Normal University, Chengdu 611130, China

* Correspondence: giser_zzy@infu.ac.cn (Z.Z.); liangxinjian@cdnu.edu.cn (X.L.)

Abstract: In view of the problems in planning and recommending tour routes, this paper constructs a feature text mining (FTM) method and spatial accessibility model (SAM) as the key factors for scenic spot recommendation (SSR) and tour route recommendation (TRR). The scenic spot clustering algorithm (SSCA) based on FTM was constructed by tourists' text evaluation data mining. Considering the spatial attributes of scenic spots, the scenic spot topology tree algorithm (SSTTA) based on dynamic buffer spatial accessibility (DBSA) was constructed. The optimal scenic spots were recommended based on interest matching and spatial accessibility optimization. As to the recommended scenic spots, this paper proposes an optimal tour route recommendation algorithm (TRRA) based on SSTTA, which aims to determine the optimal adjacent section path structure tree (ASPST) with the lowest cost under travel constraints and transportation modes. The experiment verifies that the proposed algorithm can recommend scenic spots that match tourists' interests and have optimal spatial accessibility, and the optimal tour routes with the lowest costs under certain travel constraints. Compared with the searched sub-optimal tour routes, the optimal tour route recommended by the proposed algorithm produces the lowest travel costs, and all the scenic spots in the tour route meet the tourists' interests. Compared with the commonly used BDMA and GDMA methods, the proposed algorithm can determine the optimal routes with lower travel costs.

Keywords: tour route recommendation (TRR); scenic spot clustering algorithm (SSCA); feature text mining (FTM); spatial accessibility model (SAM)



Citation: Zhou, X.; Zhang, Z.; Liang, X.; Su, M. Intelligent Geo-Tour Route Recommendation Algorithm Based on Feature Text Mining and Spatial Accessibility Model. *Electronics* **2024**, *13*, 1845. <https://doi.org/10.3390/electronics13101845>

Academic Editors: Hyunsoo Yoon and Namgi Kim

Received: 7 April 2024

Revised: 28 April 2024

Accepted: 7 May 2024

Published: 9 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Research Background and Problem Analysis

The core aim of intelligent tourism recommendation systems is to provide scenic spots and routes for tourists. There are many ways to recommend scenic spots, for instance, predicting tourists' interests based on their browsing activities on tourism websites, and recommending scenic spots based on the click frequencies. The recommendation system searches for users who have similar behavior habits to the tourists and recommends similar users' favorite scenic spots to the current users. In addition, scenic spots can be recommended through the user's subjective evaluation scores. Based on tourism big data, the system recommends the popular routes with high scores evaluated by previous tourists. However, existing tourism recommendation systems have certain deficiencies and limitations [1,2]. The user's behavior when browsing tourism websites is random, and the clicking frequencies that really represent tourists' interests account for a small proportion of a large number of clicking behaviors. Meanwhile, tourists have relatively low awareness of unfamiliar tourism cities and urban scenic spots. Clicking behaviors only indicate that tourists have a certain preference for the scenic spot's name and introduction, which does

not represent whether the feature attributes of the scenic spot can really meet the tourists' interests. By evaluating previous tourists whose behaviors are similar to the current tourists, the system recommends the scenic spots and tour routes that the previous tourists have visited. This method is designed on user similarity, and the recommendation results may not meet the tourists' requirements. Because of the individual differences, the behaviors of previous tourists cannot directly represent the real interests of the current tourists. The method to recommend scenic spots by users' evaluation scores takes scenic spots as the basic unit. Tourists' scoring on scenic spots relies on their subjective judgments, since they lack knowledge of scenic spots, such as the tourism functions and feature attributes of the scenic spots. The recommended scenic spots may not match the tourists' interests [3,4].

1.2. Related Theoretical Foundations

For the target of scenic spot recommendation, it is necessary to consider the tourist interest features and scenic spot attributes, including feature attributes and spatial attributes. By constructing the matching relationship between tourist interest needs and scenic spot attributes, a matching model between tourist demand data and scenic spot functions could be directly established, thereby recommending the optimal scenic spots for tourists. On the basis of recommending the best scenic spots, the proposed algorithm integrates urban geospatial data and searches for the most cost-effective tour routes according to different travel modes, which can minimize travel costs and improve tourist satisfaction. According to the research objectives, the theoretical foundations involved in constructing the model mainly include the following aspects.

- (1) Text-mining method. We use the text-mining algorithm to obtain the feature attributes of scenic spots. Text information, such as descriptions of scenic spots, evaluations of tourists, and related literature, serves as the data source for mining the attributes of scenic spots. The label attributes are quantified by using the text statistics method to obtain the clustering features of text labels in the different attribute categories and obtain the attribution of scenic spots.
- (2) Clustering method. The clustering method is an unsupervised learning method in data mining that clusters objects with similar attribute relationships into one class, while objects with distant attribute relationships are dispersed into different classes. By using text statistics and label quantification, the clustering degree of each scenic spot on the text attribute in each cluster is obtained, thereby achieving scenic spot clustering.
- (3) Spatial accessibility model. Spatial accessibility describes the difficulty of traveling from a certain point to a destination point, which is usually measured by the straight-line distance or travel distance between two points. We adopted the spatial accessibility method to measure the accessibility difficulty between the current starting point and the target scenic spot in order to dynamically search for the scenic spot with the best spatial accessibility.
- (4) Graph theory method. Graph theory is a method of abstracting real-world problems into graphics and solving the problems via geometric calculation methods. The urban road network used in our proposed route algorithm is suitable for abstracting as a graph. Scenic spots and road nodes can be abstracted as vertices of the graph, and the roads connecting scenic spots and nodes can be abstracted as edges of the graph. According to the algorithm definition, the constructed graphs are all closed graphs, and the weighted edges are all unidirectional flowing edges.
- (5) Route-planning method. The route-planning method is a type of method that mainly plans routes that meet preset conditions based on the spatial points as the framework. Our proposed route-planning method aims to minimize travel costs, including two categories: minimizing time costs and minimizing the cost of expenses. By searching for road nodes between the starting point and target point, we constructed the sub-intervals between points and searched for the shortest path in each sub-interval, which could iteratively generate a route with the lowest cost in the entire domain.

1.3. The Structure and Research Content of the Work

Based on the analysis of the background and theoretical foundations, we constructed an intelligent route-recommendation algorithm based on feature text mining and the spatial accessibility model. The structure and research content of this work are as follows.

- (1) Introduction. Elaborates on the background of the problem, theoretical foundations, and work structure;
- (2) Related work analysis. Summarizes and analyzes the literature closely related to the research content in recent years, compare the research content with relevant literature, and summarize the advantages of our study;
- (3) Methodology. The research methodology consists of three parts:
 - ① Constructing an unsupervised clustering algorithm for scenic spots based on feature text mining and implementing a text-mining algorithm to cluster scenic spots in feature attributes within the research domain, which provides preconditions for the scenic spot recommendation;
 - ② Constructing a scenic spot topology tree algorithm based on spatial accessibility dynamic buffer searching to recommend scenic spots for tourists from the perspectives of feature attribute matching and optimal spatial accessibility;
 - ③ Constructing an optimal tour route planning algorithm based on the scenic spot path structure tree, which searches for the route with the lowest travel cost for tourists based on the recommended scenic spots.
- (4) Experiment and result analysis. Designs a sample experiment using the tourism city of Chengdu to verify the feasibility of the proposed algorithm. At the same time, the comparative experiment is designed and compared with the electronic maps commonly used for planning tourism routes, and obtain the advantages of the proposed algorithm;
- (5) Conclusion and prospect. Summarizes the research content of the whole study and discusses the limitations of the proposed method and the improvements in the next steps of research.

2. Related Work

2.1. Analysis of Related Work

Yun et al. [5] used the Naive Bayes model to evaluate users' interests. This method finally improved the collaborative filtering algorithm's performance and increased the recommendation accuracy. Lin [6] proposed a depth CTR recommendation model based on the gradient lifting tree and factor decomposer, which improved the performance of the recommendation system. Pacheco et al. [7] brought forward a deep-learning recommendation system on scenic spots based on the Internet of Things and improved the system's performance. Han [8] proposed a scenic spot recommendation method, combining geographical labeled photos with time, space, and visual embedding. Zheng et al. [9] proposed a two-stage greedy algorithm based on crowd sensing to solve the problem of TRR. The crowd sensing score mechanism was used to explore the locations of POIs surrounding scenic spots and the influence of user evaluations of scenic spots, by which the tour routes were planned. Lyu [10] used the convolutional neural network to construct a tour route-planning algorithm, which had good learning control ability and convergence performance. It could improve the reliability of tour route planning. Niu [11] proposed a smart tour guide system based on attraction positioning and recommendation. The RankSVM + time algorithm and K-means clustering algorithm were used to obtain the scenic spot recommendation, and the route algorithm was then set up to plan tourist routes. Finally, a smart tour guide system was set up. Li et al. [12] constructed a tour route algorithm based on an improved knowledge ant colony algorithm, taking the maximization of the overall satisfaction of all tourist groups as the objective function. The algorithm can determine the optimal route with higher efficiency.

Based on the above analysis, the current research on SSR and the TRR mainly focuses on optimizing the algorithm performance and efficiency, or increasing the recommendation accuracy, searching tour routes with high algorithm efficiency, etc. Therefore, SSR and TRR still have the following problems.

- (1) They usually ignore the studies on tourists' interests and focus on mathematical issues, such as algorithm optimization and computational speed. This type of research mode has certain significance for improving the recommendation algorithm's performance, but it is difficult for individual tourists to obtain the maximum motive benefits by recommending scenic spots and tour routes while ignoring their interests;
- (2) The research on the functional attributes of scenic spots is insufficient, especially on the scenic spot categories and the scenic spots' capacities to match tourists' interests. This problem results in the fact that the recommended scenic spots do not fully match tourists' interests, further causing the planned tour routes to not satisfy their needs, which greatly reduces the tourists' motive benefits. The core issue of tour route planning is always to meet the tourists' satisfaction. However, some research has not fully considered the core issue, as well as urban geospatial conditions, urban traffic conditions, and other factors.

2.2. The Superiority of the Proposed Method

To solve the problems, we constructed an intelligent geo-tour route-recommendation algorithm (TRRA) based on feature text mining (FTM) and the spatial accessibility model (SAM). The scenic spot clustering algorithm (SSCA) was constructed by text data mining, and the scenic spots were clustered by SSCA. Using the text labels as the interest data standard, we constructed the tourist interest tendency model to obtain the matching relationship between the scenic spots and the tourists' interests, and then determined the recommended category and scenic spots for the tourists. In view of the particularity and complexity of the urban geospatial environment, and considering the tourists' travel conditions, the spatial distribution and spatial accessibility of urban scenic spots were considered as the important factors for recommending scenic spots and planning tour routes. Combined with the tourists' preferences in choosing transportation modes, the TRRA based on SAM was constructed. Through the proposed method, the tour routes could not only meet the tourists' interests, but also have an optimal geospatial distribution. On the premise of meeting tourists' interests, the recommended tour routes could reduce travel costs and improve tourists' motive benefits [13,14].

2.2.1. Comparison with Literature Work

Based on the analysis of existing studies, the proposed algorithm has significant advantages over the literature [5–12].

Reference [5] used the Naive Bayes method and collaborative filtering method to recommend scenic spots. The Naive Bayes classification algorithm itself has flaws. In the case of a missing quantity or score of certain attribute features, it might cause the calculation result of conditional probability to be 0, resulting in sample classification error and causing interest matching bias. The collaborative filtering algorithm is essentially a recommendation based on the similarity of users or scenic spots, and it does not set up the algorithm from the perspective of the current tourists' interest matching. The advantage of the proposed algorithm lies in obtaining accurate interest evaluations of tourists, and then recommending scenic spots with the best-matched feature attributes. It has no defects and higher accuracy.

The algorithms constructed in references [6,7] aim to improve the AUC (area under the ROC curve), cross entropy, recall, accuracy, etc. of recommendation algorithms, and they still make improvements from the perspective of optimizing algorithm performance, ignoring the core issues of interest matching and maximizing dynamic benefits. The advantage of our proposed algorithm lies in its exploration of tourist interests, scenic feature attributes, and route optimization, with the goal of searching for the global optimal

solution. From the perspective of satisfying tourist interests and motive interests, it searches for the best-matched scenic spot and the most cost-effective route.

Reference [8] adopted the method of mining geographical photos to obtain the historical tourism positioning and interests of tourists, thereby optimizing the tourism recommendation model. The mode of mining and collecting historical photos still relies on mining historical data to obtain recommendations that approximate interests; it cannot represent the current tourists' interests, which is a flaw in mining historical data for interest research. The advantage of our proposed algorithm lies in avoiding the mining of historical data and directly using current tourist interest data and current tourism city geospatial data as sources for matching the best scenic spots and searching for the best routes, resulting in higher accuracy.

Reference [9] recommended tour routes based on crowd perception, without recommending scenic spots from the perspective of individual tourist interest and matching with feature attributes, nor did it present a complete specific algorithm for planning the lowest-cost route based on the best-matched scenic spots. The advantage of our proposed algorithm lies in recommending the best-matched scenic spots based on tourist interests, and searching for the lowest-cost routes based on the current geographical and spatial constraints of tourism cities. It not only meets the personalized needs, but also minimizes the travel costs.

The method constructed in reference [10] aims to optimize the learning control ability and convergence performance of the route-planning algorithm, while the algorithm constructed in reference [11] aims to improve the speed and accuracy. Essentially, it is still an improvement on the computer algorithm, without considering the personalized needs of tourists and the attributes of scenic spots, and without considering the geospatial constraints to search for the route with the lowest cost. The advantage of our proposed algorithm is that it can search for the optimal scenic spots and routes to meet the interests of tourists and reduce costs.

The method constructed in reference [12] aims to meet the needs of tourist groups and improve the revenue of tourism enterprises without considering the personalized needs of individual tourists. When constructing a route algorithm, the goal is to improve the algorithm's computational speed, which is essentially optimizing the algorithm, rather than improving tourist satisfaction. The advantage of our proposed algorithm lies in meeting the interests of individual tourists and searching for the global lowest-cost route.

2.2.2. Comparison with the Commonly Used Route-Planning Methods

In actual tourism activities, the route-planning tool used by tourism enterprises, travel agencies, or tourists is an electronic map. The most commonly used electronic maps are Baidu Map, Gaode Map, etc. Due to the influence of the GPS navigation hardware usage environment and the limitations of software map data, there are some insurmountable problems in using electronic maps to plan routes:

- (1) In order to avoid certain factors, such as traffic lights, intersections, and congested road conditions, electronic maps sometimes recommend longer routes to users without considering transportation modes, resulting in increased travel costs;
- (2) The recommended route is the overall route and cannot change the shortest path and ferrying mode between route nodes based on user needs and changes in transportation modes;
- (3) It is easy to overlook secondary roads and road nodes with identical functions, resulting in the searched route potentially not being the shortest. Especially for those cities with very complex road structure networks, the recommended number of routes is generally limited to 3–4, and due to the complexity of the road network and many nodes, other alternative optimal routes might be ignored. The map-searching algorithm may output local or approximate optimal solutions, rather than global optimal solutions.

In comparison, the proposed algorithm can effectively solve these problems. It considers two modes: the optimal travel time and the lowest route cost. It takes the subspace between two scenic spots as the local area of the searching route and collects all nodes within the subspace and traverses all nodes to search for the global optimal solution. The algorithm considers different transportation modes. The experiment shows that, when the transportation modes are “taxis” and “buses + shared bicycles (or walking)”, the overall performance of the algorithm is stable, and it can determine routes with lower costs than Baidu Map and Gaode Map in each sub-interval, making the overall output route cost the lowest. According to the calculation results, when the transportation mode is “taxi”, the cost weight of the proposed algorithm is 1.079 higher than Baidu Maps and 1.051 higher than Gaode Maps; When the transportation mode is “bus + shared bicycles (or walking)”, the cost weight of the proposed algorithm is 1.126 higher than Baidu Maps and 0.943 higher than Gaode Maps. The higher the weight is, the lower the total cost of the line will be. By comparison, it can be seen that the proposed algorithm is more cost-effective than Baidu Maps and Gaode Maps in terms of travel costs.

3. Methodology

Tourists usually have a habit of evaluating their travel experience after visiting scenic spots. They often provide subjective and genuine text evaluations. The text evaluations of a scenic spot by tourists imply their interest preferences for scenic spots and can reflect the basic functional attributes of the scenic spot. By mining and analyzing a large number of historical tourists’ text evaluations, we used the unsupervised learning method to build the SSCA and obtain the functional categories of scenic spots in the form of clustering. The purpose of the SSCA is to determine the functional attributes of scenic spots to meet tourists’ interests and provide structured data for matching the tourists’ interests and designing the scenic spot recommendation algorithm. Therefore, the design of a tourist interest evaluation matrix based on FTM can help in obtaining the tendencies of tourists’ interest in the functional attributes of tourism. Through mining the interest tendencies for the functional attributes and matching with the functional attributes of scenic spots, the SSRA based on tourists’ interests was constructed. In the urban geospatial environment, tourists will be constrained by the geospatial conditions when traveling. Therefore, optimal tour route planning based on the recommended scenic spots must consider the real-world tourism conditions. Urban scenic spots are distributed in different geographical locations, with different spatial attributes and accessibilities. The higher accessibility a scenic spot has, the easier it is for tourists to get to, and the lower the travel cost will be, and vice versa. The spatial accessibilities of scenic spots have a direct impact on travel costs. For two scenic spots that both satisfy tourists’ interests, the scenic spot with higher accessibility has superiority in distance and cost. Thus, considering the actual travel conditions, it is necessary to combine the spatial attributes and accessibilities as factors to construct the SSRA to ensure that the recommended scenic spots are optimal in functional attributes and spatial accessibilities, and can well match the tourists’ interests and travel conditions [15–17]. Based on the recommended scenic spots, the TRRA was constructed with the different transportation modes selected by tourists.

3.1. The Modeling of the SSCA Based on FTM

Historical tourists’ text evaluations of scenic spots contain massive interest knowledge. Mining scenic spot evaluation text can obtain the interest tendency for the scenic spot category [18–22]. The principle of modeling is as follows: Divide several unknown clusters based on the features and functional attributes of the urban scenic spots, and each cluster will contain a certain number of scenic spots. According to the clustering features and functional attributes, expand the feature text of each unknown cluster, which represents its internal features. Based on the feature text of the different unknown clusters, the interest mining algorithm for the scenic spot evaluation text was constructed to obtain the

functional attribute tendency for the scenic spots, and then finally obtain the cluster for each scenic spot.

Definition 1. Unknown scenic spot cluster $S_{(i)}$, scenic spot $s_{(i,j)}$ in cluster $S_{(i)}$, and cluster capacity $n_{(i)}$. The categories used to distinguish the features and functional attributes of the scenic spots are defined as the unknown scenic spot cluster $S_{(i)}$, $0 < i \leq p$, $i, p \in \mathbf{N}$, and p is the number of cluster $S_{(i)}$. A sample scenic spot x in the research domain is grouped into an unknown cluster $S_{(i)}$, and the scenic spot is defined as a cluster of scenic spots $s_{(i,j)}$ in cluster $S_{(i)}$, where i is the cluster number and j is the scenic spot number in cluster $S_{(i)}$. The number of scenic spots x included in the cluster $S_{(i)}$ after the clustering process is defined as the clustering capacity $n_{(i)}$.

Definition 2. Cluster feature text $t_{(i,j)}$, cluster feature text vector $\mathbf{t}_{(i)}$, and cluster feature text matrix \mathbf{T} . Regarding the unknown scenic spot cluster $S_{(i)}$, confirm m the number of text keywords that can stand for the functional attributes and scenic spot features of cluster $S_{(i)}$. Define the text keyword as the cluster feature text, noted as $t_{(i,j)}$, $0 < i \leq p$, $i, p \in \mathbf{N}$, where i corresponds to cluster $S_{(i)}$ and j represents No. j feature text $t_{(i,j)}$ of cluster $S_{(i)}$. The $1 \times m$ dimension vector $\mathbf{t}_{(i)}$ is constructed based on cluster $S_{(i)}$ and feature text $t_{(i,j)}$. The vector contains m number of elements, and each element stores one cluster feature text $t_{(i,j)}$. This vector is defined as the cluster feature text vector. Footmark i order of unknown cluster $S_{(i)}$ is the row of vector $\mathbf{t}_{(i)}$, and footmark j of feature text $t_{(i,j)}$ is the column. A $p \times m$ dimension matrix \mathbf{T} is constructed to store p feature vectors $\mathbf{t}_{(i)}$, which is defined as the cluster feature text matrix.

According to the definition, $t_{(i,j)}$, $\mathbf{t}_{(i)}$, and \mathbf{T} meet the following constraints:

- (1) All of $t_{(i,j)}$ are the text elements, and $\mathbf{t}_{(i)}$ and \mathbf{T} are the text vector and the text matrix respectively, $t_{(i,j)} \neq 0$, $\mathbf{t}_{(i)} \neq \mathbf{0}$, and $\mathbf{T} \neq \mathbf{0}$;
- (2) Any element of $\forall t_{(i,j1)}$ and $\forall t_{(i,j2)}$ can satisfy the condition $t_{(i,j1)} \neq t_{(i,j2)}$;
- (3) The vector $\mathbf{t}_{(i)}$ must be fully ranked, and the rank satisfies $\text{rank}(\mathbf{t}_{(i)})_{co} = m$;
- (4) The row rank and the column rank of \mathbf{T} satisfy $\text{rank}(\mathbf{T})_{ro} = p$ and $\text{rank}(\mathbf{T})_{co} = m$, and any two rows and any two columns are nonlinearly correlated. The no. i row of the matrix corresponds to the vector $\mathbf{t}_{(i)}$.

Formula (1) is constructed from $t_{(i,j)}$, $\mathbf{t}_{(i)}$, and \mathbf{T} .

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_{(1)} \\ \dots \\ \mathbf{t}_{(i)} \\ \dots \\ \mathbf{t}_{(p)} \end{bmatrix}, \mathbf{T} = \begin{bmatrix} t_{(1,1)} & \dots & t_{(1,j)} & \dots & t_{(1,m)} \\ t_{(i,1)} & \dots & t_{(i,j)} & \dots & t_{(i,m)} \\ t_{(p,1)} & & t_{(p,j)} & & t_{(p,m)} \end{bmatrix} \quad (1)$$

Definition 3. Feature text word frequency $f_{(t(i,j))}$, feature text word frequency vector $\mathbf{f}_{(i)}$, and feature text word frequency matrix \mathbf{F} of the sample scenic spot x . For any scenic spot $\forall x$, a large amount of evaluation text on tourism websites is obtained and stored in a structured format to form an original text dataset T_{ext} . The frequency of the cluster feature text $t_{(i,j)}$ of the sample x in the text dataset T_{ext} is defined as the word frequency $f_{(t(i,j))}$ of the feature text. A $1 \times m$ dimension vector $\mathbf{f}_{(i)}$ is constructed corresponding to a feature text vector $\mathbf{t}_{(i)}$ of an unknown cluster. The vector contains m elements, and each element stores one feature text word frequency $f_{(t(i,j))}$ of vector $\mathbf{t}_{(i)}$. This vector is defined as the feature text word frequency vector. Relating to \mathbf{T} , a $p \times m$ dimension matrix \mathbf{F} is constructed to store p feature text word frequency vectors $\mathbf{f}_{(i)}$. The cluster footmark order i is set as the vector row and the feature text $t_{(i,j)}$ footmark j is set as the column.

Regarding the definition, $f_{(t(i,j))}$, $\mathbf{f}_{(i)}$, and \mathbf{F} meet the following constraints:

- (1) $f_{(t(i,j))}$ is a numerical element. $\mathbf{f}_{(i)}$ and \mathbf{F} are the numerical vector and matrix, $f_{(t(i,j))} \neq 0$, $\mathbf{f}_{(i)} \neq \mathbf{0}$, $\mathbf{F} \neq \mathbf{0}$;

- (2) Vector $\mathbf{f}_{(i)}$ must be a full rank, and the rank of the column satisfies $\text{rank}(\mathbf{f}_{(i)})_{co} = m$;
 (3) The row rank and the column rank of matrix \mathbf{F} satisfy $\text{rank}(\mathbf{F})_{ro} = p$ and $\text{rank}(\mathbf{F})_{co} = m$. Row no. i of the matrix corresponds to vector $\mathbf{f}_{(i)}$.

Formula (2) is constructed as the relationship among $f_{(t(i,j))}$, $\mathbf{f}_{(i)}$, and \mathbf{F} .

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}_{(1)} \\ \vdots \\ \mathbf{f}_{(i)} \\ \vdots \\ \mathbf{f}_{(p)} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} f(t_{(1,1)}) & \dots & f(t_{(1,j)}) & \dots & f(t_{(1,m)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(t_{(i,1)}) & \dots & f(t_{(i,j)}) & \dots & f(t_{(i,m)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(t_{(p,1)}) & \dots & f(t_{(p,j)}) & \dots & f(t_{(p,m)}) \end{bmatrix} \quad (2)$$

The sample scenic spots x in the urban geographical space are evaluated by the tourists to obtain the textual expression of their functional attributes. Usually, the tourists evaluate the functional attributes of the scenic spots from the perspective of interests, so that the functional attributes and categories of the scenic spots can be determined by the text knowledge. As to the sample scenic spots x , massive text evaluations are mined and then their feature text word frequencies are obtained. By modeling the feature text word frequency, the function categories of the sample scenic spots x can be obtained and the scenic spots can be classified into certain unknown clusters $S_{(i)}$ as cluster elements $s_{(i,j)}$.

Definition 4. The scenic spot cluster absolute objective function $f_{(x)^i}$ and the scenic spot cluster relative objective function $\Delta f_{(x)^i}$. The scenic spot cluster absolute objective function $f_{(x)^i}$ reflects the feature text word frequency $f_{(t(i,j))}$ statistics of unknown cluster $S_{(i)}$ corresponding to the sample scenic spots x , and it represents the absolute tendency of the sample scenic spots x being gathered to cluster $S_{(i)}$. The scenic spot cluster relative objective function $\Delta f_{(x)^i}$ reflects the proportion of the feature text word frequency $f_{(t(i,j))}$ corresponding to the unknown cluster $S_{(i)}$ of the sample scenic spots x in the total statistical amount of feature text word frequency $f_{(t(i,j))}$ of the total p number of clusters $S_{(i)}$. It represents the relative tendency of the sample scenic spots x being gathered to cluster $S_{(i)}$.

The scenic spot cluster absolute objective function $f_{(x)^i}$ and scenic spot cluster relative objective function $\Delta f_{(x)^i}$ based on FTM and feature text word frequency $f_{(t(i,j))}$ are constructed as Formulas (3) and (4).

$$f_{(x)^i} = \sum_{j=1}^m f(t_{(i,j)}) \quad (3)$$

$$\Delta f_{(x)^i} = \frac{\sum_{j=1}^m f(t_{(i,j)})}{\sum_{i=1}^p \sum_{j=1}^m f(t_{(i,j)})} \quad (4)$$

Definition 5. Scenic spot cluster tendency vector \mathbf{F}_d . The $1 \times p$ dimension descending order vector is constructed based on the scenic spot cluster relative objective function $\Delta f_{(x)^i}$. The vector stores the relative function value $\Delta f_{(x)^i}$ in the form of the maximum heap, and it is defined as the scenic spot cluster tendency vector, noted as \mathbf{F}_d . The vector element is noted as $\mathbf{F}_{d(i)}$.

The vector storage mode meets the maximum heap condition:

- (1) The vector \mathbf{F}_d is full-rank, and the rank is $\text{rank}(\mathbf{F}_d) = p$;
- (2) For any of $\forall i$ and $\forall j$, if there is $i > j$, there must be $\mathbf{F}_{d(i)} \leq \mathbf{F}_{d(j)}$;
- (3) The element $\mathbf{F}_{d(1)}$ corresponds to the maximum function value $\max \Delta f_{(x)^i}$ and the element $\mathbf{F}_{d(p)}$ corresponds to the minimum function value $\min \Delta f_{(x)^i}$.

The SSCA based on FTM is constructed as follows (Algorithm 1). Figure 1 shows the algorithm flow of the SSCA based on FTM.

Algorithm 1: The SSCA based on FTM

Input: Samples x , text data set T_{ext} .

Output: Cluster $S_{(i)}$ with $n_{(i)}$ number of samples $x \sim s_{(i,j)}$.

- Step 1** Preset clusters $S_{(i)}$, $0 < i \leq p$, and $i, p \in \mathbf{N}$. Confirm m number of $t_{(i,j)}$ for each $S_{(i)}$. Initialize $\mathbf{t}_{(i)}$ and \mathbf{T} .
- Step 2** For $S_{(1)}$, search for row no. 1's text $t_{(1,j)}$ of \mathbf{T} in set T_{ext} , traverse $j \sim (0, m]$.
- 2.1 Search for $f_{(t(1,1))}$ of $t_{(1,1)}$, store $f_{(t(1,1))}$ in $\mathbf{f}_{(1,1)}$ of $\mathbf{f}_{(1)}$ relating to $S_{(1)}$.
- 2.2 Search for $f_{(t(1,2))}$ of $t_{(1,2)}$, store $f_{(t(1,2))}$ in $\mathbf{f}_{(1,2)}$ of $\mathbf{f}_{(1)}$ relating to $S_{(1)}$.
- 2.3 Search for $f_{(t(1,j))}$ of $t_{(1,j)}$, store $f_{(t(1,j))}$ in $\mathbf{f}_{(1,j)}$ of $\mathbf{f}_{(1)}$ relating to $S_{(1)}$, till $j = m$. End.
- 2.4 Calculate $f_{(x)^1}$ for $S_{(1)}$.
- Step 3** For $S_{(2)}$, search for row no. 2's text $t_{(2,j)}$ of \mathbf{T} in set T_{ext} , traverse $j \sim (0, m]$.
- 3.1 Repeat 2.1–2.3, search for $f_{(t(2,j))}$ of $t_{(2,j)}$, store $f_{(t(2,j))}$ in $\mathbf{f}_{(2,j)}$ of $\mathbf{f}_{(2)}$ relating to $S_{(2)}$, till $j = m$. End.
- 3.2 Calculate $f_{(x)^2}$ for $S_{(2)}$.
- Repeat Step 3–Step 4. For $\forall S_{(i)}$, search for row no. i 's text $t_{(i,j)}$ of \mathbf{T} in set T_{ext} , traverse $i \sim (2, p]$, $j \sim (0, m]$,
- Step 4** $i, j, p, m \in \mathbf{N}$. Search for $f_{(t(i,j))}$ of $t_{(i,j)}$, store $f_{(t(i,j))}$ in the related $\mathbf{f}_{(i,j)}$ of $\mathbf{f}_{(i)}$ relating to $S_{(i)}$. Generate \mathbf{F} and calculate $f_{(x)^i}$ for $S_{(i)}$.
- Step 5** Based on \mathbf{F} and $f_{(x)^i}$, calculate $\Delta f_{(x)^i}$ and store in \mathbf{F}_d .
- 5.1 Compare $\Delta f_{(x)^1}$ and $\Delta f_{(x)^2}$:
- (1) If $\Delta f_{(x)^1} > \Delta f_{(x)^2}$, store $\Delta f_{(x)^1}$ in $\mathbf{F}_{d(1)}$ and $\Delta f_{(x)^2}$ in $\mathbf{F}_{d(2)}$.
- (2) If $\Delta f_{(x)^1} < \Delta f_{(x)^2}$, store $\Delta f_{(x)^2}$ in $\mathbf{F}_{d(1)}$ and $\Delta f_{(x)^1}$ in $\mathbf{F}_{d(2)}$.
- 5.2 Compare $\Delta f_{(x)^1}$, $\Delta f_{(x)^2}$, and $\Delta f_{(x)^3}$:
- (1) Store the maximum $f_{(x)^i}$ in $\mathbf{F}_{d(1)}$.
- (2) Store the sub-maximum $f_{(x)^i}$ in $\mathbf{F}_{d(2)}$.
- (3) Store the minimum $f_{(x)^i}$ in $\mathbf{F}_{d(3)}$.
- 5.3 Repeat 5.1–5.2, store $f_{(x)^i}$ and traverse $i \sim (3, p]$, till $i = p$. End.
- 5.4 Take the no. 1 $\mathbf{F}_{d(1)}$ as the $\max \Delta f_{(x)^i}$, sample $x \sim s_{(i,j)}$ is clustered into $S_{(i)}$ of the $\mathbf{F}_{d(1)}$. Store x into $S_{(i)}$ and note it as $s_{(i,1)}$.
- Step 6** Repeat Step 2–Step 5, and calculate and cluster all samples $x \sim s_{(i,j)}$ into related $S_{(i)}$. Note $s_{(i,j)}$ in each $S_{(i)}$, $0 < j < n_{(i)}$, and $j, n_{(i)} \in \mathbf{N}$. End.
-

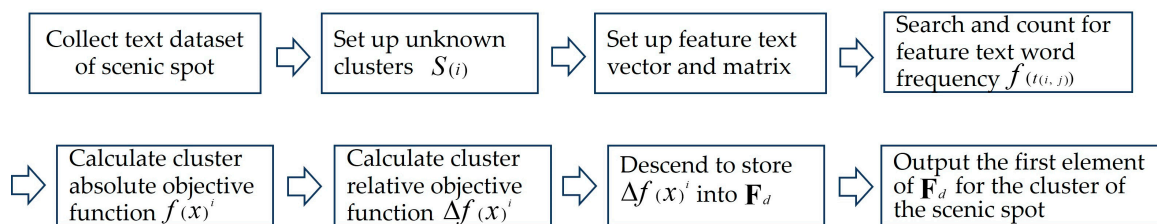


Figure 1. The algorithm flow of the SSCA based on FTM.

3.2. Modeling of the SSTTA Based on the DBSA

The scenic spot cluster knowledge and tourist features mined from the historical tourists' text evaluations reflect the degree to which the tourists prefer an unknown cluster in the aspect of the interest tendency and also reflect the capacity of the scenic spots' functional attributes to match tourists' interests. First, based on the cluster feature text vector $\mathbf{t}_{(i)}$, the feature text is used as the quantitative label. The current tourists select and score the quantitative labels, by which the current tourists' interest tendencies for the scenic spot cluster are obtained. Second, in the aspect of recommending scenic spots, it is necessary to further explore the tourists' interests in a single scenic spot in each cluster. Regarding the feature text of the vector $\mathbf{t}_{(i)}$, the feature text function measurement values of the sample scenic spots are obtained from the urban tourism information service platform and the geographic information service platform. It is then matched with the

tourists' scoring on the feature text so as to obtain the matching degree of the functional attributes between a single scenic spot and the tourists' interests. From the perspective of spatial features, the research domain is a certain geospatial range and the distributed scenic spots have spatial attributes, such as coordinates, azimuth, straight-line distance, travel distance, etc. Therefore, the spatial attributes must be taken into account when recommending tour routes. The process of tourists traveling among scenic spots will arouse time and fee costs, which are determined by the spatial accessibility of scenic spots [23–28]. Regarding the modeling idea, the relationship model between the tourists' interests and the cluster of scenic spots is constructed, and the scenic spot searching model based on DBSA is constructed.

Definition 6. Tourist feature evaluation matrix S_T^{user} and cluster scoring function $f(S_T^{user})^i$. Based on the cluster feature text matrix T , the feature text $t_{(i,j)}$ corresponding to each cluster feature text vector $t_{(i)}$ in matrix T is assigned by the hundred percentage point system score of 100 for pre scoring. The score is $S_{c(i,j)}$, $0 < S_{c(i,j)} \leq 100$, $S_{c(i,j)} \in \mathbf{R}^+$. Construct a $p \times m$ dimension matrix S_T^{user} with the same dimensions as matrix T . Row no. i of matrix S_T^{user} corresponds to the scoring of cluster No. i 's feature text vector $t_{(i)}$, and column no. j represents the scoring of feature text no. j in cluster no. i . The matrix S_T^{user} is defined as the tourist feature evaluation matrix. The scores in matrix S_T^{user} are normalized and iterated to obtain the average statistical score of the tourists on the cluster. The function to obtain the average statistical score is defined as the cluster-scoring function, noted as $f(S_T^{user})^i$. The cluster-scoring function value determines the tourists' degree of interest in cluster $S_{(i)}$. Formulas (5) and (6) are constructed as the tourist feature evaluation matrix S_T^{user} and the cluster scoring function $f(S_T^{user})^i$.

$$S_T^{user} = \begin{bmatrix} S_{c(1,1)} & \cdots & S_{c(1,j)} & \cdots & S_{c(1,m)} \\ \vdots & & \vdots & & \vdots \\ S_{c(i,1)} & \cdots & S_{c(i,j)} & \cdots & S_{c(i,m)} \\ \vdots & & \vdots & & \vdots \\ S_{c(p,1)} & \cdots & S_{c(p,j)} & \cdots & S_{c(p,m)} \end{bmatrix} \quad (5)$$

$$f(S_T^{user})^i = \frac{\sum_{j=1}^m S_{c(i,j)} / 100}{m} \quad (6)$$

Definition 7. Scenic spot $x \sim s_{(i,j)}$ feature measurement matrix $S_T^{s(i,j)}$. Based on the cluster feature text matrix T , the feature text $t_{(i,j)}$ function measurement values $S_{c(i,j)}^t$ of the sample scenic spots $x \sim s_{(i,j)}$ are obtained from the urban tourism information service platform and the geographic information service platform. Each value $S_{c(i,j)}^t$ is scored by the hundred percentage point system, $0 < S_{c(i,j)}^t \leq 100$, $S_{c(i,j)}^t \in \mathbf{R}^+$. Construct a $p \times m$ dimension matrix $S_T^{s(i,j)}$ with the same dimensions as matrix T . Row no. i of matrix $S_T^{s(i,j)}$ corresponds to the measurement value $S_{c(i,j)}^t$ of the no. i cluster feature text vector $t_{(i)}$ of the sample scenic spot $x \sim s_{(i,j)}$, and column no. j represents the measurement value of feature text no. j in cluster no. i . Matrix $S_T^{s(i,j)}$ is defined as the feature measurement matrix of scenic spot $x \sim s_{(i,j)}$.

Definition 8. Scenic spot $x \sim s_{(i,j)}$ feature-measurement function $f(S_T^{user}, S_T^{s(i,j)})^i$. The number of scenic spots in cluster $S_{(i)}$ is $n_{(i)}$. When the sample scenic spot x is listed as a cluster element $s_{(i,j)}$, the measurement value of its cluster $S_{(i)}$ is extracted from the determined scenic spot feature-measurement matrix $S_T^{s(i,j)}$. The measurement value in row no. i of matrix $S_T^{s(i,j)}$ expresses the feature attribute of scenic spot $s_{(i,j)}$ in its cluster. The measurement value that satisfies the tourists' interests is defined as the feature measurement function $S_T^{s(i,j)}$ of scenic spot $x \sim s_{(i,j)}$.

Function $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ is constructed in Formula (7). By calculating the function values, the ability of each scenic spot $s_{(i,j)}$ in the same cluster $S_{(i)}$ to match tourists' interests can be obtained. According to the definition, the higher the measurement function value is, the stronger the ability of a scenic spot to match tourists' interests will be, and vice versa.

$$f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i = \left\{ \sum_{j=1}^m \left[(S_{c(i,j)} - S_{c(i,j)}^t) / 100 \right]^2 \right\}^{-1/2} \quad (7)$$

In the tour route, each scenic spot can be regarded as a space point x . When searching for scenic spots in a tour route, the current scenic spot is reference point y when the tourists have finished visiting it and are preparing to search for the next scenic spot. Therefore, on the basis of ensuring that the scenic spot meets the tourists' interests, the current scenic spot is taken as the reference point to search for the next one, which both meets the tourists' interests and has the best spatial accessibility. According to the determinants of spatial accessibility, tourists are constrained by the urban geospatial conditions when they travel in an urban area. When tourists choose different kinds of transportation modes, the travel time, travel distance, and fee costs will be different.

Definition 9. *Dynamic spatial accessibility function $f_{A(x,y)}$. In urban space, the straight-line distance between space point x and space point y is $d_{(x,y)}$. According to Definition 1, the number of scenic spots in the city meets $n = \sum_{i=1}^p n_{(i)}$. The sample scenic spot in the city is noted as $x_{(i)}$, $0 < i < n$, $i, n \in \mathbf{N}$. If the current reference point y is an arbitrary scenic spot $\forall x_{(i)}$, then the dynamic spatial accessibility function $f_{A(x,y)}$ for scenic spot $\forall x_{(j)}$ in the scenic spot set is constructed as Formula (8), $0 < j < n$, $j, n \in \mathbf{N}$, and $i \neq j$.*

$$f_{A(x(i),x(j))} = \frac{d_{(x(i),x(j))}}{\sum_{j=1}^n d_{(x(i),x(j))}}, j \neq i \quad (8)$$

When the reference scenic spot $x_{(i)}$ changes dynamically, the function value $f_{A(x,y)}$ will change. The search for the best scenic spot is a dynamic process, with the optimal value eliminating the sub-optimal one. Function value $f_{A(x,y)}$ reflects the accessibility measurement between point $x_{(j)}$ and reference point $x_{(i)}$. The higher the function value is, the weaker the accessibility of point $x_{(j)}$ will be, and vice versa. When the dynamic reference point is always the same point, it is a special case for the spatial accessibility function $f_{A(x,y)}$; that is, the dynamic reference point is permanent for each iteration.

Definition 10. *Dynamic searching buffer radius $r_{(a)}$. When searching for the scenic spots for tour route, the current reference point $x_{(i)}$ is taken as the circle center, the radius $r_{(a)}$ is taken as the range to radiate the circular buffer, and the next point $x'_{(i)}$ is searched in the buffer. When the best reference point $x'_{(i)}$ cannot be found in the buffer, the radius is expanded to continue searching.*

Definition 11. *Reference point dynamic storage vector \mathbf{G} . Starting from point A , the dynamic searching is carried out for the n number of scenic spots in all clusters $S_{(i)}$. Under the constraints of cluster $S_{(i)}$, the feature measurement function $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ and the dynamic space accessibility function $f_{A(x,y)}$, a vector with dimension $1 \times k$ is constructed according to the tourists' expected number of scenic spots. The vector is used to store the qualified reference scenic spots, and it is defined as the reference point dynamic storage vector. Finally, the scenic spots stored in the vector will be recommended to the tourists.*

According to the modeling idea and the above definitions, the SSTTA based on DBSA is constructed as follows (Algorithm 2). Figure 2 is the modeling process to construct the SSTTA. Figure 2a represents the scenic spot distribution map, Figure 2b represents the extracted scenic spots and starting point, Figure 2c is the constructed SSTT found by the

searching algorithm, and Figure 2d–f represent the process to search the SSTT in multiple layers of buffers. Figure 3 shows the algorithm flow of the SSTTA based on DBSA.

Algorithm 2: The SSTTA based on DBSA

Input: Clusters $S_{(i)}$, scenic spots $s_{(i,j)}$, matrix \mathbf{S}_T^{user} , radius $r_{(a)}$, k number of to be visited scenic spots, and starting point A .

Output: Vector \mathbf{G} .

- Step 1** Based on \mathbf{T} , tourists make scores $S_{c(i,j)}$ on $t_{(i,j)}$ to form \mathbf{S}_T^{user} . Calculate $f(\mathbf{S}_T^{user})^i$ for $S_{(i)}$, traversing $i \sim (0, p]$, $i, p \in \mathbf{N}$. Select $k_{(i)}$ number of scenic spots to be visited in each $S_{(i)}$.
- Step 2** In $S_{(i)}$, confirm $\mathbf{S}_T^{s(i,j)}$ for each $s_{(i,j)}$. Calculate $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ of $s_{(i,j)}$ based on \mathbf{S}_T^{user} .
- Step 3** Input $k, k_{(i)}, r_{(a)}$, and A , search for the feasible $s_{(i,j)}$.
- 3.1 Confirm no. 1 buffer with $r_{(1)}$ and A . Mark the w_1 number of $s_{(i,j)(v)}$ in the buffer. i is the $S_{(i)}$ number, j is the number for $s_{(i,j)(v)}$ in $S_{(i)}$, v is the number of $s_{(i,j)(v)}$ in buffer $r_{(1)}$, $0 < v \leq w_1$, $v, w_1 \in \mathbf{N}$.
- 3.2 Calculate $f_{A(s(i,j)(v),A)}$ for the w_1 number of $s_{(i,j)(v)}$.
- 3.3 Search for $s_{(i1,j)(1)}$ and make judgement:
- (1) If $s_{(i1,j)(1)} \in \forall S_{(i1)}, k_{(i1)} \neq 0$, store $s_{(i1,j)(1)}$ in $\mathbf{G}_{(1)}$ of the \mathbf{G} .
 - (2) If $s_{(i1,j)(1)} \in \forall S_{(i1)}, k_{(i1)} = 0$, delete $s_{(i1,j)(1)}$. Turn to step 3.4
- 3.4 Search for $s_{(i2,j)(2)}$ and make judgement:
- (1) If $s_{(i2,j)(2)} \in \forall S_{(i2)}, i_1 \neq i_2, k_{(i2)} \neq 0$, store $s_{(i2,j)(2)}$ in $\mathbf{G}_{(2)}$ of the \mathbf{G} .
 - (2) If $s_{(i2,j)(2)} \in \forall S_{(i2)}, i_1 = i_2, k_{(i2)} \neq 0$, compare $s_{(i1,j)(1)}$ and $s_{(i2,j)(2)}$ in $f_{A(s(i,j)(v),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$.
 - ① If $k_{(i2)} = 1$:
 - (I) $f_{A(s(i1,j)(1),A)} >> f_{A(s(i2,j)(2),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ has a slight difference, delete $s_{(i1,j)(1)}$ in \mathbf{G} , store $s_{(i2,j)(2)}$ in $\mathbf{G}_{(1)}$; $f_{A(s(i1,j)(1),A)} << f_{A(s(i2,j)(2),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ has a slight difference, keep $s_{(i1,j)(1)}$ in \mathbf{G} and delete $s_{(i2,j)(2)}$.
 - (II) $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i1,j)(1)})^i >> f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i2,j)(2)})^i$ and $f_{A(s(i,j)(v),A)}$ has a slight difference, keep $s_{(i1,j)(1)}$ in \mathbf{G} and delete $s_{(i2,j)(2)}$; $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i1,j)(1)})^i << f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i2,j)(2)})^i$ and $f_{A(s(i,j)(v),A)}$ has a slight difference, delete $s_{(i1,j)(1)}$ in \mathbf{G} and store $s_{(i2,j)(2)}$ in $\mathbf{G}_{(1)}$.
 - (III) If one higher value of $f_{A(s(i,j)(v),A)}$ or $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ exists, randomly select one scenic spot and store it in $\mathbf{G}_{(1)}$. Delete the other one.
 - ② If $k_{(i2)} > 1$, store $s_{(i2,j)(2)}$ in $\mathbf{G}_{(2)}$ of \mathbf{G} .
 - (3) Judge the number σ of stored scenic spots in \mathbf{G} . If $\sigma = k$, End, output \mathbf{G} ; If $\sigma \neq k$, turn to step 3.5.
- 3.5 Search no. b $s_{(ib,j)(b)}$. Judge $S_{(ib)}$, number $k_{(ib)}$ of scenic spots, $f_{A(s(ib,j)(b),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(ib,j)(b)})^i$, traversing $b \sim (0, w_1]$, $b, w_1 \in \mathbf{N}$.
- (1) If $s_{(ib,j)(b)} \in \forall S_{(ib)}, i_b \neq i_{b'}, k_{(ib)} \neq 0$, store $s_{(ib,j)(b)}$ in $\mathbf{G}_{(\sigma+1)}$ of \mathbf{G} ;
 - (2) If $s_{(ib,j)(b)} \in \forall S_{(ib)}, i_b = i_{b'}, k_{(ib)} \neq 0$, compare $s_{(ib,j)(b)}$ of $S_{(ib)}$ with the other $s_{(ib',j)(b')}$ in $f_{A(s(i,j)(v),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ within the buffer $r_{(1)}$:
 - ① $k_{(ib)}$ reaches the maximum.
 - (I) $f_{A(s(ib,j)(b),A)} << \forall f_{A(s(ib',j)(b'),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ has a slight difference. Delete the maximum $f_{A(s(ib',j)(b'),A)}$ in $s_{(ib',j)(b')}$ of \mathbf{G} , store $s_{(ib,j)(b)}$. Else, delete $s_{(ib,j)(b)}$.
 - (II) $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(ib,j)(b)})^i \geq \forall f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(ib',j)(b')})^i$ and $f_{A(s(i,j)(v),A)}$ has a slight difference. Delete the minimum $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ in $s_{(ib',j)(b')}$ of \mathbf{G} , store $s_{(ib,j)(b)}$. Else, delete $s_{(ib,j)(b)}$.
 - (III) If $f_{A(s(i,j)(v),A)}$ and $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ of $s_{(ib,j)(b)}$ are among the values of the other $s_{(ib',j)(b')}$, delete the maximum $f_{A(s(i,j)(v),A)}$ or the minimum $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$, store $s_{(ib,j)(b)}$.
 - ② $k_{(ib)}$ does not reach the maximum. Store $s_{(ib,j)(b)}$ in $\mathbf{G}_{(\sigma+1)}$ of \mathbf{G} .
-

(3) If arbitrarily one following condition is not satisfied, turn to step 3.6. Else, End, output **G**.

- ① σ reaches $\sigma = k$;
- ② $\forall k_{(i)}$ reaches the maximum.

3.6 Confirm no. 2 buffer with A and $r_{(2)}, r_{(2)} > r_{(1)}$. Mark the w_2 number of $s_{(i,j)(v)}$ in the area between $r_{(1)}$ and $r_{(2)}$. In $s_{(i,j)(v)}$, i is the cluster $S_{(i)}$ number, j is the number of $s_{(i,j)(v)}$ in $S_{(i)}$, v is the number of $s_{(i,j)(v)}$ in the area, $0 < v \leq w_2$, $v, w_2 \in \mathbf{N}$. Repeat 3.1~3.5, search scenic spots, and store in **G**.

- (1) If **G**: σ reaches $\sigma = k$ and $\forall k_{(i)}$ reaches the maximum. End, output **G**;
- (2) Else, expand $r_{(a)}$ and continue searching until σ reaches $\sigma = k$ and $\forall k_{(i)}$ reaches maximum. End, output **G**;

Step 4 Output **G**, extract k number of scenic spots. The scenic spots in vector **G** meet the conditions:

- (1) All $f_{A(x,y)}$ are the minimum ones.
- (2) All $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ are the maximum ones.

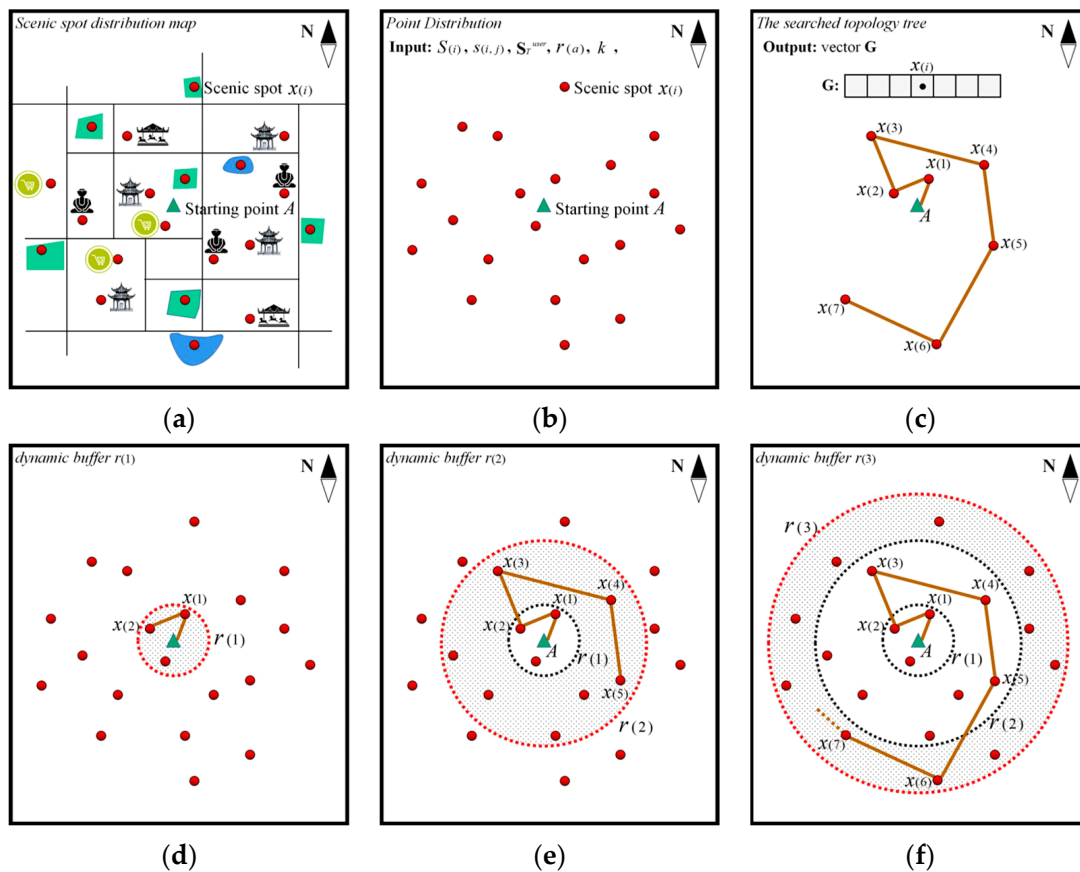


Figure 2. Modeling process for constructing the SSTTA.

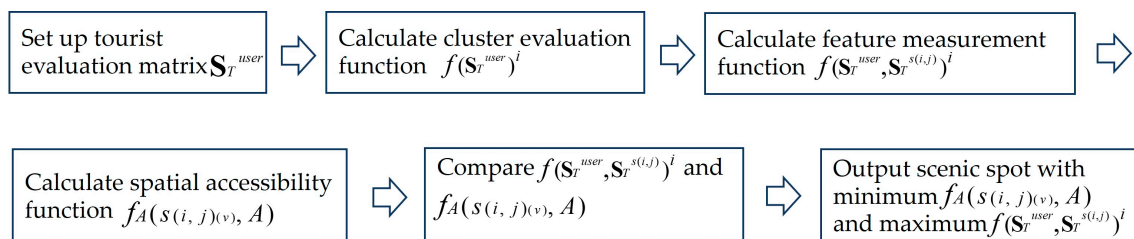


Figure 3. Algorithm flow of the SSTTA based on DBSA.

3.3. Modeling of the OTRA Based on SSTTA

Urban scenic spots are located at different geographical coordinates in the city. When tourists start from starting point A to visit k scenic spots, they are constrained by urban geospatial factors, including:

- (1) The urban roads between the points;
- (2) The transportation mode chosen by the tourists;
- (3) Travel time;
- (4) Travel fees.

The tour route recommendation algorithm (TRRA) is constructed based on the constraints of urban geographical space, and the time cost and fee cost are taken into account. According to the vector \mathbf{G} , the tourists determine one kind of travel sequence according to their interests. The travel sequence is constrained by the geographical conditions and cost factors. When the tourists consider the time cost or fee cost, the tour routes will be different, or the costs will be different [28–31]. According to the analysis, the tour route issue could be transformed into the issue of searching for the route with the lowest time cost or the lowest fee cost under the condition of the travel sequence, the geospatial factors as well as the tourists' travel budgets. The travel sequence is as follows: visit scenic spot $\mathbf{G}_{(i,1)}$ first, then scenic spot $\mathbf{G}_{(i,2)}$, until the following k scenic spots $\mathbf{G}_{(i,k)}$ are visited. The footmark x in the scenic spot $\mathbf{G}_{(i,x)}$ represents an arbitrary scenic spot selected by the tourists, $0 < x \leq k$, $x, k \in \mathbf{N}$.

Definition 12. Tour route structure chain \mathbf{R} and structure chain-adjacent section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$. From the starting point A , a chain is formed by the $k + 1$ points from $\mathbf{G}_{(i,1)}$ to $\mathbf{G}_{(i,k)}$ in the travel sequence; it is defined as the tour route structure chain \mathbf{R} . If element $\mathbf{G}_{(i,x)}$ in chain \mathbf{R} is the current scenic spot, then the next scenic spot to be visited is $\mathbf{G}_{(i,y)}$. The section composed of $\mathbf{G}_{(i,x)}$ and $\mathbf{G}_{(i,y)}$ is defined as the structure chain-adjacent section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$. The adjacent section is the basic unit of the route chain, which is composed of the urban roads and road nodes.

Definition 13. Adjacent section control point $P_{(i)}$ and adjacent section control point set $P_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$. The tourists pass through several urban roads and road nodes in the process of traveling from point $\mathbf{G}_{(i,x)}$ to point $\mathbf{G}_{(i,y)}$. The key road nodes in the adjacent section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$ of the structure chain between points $\mathbf{G}_{(i,x)}$ and $\mathbf{G}_{(i,y)}$ are defined as the adjacent section control points $P_{(i)}$. A set of control points $P_{(i)}$ in section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$ is defined as the adjacent-section control point set $P_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$. Adjacent-section control points are determined by urban geospatial information, and their locations are abstracted from the city map.

Section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$ which, is formed by the arbitrary two adjacent points $\mathbf{G}_{(i,x)}$ and $\mathbf{G}_{(i,y)}$ in chain \mathbf{R} , is regarded as the basic unit, and its coordinates on the map are confirmed. The n number of control points in section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$ are searched and confirmed, $0 < i < n$, $i, n \in \mathbf{N}$. Then, search to confirm the connecting path of the adjacent control points and determine the travel distances between two control points. Abstract the section into the graph edge and abstract the points into the graph vertices. Map the graph onto the two-dimensional space.

Definition 14. Adjacent-section-directed weighted-edge graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$. The graph that is used to describe the relationship between points and paths, adjacent relationship, and quantitative relationship is defined as the adjacent-section-directed weighted-edge graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$. The graph meets $\mathbf{V} = |n + 2|$, $\mathbf{A} = |m|$. The adjacent points represent the urban path. Figure 4 shows the modeling process of the graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$ for the structure chain-adjacent section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$.

Definition 15. Adjacent section feasible path $P_{\text{ath}(c)}$, adjacent section path structure tree $T_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$, adjacent graph edge weight $E_{(\alpha, k)}$ and path weight $E(\text{Path}(c))$. The adjacent section $R_{(\mathbf{G}_{(i,x)}, \mathbf{G}_{(i,y)})}$ is the basic unit. Based on the directed weighted graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$, set the starting point $\mathbf{G}_{(i,x)}$ as the root node and terminal point $\mathbf{G}_{(i,y)}$ as the target node. One feasible path that is formed by the

searching algorithm between $G_{(i,x)}$ and $G_{(i,y)}$ is defined as the adjacent section feasible path $P_{ath(c)}$. Suppose that path $P_{ath(c)}$ has α edge weights. The quantity of α differs along with the path quantity, and the related weights $E_{(\alpha,k)}$ will be different. The structure tree formed by all feasible paths $P_{ath(c)}$ is defined as the adjacent section path structure tree (ASPST), noted as $T_{(G_{(i,x)},G_{(i,y)})}$. Figure 5 shows the modeling process of the ASPST of $T_{(G_{(i,x)},G_{(i,y)})}$ by using Figure 4c as an example.

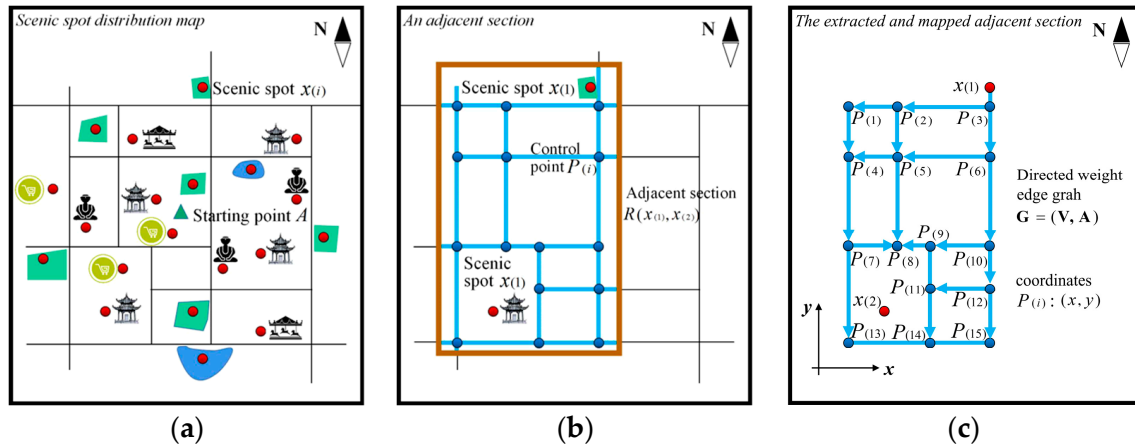


Figure 4. Modeling of the directed-weight edge graph of the structure chain-adjacent section. (a) Original distribution of scenic spots and road network. (b) Adjacent section. (c) Adjacent section control point set.

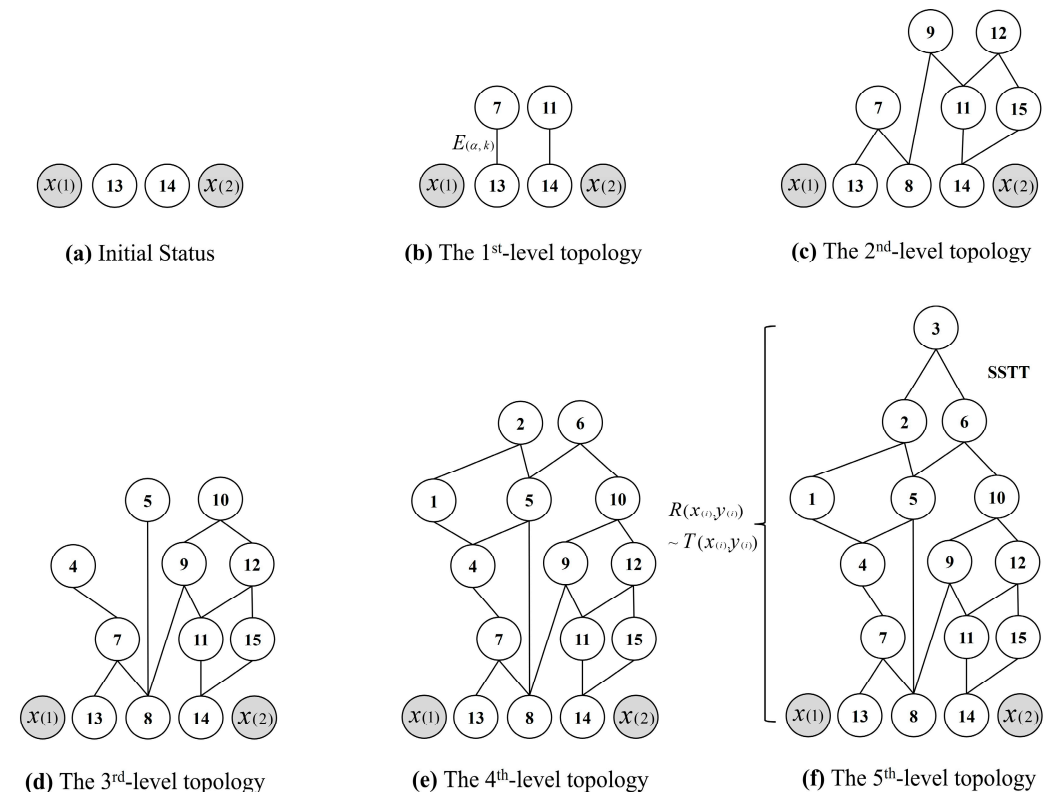


Figure 5. Modeling process of ASPST of $T_{(G_{(i,x)},G_{(i,y)})}$.

Suppose the condition that the ASPST has g kinds of feasible paths, then there is $0 < c \leq g, c, g \in \mathbf{N}$ in $P_{ath(c)}$, in which there is only one path representing the optimal route with the minimum distance cost. In the weighted graph, if the travel distance between the two adjacent points $P_{(i)}$ and $P_{(j)}$ is set as $d_{is(P(i),P(j))(\alpha)}$, the edge weight $E_{(\alpha,k)}$ between the two points is constructed as Formula (9), where α represents the edge of the graph

and k represents $R_{(G(i,x),G(i,y))}$. In a path $P_{ath(c)}$, the total weight calculated by all edge weights $E_{(\alpha,k)}$ is defined as the path weight $E(Path_{(c)})$, constructed as Formula (10). The weight for the tour route formed by k adjacent sections $R_{(G(i,x),G(i,y))}$ is constructed as the Formula (11).

$$E_{(\alpha,k)} = \frac{1}{d_{is(P(i),P(j))(\alpha)}} \quad (9)$$

$$E(Path_{(c)}) = \frac{1}{\sum_{\alpha=1}^{\max \alpha} E_{(\alpha,k)}} \quad (10)$$

$$E_{route} = \frac{1}{\sum_{\alpha=1}^{\max \alpha} E_{(\alpha,1)}} + \dots + \frac{1}{\sum_{\alpha=1}^{\max \alpha} E_{(\alpha,k)}} + \dots + \frac{1}{\sum_{\alpha=1}^{\max \alpha} E_{(\alpha,\max k)}} \quad (11)$$

Definition 16. *Adjacent section feasible path vector P_{ath} . Based on $T_{(G(i,x),G(i,y))}$, the searching algorithm will generate g kinds of feasible paths. Store the path weights $E(Path_{(c)})$ that relate to the g kinds of feasible paths in a $1 \times g$ dimension vector according to the sorting algorithm, the vector is defined as the adjacent section feasible path vector P_{ath} . The vector reflects the path conditions between points $G_{(i,x)}$ and $G_{(i,y)}$, and the optimal route is extracted from the vector.*

According to the definition, after the tourists confirm the starting point A , the reference point storage vector G , k number of scenic spots to be visited, and a kind of traveling sequence, the recommendation system will generate k adjacent sections $R_{(G(i,x),G(i,y))}$, corresponding to the k number of trees $T_{(G(i,x),G(i,y))}$. The k number of adjacent sections $R_{(G(i,x),G(i,y))}$ forms a complete tour route. The optimal TRRA based on SSTTA is constructed (Algorithm 3). Bid is the two-way pass, Uid^+ is the positive one-way pass, and Uid^- is the reversal one-way pass. Figure 6 shows the algorithm flow of the optimal TRRA based on SSTTA.

Algorithm 3: The optimal TRRA based on SSTTA

Input: Starting point A , vector G , and k number of scenic spots to be visited.

Output: Vector P_{ath} , the optimal $P_{ath(c)}$, and E_{route} .

- Step 1** Tourists confirm a tour sequence based on A and G , obtain R . Output k number of $R_{(G(i,x),G(i,y))}$, $0 < i \leq k$, $i, k \in \mathbf{N}$.
- Step 2** For $R_{(G(i,x),G(i,y))}(i)$, take $i = 1$, confirm $R_{(A,G(i,x))}(1)$ and $G_{(1)} = (V, A)$.
- 2.1 Confirm coordinates of A and $G_{(i,x)}$. Search urban roads connecting A and $G_{(i,x)}$. Confirm the n number of $P_{(i)}$.
- 2.2 Map A , $G_{(i,x)}$, and $P_{(i)}$ onto 2D coordinate and form the basic structure $G_{(1)} = (V, A)$.
- 2.3 Search $dis_{(P(i),P(j))(\alpha)}$ among A , $G_{(i,x)}$, and $P_{(i)}$. Calculate $E_{(\alpha)}$ of graph $G_{(1)}$.
- 2.4 Quantify the graph $G_{(1)} = (V, A)$. The quantity vertex is $V = |n + 2|$, edge $A = |m|$.
- Step 3** Construct the tree $T_{(G(i,x),G(i,y))}$.
- 3.1 In $R_{(A,G(i,x))}(1)$, confirm the root node A and the target node $G_{(i,x)}$.
- 3.2 Search nearby $P_{(t)}$ to A , $0 < t \leq \max t < n$, and $t, n \in \mathbf{N}$. Judge the *Access* between point A and $P_{(t)}$:
- (1) If *Access* = *Bid* or *Access* = *Uid*⁺, connect A and $P_{(t)}$ to form the weight edge and assign value $E_{(\alpha)}$;
- (2) If *Access* = *Uid*[−], A and $P_{(t)}$ do not form a positive weight edge and do not connect.
- 3.3 Inspect $P_{(t)}$, $t \sim (0, \max t]$.
- (1) Take $P_{(1)}$ as a vertex, search all $P_{(1,v)}$ connecting $P_{(1)}$, $0 < v \leq \max v < n$, and $v, n \in \mathbf{N}$. Judge *Access* between A and each $P_{(1,v)}$:
- ① If *Access* = *Bid* or *Access* = *Uid*⁺, connect $P_{(1)}$ and $P_{(1,v)}$ to form the weight edge and assign value $E_{(\alpha)}$;
- ② If *Access* = *Uid*[−], $P_{(1)}$ and $P_{(1,v)}$ do not form a positive weight edge and do not connect.
- (2) Take $P_{(2)}$ as a vertex, search all $P_{(2,v)}$ connecting $P_{(2)}$, $0 < v \leq \max v < n$, and $v, n \in \mathbf{N}$. Judge *Access* between $P_{(2)}$ and each $P_{(2,v)}$:
- ① If *Access* = *Bid* or *Access* = *Uid*⁺, connect $P_{(2)}$ and $P_{(2,v)}$ to form the weight edge and assign value $E_{(\alpha)}$;
- ② If *Access* = *Uid*[−], $P_{(2)}$ and $P_{(2,v)}$ do not form a positive weight edge and do not connect.
-

- (3) Take $P_{(t)}$ as a vertex, search all $P_{(t,v)}$ connecting $P_{(t)}$, $0 < v \leq \max v < n$, and $v, n \in \mathbf{N}$. Judge *Access* between $P_{(t)}$ and each $P_{(t,v)}$:
- ① If $Access = Bid$ or $Access = Uid^+$, connect $P_{(t)}$ and $P_{(t,v)}$ to form the weight edge and assign value $E_{(\alpha)}$;
 - ② If $Access = Uid^-$, $P_{(t)}$ and $P_{(t,v)}$ do not form a positive weight edge and do not connect.
- (4) Traverse $t \sim (0, \max t]$, connect all lower level nodes $P_{(t,v)}$ with all $P_{(t)}$.
- 3.4 Expand to the lower-level nodes. Inspect $P_{(t,v)}$, $t \sim (0, \max t]$, and $v \sim (0, \max v]$. Search all lower-level nodes $P_{(t,v,c)}$ of each $P_{(t,v)}$. Judge *Access* between $P_{(t,v)}$ and $P_{(t,v,c)}$:
- (1) If $Access = Bid$ or $Access = Uid^+$, connect $P_{(t,v)}$ and $P_{(t,v,c)}$ to form a weight edge and assign value $E_{(\alpha)}$;
 - (2) If $Access = Uid^-$, $P_{(t,v)}$ and $P_{(t,v,c)}$ do not form a positive weight edge and do not connect.
- 3.5 Traverse all-level $P_{(i)}$, $0 < t \leq n$, and $i, n \in \mathbf{N}$ until target $\mathbf{G}_{(i,x)}$ is searched, End. Output g kinds of $T_{(A,G(i,x))}$.
- Step 4** Store and form vector \mathbf{P}_{ath} with the following steps.
- 4.1 Initialize $\mathbf{P}_{ath} = 0$, element $\mathbf{P}_{ath(c)}$, $0 < c \leq g$, and $c, g \in \mathbf{N}$.
- 4.2 Find $P_{ath(1)}$ and $P_{ath(2)}$, extract $E_{(\alpha)}$ of $P_{ath(1)}$ and $P_{ath(2)}$, calculate $E(Path_{(1)})$ and $E(Path_{(2)})$.
- (1) If $E(Path_{(1)}) \geq E(Path_{(2)})$, store $P_{ath(1)}$ and $P_{ath(2)}$ in $\mathbf{P}_{ath(1)}$ and $\mathbf{P}_{ath(2)}$;
 - (2) If $E(Path_{(1)}) < E(Path_{(2)})$, store $P_{ath(1)}$ and $P_{ath(2)}$ in $\mathbf{P}_{ath(2)}$ and $\mathbf{P}_{ath(1)}$.
- 4.3 Find $P_{ath(3)}$ and calculate $E(Path_{(3)})$, compare $E(Path_{(1)})$, $E(Path_{(2)})$ and $E(Path_{(3)})$.
- (1) Store the maximum $E(Path_{(c)})$ in $\mathbf{P}_{ath(1)}$.
 - (2) Store the sub-maximum $E(Path_{(c)})$ in $\mathbf{P}_{ath(2)}$.
 - (3) Store the minimum $E(Path_{(c)})$ in $\mathbf{P}_{ath(3)}$.
- 4.4 Find other $P_{ath(c)}$ and calculate $E(Path_{(c)})$. Descend to store $P_{ath(c)}$ in \mathbf{P}_{ath} . Traverse $c \sim (0, g]$. The path of $\mathbf{P}_{ath(1)}$ relates to the optimal one in $R_{(A,G(i,x))(1)}$.
- Step 5** Repeat Step 2–Step 4, traverse $i \sim (1, k]$, and find \mathbf{P}_{ath} of $R_{(A,G(i,x))(i)}$ and optimal $P_{ath(c)}$ for each $R_{(G(i,x)A,G(i,y))(i)}$. All $P_{ath(c)}$ form an optimal tour route.
- Step 6** Calculate E_{route} formed by k number of $R_{(G(i,x),G(i,y))}$.

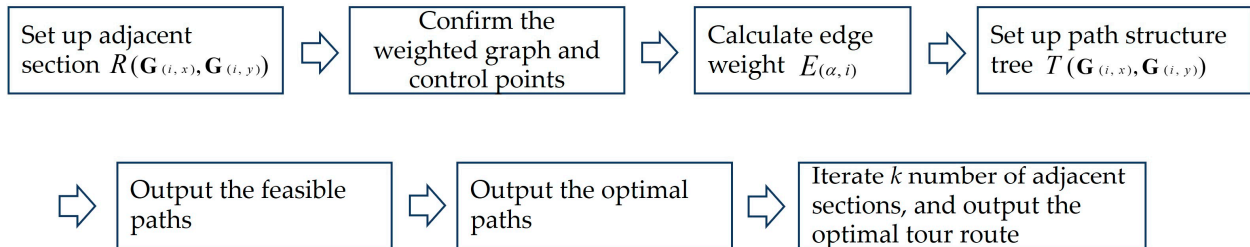


Figure 6. Algorithm flow of the optimal TRRA based on SSTTA.

4. Experiment and Result Analysis

4.1. Basic Principle and Process of the Experiment

The minimum cost path $P_{ath(c)}$ generated by the ASPST of $T_{(G(i,x),G(i,y))}$ represents the minimum cost distance for the tourists to travel between points in the geographic space. When the tourists choose the traveling mode, the system will recommend different transportation modes and adopt the optimal path $P_{ath(c)}$ in each section. Tourists usually consider the time cost and fee cost in the traveling process, which can be divided into two modes:

(1) Only consider the lowest time cost, and do not consider the fee cost. The system recommends “Taxi” as the transportation mode. Based on the tour route structure chain \mathbf{R} and the route algorithm, it outputs the optimal distance path $P_{ath(c)}$ for the taxis in each adjacent section to determine the optimal route.

(2) Only consider the lowest fee cost, and do not consider the time cost. The system recommends the “shared bicycle” mode, and outputs the optimal distance path $P_{ath(c)}$ of the shared bicycle (or walking) in each adjacent section based on the tour route structure chain \mathbf{R} and route algorithm, and then outputs the overall optimal route.

We chose the two modes as examples. The experimental process is as follows.

(1) Confirm the starting point A of the tour and the number of scenic spots k to be visited. Collect the evaluation data of urban scenic spots, and the quantity of text evaluations or introductions is more than 1000 pieces. The constructed SSCA is used to cluster the urban scenic spots to obtain the cluster distribution and cluster capacity.

(2) The tourists score the feature text and the system calculates the cluster scoring function $f(\mathbf{S}_T^{user})^i$, collect the feature measurement matrix $\mathbf{S}_T^{s(i,j)}$ of each scenic spot and calculate their feature measurement function $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$. By searching the geospatial data, the values of the spatial accessibility functions $f_{A(x,y)}$ for scenic spots are calculated based on point A . According to SSTTA, search for a total k number of scenic spots to be visited in the clusters.

(3) The tourists confirm the route structure chain, the system collects the adjacent sections and control points of the structure chain. Generate directed weighted-edge graphs for each adjacent section based on the set of control points, build ASPSTs, and output the feasible path vectors based on the constructed optimal route algorithm to determine the optimal path for each adjacent section. Based on the transportation mode selected by tourists, the route decision schemes for tourists are finally confirmed.

4.2. Scenic Spot Clustering Result

Take the tourism city Chengdu as an example. The collected scenic spots are:

- $x_{(1)}$: Kuanzhai Alley;
- $x_{(2)}$: Chunxi Road;
- $x_{(3)}$: Giant Panda Breeding Base;
- $x_{(4)}$: Qinglong Lake;
- $x_{(5)}$: Temple of Marquis;
- $x_{(6)}$: Happy Valley;
- $x_{(7)}$: Shengxian Lake Park;
- $x_{(8)}$: Jinsha Site;
- $x_{(9)}$: Wangjianglou Park;
- $x_{(10)}$: East Lake Park;
- $x_{(11)}$: Du Fu Cottage;
- $x_{(12)}$: Chengdu Zoo;
- $x_{(13)}$: GuoSeTianXiang Paradise;
- $x_{(14)}$: Eastern Suburb Memory;
- $x_{(15)}$: TaZiShan Park;
- $x_{(16)}$: Jincheng Lake;
- $x_{(17)}$: Sichuan Museum;
- $x_{(18)}$: Chengdu Yongling Museum;
- $x_{(19)}$: The People's Park;
- $x_{(20)}$: South Lake Amusement Park.

The feature text vector is:

- $\mathbf{t}_{(1)}$: Walking, flower appreciation, cruise, leisure, and drinking tea;
- $\mathbf{t}_{(2)}$: The knowledge, the history, the humanity, investigation, and culture;
- $\mathbf{t}_{(3)}$: Theme, entertainment, games, sports, and outdoor.

The feature text matrix is constructed with the feature text vector as the basic element. Collect the text evaluation data of each scenic spot $x_{(i)}$ from Ctrip, Tong Cheng, Qunar, and other tourism booking portals and gather them in the text dataset. Count the total number of vector $\mathbf{t}_{(i)}$ word frequencies for each scenic spot and calculate the scenic spot clustering relative objective function value $\Delta f_{(x)}^i$, shown in Table 1. From the values, the clustering

tendency vector of each scenic spot $x_{(i)}$ is output and the cluster for each scenic spot is obtained, as shown in Table 2.

Table 1. Result of scenic spot clustering relative objective function value $\Delta f_{(x)}^i$.

	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
$\mathbf{t}_{(1)}$	0.657	0.659	0.255	0.701	0.177	0.305	0.660	0.066	0.581	0.620
$\mathbf{t}_{(2)}$	0.234	0.244	0.653	0.088	0.739	0.076	0.093	0.840	0.329	0.109
$\mathbf{t}_{(3)}$	0.109	0.097	0.092	0.211	0.084	0.618	0.246	0.093	0.090	0.272
	$x_{(11)}$	$x_{(12)}$	$x_{(13)}$	$x_{(14)}$	$x_{(15)}$	$x_{(16)}$	$x_{(17)}$	$x_{(18)}$	$x_{(19)}$	$x_{(20)}$
$\mathbf{t}_{(1)}$	0.101	0.178	0.315	0.259	0.493	0.676	0.094	0.120	0.537	0.284
$\mathbf{t}_{(2)}$	0.810	0.624	0.107	0.393	0.205	0.116	0.825	0.815	0.185	0.139
$\mathbf{t}_{(3)}$	0.089	0.198	0.578	0.347	0.302	0.208	0.081	0.065	0.278	0.577

Table 2. Clustering result for scenic spot $x_{(i)}$.

$S_{(i)}$	Cluster of Scenic Spots $s_{(i,j)}$	Cluster Capacity $n_{(i)}$
$S_{(1)} \sim \mathbf{t}_{(1)}$	$s_{(1,1)} \sim x_{(1)}$: Kuanzhai Alley; $s_{(1,2)} \sim x_{(2)}$: Chunxi Road; $s_{(1,3)} \sim x_{(4)}$: Qinglong Lake; $s_{(1,4)} \sim x_{(7)}$: Shengxian Lake Park; $s_{(1,5)} \sim x_{(9)}$: Wangjianglou Park; $s_{(1,6)} \sim x_{(10)}$: East Lake Park; $s_{(1,7)} \sim x_{(15)}$: TaZiShan Park; $s_{(1,8)} \sim x_{(16)}$: Jincheng Lake; $s_{(1,9)} \sim x_{(19)}$: The People's Park.	$n_{(1)} = 9$
$S_{(2)} \sim \mathbf{t}_{(2)}$	$s_{(2,1)} \sim x_{(3)}$: Giant Panda Breeding Base; $s_{(2,2)} \sim x_{(5)}$: Temple of Marquis; $s_{(2,3)} \sim x_{(8)}$: Jinsha Site; $s_{(2,4)} \sim x_{(11)}$: Du Fu Cottage; $s_{(2,5)} \sim x_{(12)}$: Chengdu Zoo; $s_{(2,6)} \sim x_{(14)}$: Eastern Suburb Memory; $s_{(2,7)} \sim x_{(17)}$: Sichuan Museum; $s_{(2,8)} \sim x_{(18)}$: Chengdu Yongling Museum.	$n_{(2)} = 8$
$S_{(3)} \sim \mathbf{t}_{(3)}$	$s_{(3,1)} \sim x_{(6)}$: Happy Valley; $s_{(3,2)} \sim x_{(13)}$: GuoSeTianXiang Paradise; $s_{(3,3)} \sim x_{(20)}$: South Lake Amusement Park.	$n_{(2)} = 3$

4.3. Recommendation Result of the Dynamic Buffer Optimal Scenic Spots

Condition 1: Tourists provide scores for the interest degree of each scenic spot text label, from which the system obtains the tourist feature evaluation matrix \mathbf{S}_T^{user} and then calculates the cluster scoring function values $f(\mathbf{S}_T^{user})^i$ to obtain the tourists' interest tendency for each cluster. Based on the feature measurement matrix $\mathbf{S}_T^{s(i,j)}$ of each scenic spot $x_{(i)}$, the scenic spot feature measurement function is used to calculate the measurement function values $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ of each cluster of scenic spots.

Condition 2: Taking the starting point A selected by the tourists as the center, the spatial distance of each scenic spot $x_{(i)}$ is searched and the values of the spatial accessibility functions $f_{A(A,s(i,j))}$ for the scenic spots are calculated. In the experiment, suppose that the starting point A chosen by tourists is Tianfu Square, and the expected scenic spot quantity is $k = 5$.

The calculated results of the label quantitative scoring values for each scenic spot feature text and tourist feature evaluation function values corresponding to clusters are shown in Table 3. Table 4 presents the feature measurement function values $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ of the scenic spots' cluster and the spatial accessibility function values $f_{A(A,s(i,j))}$ with starting point A . In Table 4, value (1) is the function value $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ and value (2) is the function value $f_{A(A,s(i,j))}$.

Table 3. Calculated results of the label quantitative scoring values for each scenic spot feature text and tourist feature evaluation function values corresponding to clusters.

$S_{(i)}$	Label Quantitative Scoring Values for Each Scenic Spot Feature Text					Tourist Feature Evaluation Function Values
$S_{(1)} \sim \mathbf{t}_{(1)}$	walking	flower appreciation	cruise	leisure	drinking tea	0.582
	0.86	0.12	0.15	0.9	0.88	
$S_{(2)} \sim \mathbf{t}_{(2)}$	knowledge	history	humanity	investigation	culture	0.894
	0.96	0.92	0.90	0.81	0.88	
$S_{(3)} \sim \mathbf{t}_{(3)}$	theme	entertainment	games	sports	outdoor	0.324
	0.40	0.32	0.20	0.20	0.50	

Table 4. Feature measurement function values $f(\mathbf{S}_T^{user}, \mathbf{S}_T^{s(i,j)})^i$ of the scenic spots' cluster and spatial accessibility function values $f_{A(A,s(i,j))}$.

$S_{(1)} \sim \mathbf{t}_{(1)}$	$x_{(1)}$	$x_{(2)}$	$x_{(4)}$	$x_{(7)}$	$x_{(9)}$	$x_{(10)}$	$x_{(15)}$	$x_{(16)}$	$x_{(19)}$
Value (1)	6.3372	4.1922	1.0138	1.5199	0.9868	0.9839	0.9868	1.0997	0.9868
Value (2)	0.0101	0.0116	0.0891	0.0428	0.0297	0.0362	0.0442	0.0703	0.0059
$S_{(2)} \sim \mathbf{t}_{(2)}$	$x_{(3)}$	$x_{(5)}$	$x_{(8)}$	$x_{(11)}$	$x_{(12)}$	$x_{(14)}$	$x_{(17)}$	$x_{(18)}$	
Value (1)	0.9541	8.9443	7.7850	8.9443	0.8942	1.2000	8.9443	8.9443	
Value (2)	0.0841	0.0152	0.0413	0.0261	0.0507	0.0406	0.0225	0.0188	
$S_{(3)} \sim \mathbf{t}_{(3)}$	$x_{(6)}$	$x_{(13)}$	$x_{(20)}$						
Value (1)	0.8344	0.8344	0.8344						
Value (2)	0.0565	0.1768	0.1275						

Based on the constructed DBSA, the interested clusters for the sample tourist are $S_{(2)}$, $S_{(1)}$, and $S_{(3)}$. The output scenic spots that satisfy the sample tourist's interests and have the optimal spatial accessibility are:

- $s_{(2,2)}$: Wuhou Temple;
- $s_{(2,8)}$: Chengdu Yongling Museum;
- $s_{(1,1)}$: Kuan Zhai Alley;
- $s_{(1,2)}$: Chunxi Road;
- $s_{(2,7)}$: Sichuan Museum.

4.4. Result of the Optimal Tour Routes and Spatial Decision-Making

Based on the recommended five scenic spots, suppose that the sample tourist chooses one of the following sequences: A : Tianfu Square, $s_{(2,2)}$: the Temple of Marquis, $s_{(2,8)}$: Chengdu Yongling Museum, $s_{(1,1)}$: Kuan Zhai Alley, $s_{(1,2)}$: Chunxi Road, and $s_{(2,7)}$: Sichuan Museum. The starting point of the tour is the Tianfu Square. According to the traveling sequence selected by the tourist, the system determines five adjacent sections and determines the set of control points within each section according to the geographical relationship of the scenic spots and the distributions of urban roads. Then, the system outputs the directed weighted-edge graphs and ASPSTs. Figure 7 shows the location distribution, path structure, generated ASPSTs and the optimal path for each adjacent section.

(1) Figure 7(a-1)–(a-3) represent the distribution of starting point and control points, the path among points, the ASPST of $T_{(A,s(2,2))}$, and the optimal path $P_{ath(c)}$ in section $R_{(A,s(2,2))}$;

(2) Figure 7(b-1)–(b-3) represent the distribution of starting point and control points, the path among points, the ASPST of $T_{(s(2,2),s(2,8))}$, and the optimal path $P_{ath(c)}$ in section $R_{(s(2,2),s(2,8))}$;

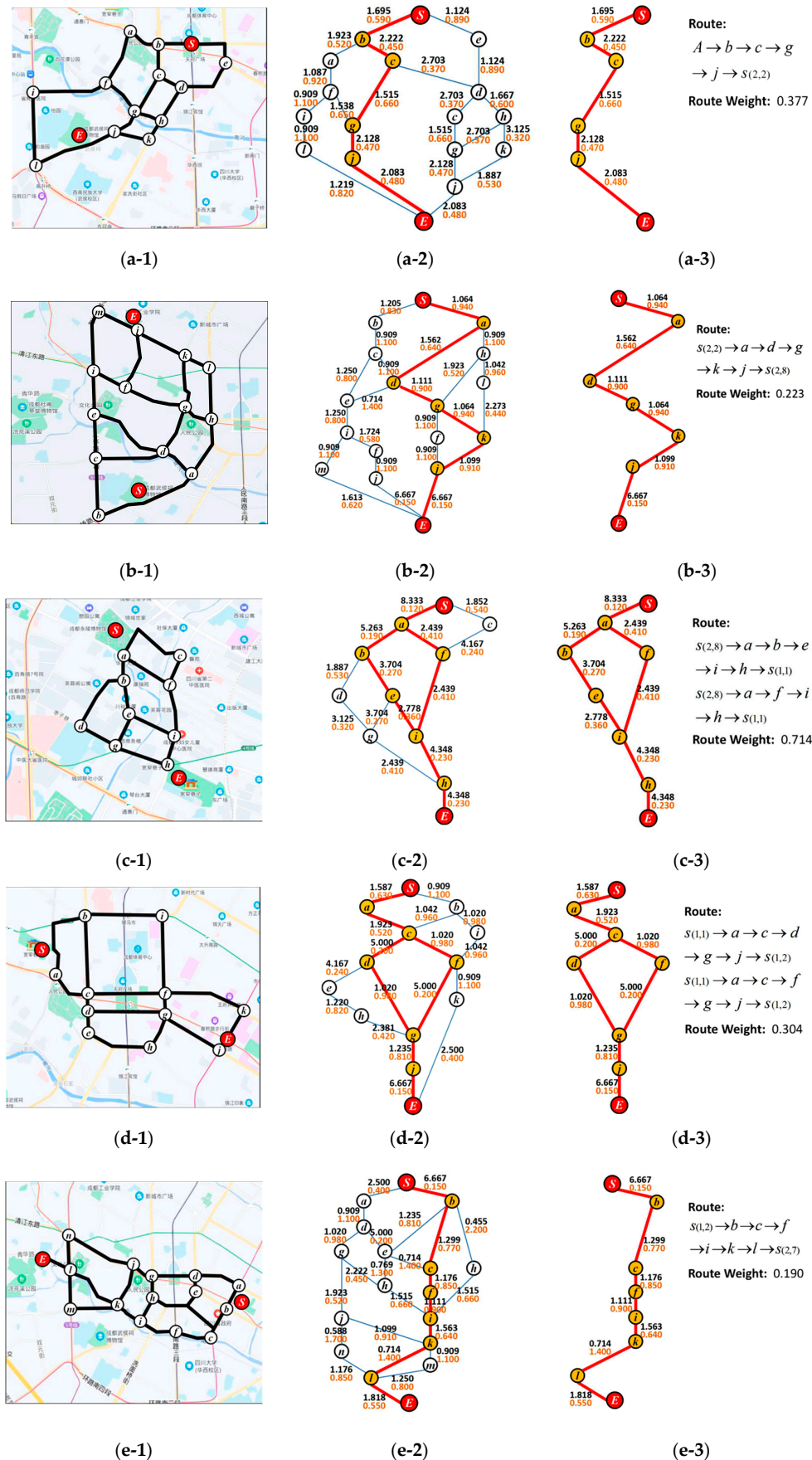


Figure 7. Location distribution, path structure, ASPST, and optimal path for adjacent section.

(3) Figure 7(c-1)–(c-3) represent the distribution of starting point and control points, the path among points, the ASPST of $T_{(s(2,8),s(1,1))}$, and the optimal path $P_{ath(c)}$ in section $R_{(s(2,8),s(1,1))}$;

(4) Figure 7(d-1)–(d-3) represent the distribution of starting point and control points, the path among points, the ASPST of $T_{(s(1,1),s(1,2))}$, and the optimal path $P_{ath(c)}$ in section $R_{(s(1,1),s(1,2))}$;

(5) Figure 7(e-1)–(e-3) represent the distribution of starting point and control points, the path among points, the ASPST of $T_{(s(1,2),s(2,7))}$, and the optimal path $P_{ath(c)}$ in section $R_{(s(1,2),s(2,7))}$.

The control points in each ASPST are represented by letters a, b, c, ..., etc. In each section, the initial point is A: Tianfu Square or scenic spot $G_{(i,x)}$, the terminal point is scenic spot $G_{(i,y)}$, and the two scenic spots are determined by the section $R_{(G_{(i,x)},G_{(i,y)})}$ point mark. In each structure tree, the path drawn in red is the searched optimal path. Extract the optimal path to obtain the path model in each section, mark the path's two endpoints and control points, and calculate the path weight for each ASPST. The optimal path weights are:

- Section $R_{(A,s(2,2))}$: 0.377;
- Section $R_{(s(2,2),s(2,8))}$: 0.223;
- Section $R_{(s(2,8),s(1,1))}$: 0.714;
- Section $R_{(s(1,1),s(1,2))}$: 0.304;
- Section $R_{(s(1,2),s(2,7))}$: 0.190.

According to the ASPSTs and the optimal paths, the system makes the spatial decision for the tourist. Analyze the travel schedule as follows.

- Start from A: Tianfu Square;
- Pass through the first section $R_{(A,s(2,2))}$;
- Arrive at the scenic spot $s_{(2,2)} \sim x_{(5)}$: Wuhou Temple, pay a visit;
- Pass through the second section $R_{(s(2,2),s(2,8))}$;
- Arrive at $s_{(2,8)} \sim x_{(18)}$: Chengdu Yongling Museum, pay a visit;
- Pass through the third section $R_{(s(2,8),s(1,1))}$;
- Arrive at $s_{(1,1)} \sim x_{(1)}$: Kuanzhai Alley, pay a visit;
- Pass through the fourth section $R_{(s(1,1),s(1,2))}$;
- Arrive at $s_{(1,2)} \sim x_{(2)}$: Chunxi Road, pay a visit;
- Pass through the fifth section $R_{(s(1,2),s(2,7))}$;
- Arrive at $s_{(2,7)} \sim x_{(17)}$: Sichuan Museum, pay a visit. The tour ends.

In the whole tour, visiting the scenic spots will create a fixed cost, with fixed time and fee costs, but moving in each section will generate different costs depending on the transportation mode. According to the travel conditions, the system provides the following two modes:

- (1) Only consider the lowest time cost, and do not consider the fee cost. It recommends "taxi" as the transportation mode and provides the optimal path in each section;
- (2) Only consider the lowest fee cost, but do not consider the time cost. It recommends the "shared bicycle" as the transportation mode.

Table 5 shows tour routes I, II, and III composed of the optimal and sub-optimal paths in each section, as well as the path total weight for each section and tour route total weight. In each section, letters are used to express the control points to form the path. The endpoints of section $R_{(G_{(i,x)},G_{(i,y)})}$ are the two points $G_{(i,x)}$ and $G_{(i,y)}$. We obtain the following conclusion: of the three tour routes, tour route I is the optimal route, while tour route II and tour route III are the sub-optimal routes.

The travel time comparison of the optimal tour route and sub-optimal tour routes under the first mode in Table 6 and the fee cost comparison of the optimal tour route and sub-optimal tour routes under the second mode in Table 7 were obtained. The visiting time of each scenic spot is the optimal tour time recommended by the tourism information

platform, and the visiting cost of each scenic spot is the minimum cost (i.e., ticket for scenic spot entrance). Taxis and shared bicycle modes are assessed based on the average speed and average price on the urban roads of Chengdu.

Table 5. Tour routes I, II, and III composed of the optimal and sub-optimal paths in each section, as well as the path total weight for each section and tour route total weight.

Section	$R_{(A,s(2,2))}$	$R_{(s(2,2),s(2,8))}$	$R_{(s(2,8),s(1,1))}$	$R_{(s(1,1),s(1,2))}$	$R_{(s(1,2),s(2,7))}$	E_{route}
Route I	b, c, g, j	a, d, g, k, j	a, b, e, i, h a, f, i, h	a, c, d, g, j a, c, f, g, j	b, c, f, i, k, l	1.808
$E(Path_{(c)})$	0.377	0.223	0.714	0.304	0.190	
Route II	b, c, d, h, g, j	a, h, l, k, j	a, b, e, g, h	a, c, f, k	b, e, h, i, k, l	1.649
$E(Path_{(c)})$	0.300	0.222	0.671	0.275	0.181	
Route III	b, c, d, h, k, j	a, h, g, k, j	c, f, i, h	a, c, d, e, h, g, j	b, h, i, k, l	1.567
$E(Path_{(c)})$	0.299	0.219	0.606	0.264	0.179	

Table 6. Time cost comparison of the optimal tour route and sub-optimal tour routes under the first mode (unit: hours).

	$R_{(A,s(2,2))}$	$s_{(2,2)}$	$R_{(s(2,2),s(2,8))}$	$s_{(2,8)}$	$R_{(s(2,8),s(1,1))}$	$s_{(1,1)}$
Route I	0.134	2.0	0.226	2.0	0.071	2.0
Route II	0.167	2.0	0.227	2.0	0.075	2.0
Route III	0.169	2.0	0.231	2.0	0.083	2.0
	$R_{(s(1,1),s(1,2))}$	$s_{(1,2)}$	$R_{(s(1,2),s(2,7))}$	$s_{(2,7)}$	Total time (hours)	Total fee (yuan)
Route I	0.166	2.0	0.266	2.0	10.863	124.61
Route II	0.184	2.0	0.279	2.0	10.932	127.1
Route III	0.191	2.0	0.282	2.0	10.956	127.64

Table 7. Fee cost comparison of the optimal tour route and sub-optimal tour routes under the second mode (unit: yuan).

	$R_{(A,s(2,2))}$	$s_{(2,2)}$	$R_{(s(2,2),s(2,8))}$	$s_{(2,8)}$	$R_{(s(2,8),s(1,1))}$	$s_{(1,1)}$
Route I	1.5	50	3.0	20	1.5	0
Route II	3.0	50	3.0	20	1.5	0
Route III	3.0	50	3.0	20	1.5	0
	$R_{(s(1,1),s(1,2))}$	$s_{(1,2)}$	$R_{(s(1,2),s(2,7))}$	$s_{(2,7)}$	Total time (hours)	Total fee (yuan)
Route I	1.5	0	3.0	0	11.422	80.5
Route II	3.0	0	3.0	0	11.545	83.5
Route III	3.0	0	3.0	0	11.580	83.5

4.5. Comparison with Other Methods

We selected BaiDu Map (BDMA) and GaoDe Map (GDMA), which are commonly used in tour route planning, as the control group. The proposed algorithm (TRRA) was set as the experimental group. Under the constraint of the shortest travel time, search for the optimal route between the scenic spots by BDMA and GDMA. Based on the searched optimal route by BDMA and GDMA, confirm the node nodes on the route and generate related ASPSTs. Calculate the weight $E(Path_{(c)})$ of each road section and the total weight E_{route} of the route by obtaining the distance between the road nodes. The comparison results are shown in Table 8, where $t_m = 1$ represents the transportation mode “Taxi” and $t_m = 2$ represents the transportation mode “shared bicycle”.

Table 8. Comparison of the $E(Path_{(c)})$ and the E_{route} of the three methods.

$t_m = 1$	$E(Path_{(c)})$					E_{route}
	$R_{(A,s(2,2))}$	$R_{(s(2,2),s(2,8))}$	$R_{(s(2,8),s(1,1))}$	$R_{(s(1,1),s(1,2))}$	$R_{(s(1,2),s(2,7))}$	
BDMA	0.085	0.248	0.188	0.118	0.090	0.729
GDMA	0.084	0.330	0.096	0.130	0.117	0.757
TRRA	0.377	0.223	0.714	0.304	0.190	1.808

$t_m = 2$	$E(Path_{(c)})$					E_{route}
	$R_{(A,s(2,2))}$	$R_{(s(2,2),s(2,8))}$	$R_{(s(2,8),s(1,1))}$	$R_{(s(1,1),s(1,2))}$	$R_{(s(1,2),s(2,7))}$	
BDMA	0.092	0.237	0.076	0.114	0.163	0.682
GDMA	0.115	0.326	0.131	0.132	0.161	0.865
TRRA	0.377	0.223	0.714	0.304	0.190	1.808

4.6. Experimental Result Analysis

The constructed SSCA was used to cluster the representative scenic spots of the tourism city Chengdu and recommend scenic spots to the sample tourist based on their interests. According to the tourism mode selected by the tourist, tour routes are searched for the tourist based on the recommended scenic spots to meet the condition of an optimal traveling time or an optimal traveling cost for the tourist. The experimental results were analyzed based on the clustering results, SSR results, TRR results, and method comparison.

4.6.1. Analysis of the SSCA Results

The feature text vector collected in the experiment met the modeling conditions, and the selected feature text represented one kind of scenic spot clustering. It had obvious differentiation with the other clusters. By mining the evaluation text, the clustering relative objective function value of each scenic spot is calculated, and then each scenic spot's tendency for the feature text clusters is obtained. Analyzing the calculation results of the clustering relative objective function values in Table 1 and the clustering results of scenic spots in Table 2, each scenic spot has an apparent tendency for one cluster feature text vector, and the tendency for its own cluster is the highest.

(1) Scenic spots $s_{(1,1)}$: Kuanzhai Alley, $s_{(1,2)}$: Chunxi Road, $s_{(1,3)}$: Qinglong Lake, $s_{(1,4)}$: Shengxian Lake Park, $s_{(1,5)}$: Wangjianglou Park, $s_{(1,6)}$: East Lake Park, $s_{(1,7)}$: Tazishan Park, $s_{(1,8)}$: Jincheng Lake, and $s_{(1,9)}$: People's Park have the highest tendencies for cluster $S_{(1)} \sim \mathbf{t}_{(1)}$, with values of $s_{(1,1)}$: 0.657, $s_{(1,2)}$: 0.659, $s_{(1,3)}$: 0.701, $s_{(1,4)}$: 0.660, $s_{(1,5)}$: 0.581, $s_{(1,6)}$: 0.620, $s_{(1,7)}$: 0.493, $s_{(1,8)}$: 0.676, and $s_{(1,9)}$: 0.537, respectively.

(2) Scenic spots $s_{(2,1)}$: The Giant Panda Breeding Base, $s_{(2,2)}$: Wuhou Temple, $s_{(2,3)}$: The Jinsha Site, $s_{(2,4)}$: Du Fu Thatched Cottage, $s_{(2,5)}$: Chengdu Zoo, $s_{(2,6)}$: Eastern Suburb Memory, $s_{(2,7)}$: Sichuan Museum, and $s_{(2,8)}$: Chengdu Yongling Museum have the highest tendencies for cluster $S_{(2)} \sim \mathbf{t}_{(2)}$, which are $s_{(2,1)}$: 0.653, $s_{(2,2)}$: 0.739, $s_{(2,3)}$: 0.840, $s_{(2,4)}$: 0.810, $s_{(2,5)}$: 0.624, $s_{(2,6)}$: 0.393, $s_{(2,7)}$: 0.825, and $s_{(2,8)}$: 0.815 respectively.

(3) Scenic spots $s_{(3,1)}$: Happy Valley, $s_{(3,2)}$: Guose Tianxiang Paradise, and $s_{(3,3)}$: South Lake Amusement Park have the highest tendencies for cluster $S_{(3)} \sim \mathbf{t}_{(3)}$, which are $s_{(3,1)}$: 0.618, $s_{(3,2)}$: 0.578, and $s_{(3,3)}$: 0.577, respectively.

According to the clustering results, the capacity of cluster $S_{(1)} \sim \mathbf{t}_{(1)}$ was $n_{(1)} = 9$, the capacity of cluster $S_{(2)} \sim \mathbf{t}_{(2)}$ was $n_{(2)} = 8$, and the capacity of cluster $S_{(3)} \sim \mathbf{t}_{(3)}$ was $n_{(3)} = 3$. We obtained the following conclusions:

(1) The clustering results indicate that our proposed algorithm can extract the feature attributes of scenic spots based on the tourist text evaluation data and can accurately cluster the scenic spots according to the feature attributes;

(2) The modeling process quantifies the tourists' interests and matches them with the feature attributes of the scenic spots so that the scenic spots that meet a certain type of tourists' interests can be grouped into one cluster;

(3) When the system recommends scenic spots, it takes the cluster as a basic unit to ensure that the recommended scenic spots are more accurate.

4.6.2. Analysis of the SSR Results

1. Results Analysis of Table 3

Analyzing the quantitative value of the sample tourist's label scoring on the scenic spot feature text in Table 3, we can conclude the following:

(1) The sample tourist provided different scores for the corresponding feature text labels for each cluster, indicating that the tourist had different interest degrees for each label;

(2) By calculating the tourist feature evaluation function values, the tourist's interest tendency for each cluster is obtained. The evaluation function value of the cluster $S_{(2)} \sim t_{(2)}$ is the highest, with a value of 0.894, followed by cluster $S_{(1)} \sim t_{(1)}$, with a value of 0.582, and finally cluster $S_{(3)} \sim t_{(3)}$, with a value of 0.324;

(3) The sample tourist had the highest degree of the interest in cluster $S_{(2)} \sim t_{(2)}$ scenic spots, followed by cluster $S_{(1)} \sim t_{(1)}$ scenic spots, and had the lowest degree of interest in cluster $S_{(3)} \sim t_{(3)}$ scenic spots;

(4) It can be inferred that the tourist preferred parks and green land, leisure landscapes, museums, memorial halls, and similar scenic spots, but was not interested in amusement parks and theme parks.

2. Results Analysis of Table 4

Analyzing the feature measurement function values of the scenic spots in the clusters in Table 4 and the spatial accessibility function value of starting point A: Tianfu Square, we can conclude the following:

(1) The feature measurement function value for each scenic spot was different. The larger the feature measurement function value is, the higher the matching degree between the scenic spot and the tourist interests will be. The smaller the spatial accessibility function value is, the stronger the accessibility of scenic spot will be, and vice versa;

(2) The scenic spots that best matched the sample tourist's interests were:

- $x_{(5)}$: Wuhou Temple;
- $x_{(11)}$: Du Fu Thatched Cottage;
- $x_{(17)}$: Sichuan Museum;
- $x_{(18)}$: Chengdu Yongling Museum.

(3) The most accessible scenic spot was $x_{(19)}$: The People's Park. The system recommends scenic spots by confirming the interest matching degree and the spatial accessibility at the same time. Combined with the conditions proposed by the tourist, the final scenic spots recommended were:

- $s_{(2,2)}$: Wuhou Temple;
- $s_{(2,8)}$: Chengdu Yongling Museum;
- $s_{(1,1)}$: Kuan Zhai Alley;
- $s_{(1,2)}$: Chunxi Road;
- $s_{(2,7)}$: Sichuan Museum.

Based on the comprehensive analysis, the recommended five scenic spots had the features of high feature measurement function values and strong spatial accessibility. The results verify that the algorithm can determine the scenic spots that best match tourists' interests and have optimal space accessibility.

4.6.3. Analysis of the TRR Results

1. Results Analysis of Figure 7

According to the constructed tour route searching algorithm, the tour route was divided into five adjacent sections based on the constraints provided by the sample tourist.

Figure 7 shows the point distribution of each adjacent section, path structure, ASPST, and optimal path for each adjacent section output by the algorithm. The secondary icon letters a, b, c, d, and e in the figure represent adjacent sections. We can conclude the following:

(1) Figure 7(a-1,b-1,c-1,d-1,e-1) show the distribution of the section endpoints and the control points. The output results indicate that the control points cover all the key road nodes in the section and involve all roads;

(2) Figure 7(a-2,b-2,c-2,d-2,e-2) show the section of ASPST. The red circles in the figure represent the starting point and terminal point of the section, the white circles represent the control point, and the blue connecting lines represent the paths between points. Regarding the values on the lines, the black values represent the calculated section edge weights between two points and the brown values represent the distance of the path between two points on the map. The red lines represent the path with the shortest searching distance and the largest total route weight, corresponding to the optimal path between the starting point and the terminal point. The route will cause the lowest costs when the tourist travels along it;

(3) Figure 7(a-3,b-3,c-3,d-3,e-3) show the optimal path extracted from ASPST, and the corresponding path, nodes, and total path weight;

(4) The experimental results indicate that the algorithm covers all paths and control points in the sections, and can determine the optimal path.

2. Results Analysis of Table 5

Analyzing the output results of the tour routes in Table 5, we obtain the following conclusions:

(1) Under the condition of the selected travel sequence, tour route I is the optimal route searched by the algorithm, and tour route II and tour route III are the sub-optimal routes searched by the algorithm;

(2) In each route, the path total weight in any section $R_{(G(i,x),G(i,y))}$ of route I is larger than that of route II and route III, and the final result is that the total weight of route I is larger than that of route II and route III;

(3) The total weights of route I, route II, and route III were 1.808, 1.649, and 1.567, respectively. This indicates that the cost caused by the sample tourist when choosing route I was the lowest, followed by the route II and route III. The total weights of other routes in the ASPST were larger than those of route I, route II, and route III in the same section. If tourists choose other routes, it will cause higher costs. Therefore, route I is recommended by the system, followed by route II and route III.

3. Result Analysis of Table 6

Analyzing the data in Table 6, we obtain the following conclusions:

(1) Under the condition of the route recommendation results, the tourist choosing different tourism modes will generate different costs;

(2) When the tourist chooses the first mode and only considers the time cost while neglecting the fee cost, the traveling time cost generated by route I is 10.863 h, the traveling time cost generated by route II is 10.932 h, and the traveling time cost generated by route III is 10.956 h;

(3) Route II and route III cost 0.069 h and 0.093 h more than route I, respectively. At the same time, the fee cost of route I is 124.61 yuan, the fee cost of route II is 127.10 yuan, and the fee cost of route III is 127.64 yuan. The fee costs of route II and route III are 2.49 yuan and 3.03 yuan more than that of route I, respectively.

4. Result Analysis of Table 7

Analyzing the data in Table 7, we obtained the following conclusions:

(1) When the tourist chooses the second mode, only considering the fee cost while neglecting the time cost, the cost of the traveling fee generated by route I is 80.5 yuan, the fee cost of route II is 83.5 yuan, and the fee cost of route III is 83.5 yuan. The costs of route II and route III are 3.5 yuan and 3.5 yuan more than that of route I, respectively;

(2) The traveling time cost of route I is 11.422 h, the traveling time cost of route II is 11.545 h, and the traveling time cost of route III is 11.580 h. Route II and route III cost 0.123 h and 0.158 h more than route I, respectively;

(3) The optimal route is the one that causes the lowest time cost or the lowest fee cost, and the algorithm constructed in this study can recommend the optimal route to the tourists.

4.6.4. Analysis of the Algorithm Effectiveness and Performance

It can be concluded from the above analysis that, no matter which mode the sample tourist chooses, the time cost and fee cost of optimal tour route I are the lowest. The system recommends route I as the optimal tour route to the tourist for a tourism day. Both route II and route III generate more travel costs than route I, and the travel costs of other feasible routes are much higher. When the number of scenic spots and tourist days increases, the other routes will generate more time costs and fee costs than the optimal one.

(1) The proposed algorithm had good clustering ability. The scenic spot clustering was based on data mining of tourists' text evaluations and the scoring on the feature text by the experimental sample tourist. The tourists' interests were matched with the feature attributes of the scenic spots, which made the scenic spot clustering highly targeted and closely related to the tourists' interest tendencies, finally improving the recommendation algorithm accuracy;

(2) The algorithm has the ability to accurately recommend scenic spots. The interest tendency is determined by the tourists' evaluation scores, and is matched with the scenic spots in the cluster to obtain their abilities to meet the tourists' interests in the aspect of feature attributes. The spatial distributions of the scenic spots that satisfy the interests are optimized by combining with spatial accessibility measurement. Then, the scenic spots satisfying the tourists' interests and having the optimal spatial accessibility are obtained by the searching algorithm and recommended to tourists. The experiment proves that the algorithm constructed in this paper can recommend the best scenic spots for tourists, not only conforming to the tourists' interests, but also being optimal in space accessibility, saving traveling costs;

(3) The algorithm can determine the optimal tour route that meets the needs of tourists under the constraint conditions, so that the time cost and fee cost of the tour route are both the lowest. The proposed tour route algorithm is based on the tourists' interests and needs, the selected traveling modes, and the geospatial data. It can progressively determine the path with the lowest cost in each road adjacent section; thus, it has the characteristics of global searching. When the number of scenic spots and the traveling days increases, tourists will save much more traveling costs by choosing the optimal route, which will save the total energy consumed in the tour and help to achieve low-carbon and green tourism.

4.6.5. Analysis of the Comparison of Three Methods

Analyzing the results in Table 8, compared with BDMA and GDMA, when the transportation modes were "taxi" and "shared bicycle", the TRRA had higher section weight $E(Path_{(c)})$ in $R_{(A,s(2,2))}$, $R_{(s(2,8),s(1,1))}$, $R_{(s(1,1),s(1,2))}$, and $R_{(s(1,2),s(2,7))}$. Only in $R_{(s(2,2),s(2,8))}$ was the weight $E(Path_{(c)})$ lower than BDMA and GDMA, indicating that the overall performance of the TRRA was stable and it could determine paths with lower costs than BDMA and GDMA in total.

From the perspective of algorithm logic, the TRRA adopts a greedy method when searching for the ASPST, traversing all road nodes in the ASPST, so it can always determine the ASPST with the lowest cost and obtain the global solution. BDMA and GDMA search for urban main road nodes, which might overlook the accessible secondary roads and certain road nodes; thus, the solution might not be the optimal one.

Regarding the optimal route weights E_{route} output by the three methods, when $t_m = 1$, the TRRA was 1.079 higher than BDMA and 1.051 higher than GDMA; when $t_m = 2$, the

TRRA was 1.126 higher than BDMA and 0.943 higher than GDMA. It can be concluded that the TRRA saves more travel costs than BDMA and GDMA.

4.6.6. Superiority of the Proposed Algorithm

Through analysis, it can be concluded that the proposed algorithm has significant advantages over the methods used in the literature [5–12] and the commonly used electronic map methods used in tour route planning, as follows:

- (1) The algorithm can achieve precise clustering of scenic spots within the research scope based on feature attributes. It can group the scenic spots that meet approximate interests and needs into one category cluster. And it has high clustering accuracy;
- (2) It can effectively capture the tourist interests and recommend scenic spots with interest data that best match their feature attributes, achieving accurate recommendations;
- (3) It can determine the globally optimal route within the research scope, that is, the route with the lowest travel cost, to avoid outputting an approximate optimal solution;
- (4) It overcomes the problems in electronic map route planning and can determine the optimal route from the perspective of meeting tourist demands and minimizing travel costs. Compared with the routes planned by electronic maps, it can save more travel costs.

5. Conclusions and Prospect

5.1. Conclusions of This Research

Based on the analysis of the existing problems for TRR, we constructed a tour route-recommendation algorithm based on feature text mining (FTM) and the spatial accessibility model (SAM). In this research, we introduce tourists' text evaluation data mining and SA measurement into scenic spot recommendation (SSR). First, a scenic spot clustering algorithm (SSCA) based on FTM is constructed. The tendency of a scenic spot being grouped into an unknown cluster is obtained by evaluating text mining for tourists in each scenic spot, and then the category of each scenic spot is obtained based on tourists' interest text mining. Regarding the SSCA, a scenic spot topology tree algorithm (SSTTA) based on dynamic buffer spatial accessibility (DBSA) was constructed, through which the scenic spot spatial distribution is obtained; then, the optimal scenic spots with the best-matched interests and spatial distributions are recommended to the tourists. Based on the recommended scenic spots, tourists confirm the travel sequence and transportation mode, and an optimal tour route recommendation algorithm (OTRRA) based on SSTTA was constructed to recommend optimal tour routes for tourists. The experiment showed that the proposed algorithm had good clustering ability and could obtain scenic spot clusters based on tourists' interests. The algorithm has the ability to accurately recommend scenic spots, which relies on the spatial accessibility and ability to match tourists' interests, resulting in high recommendation accuracy. At the same time, the algorithm can determine the optimal tour route that meets tourists' needs under constraint conditions, making the travel time and travel costs the lowest.

5.2. Application Scenarios

The application areas of the proposed algorithm mainly include several aspects.

Firstly, it can be used as an embedded algorithm to build scenic spot recommendation systems and route-recommendation systems, and develop handheld APPs or PC application systems. It can provide process methods for developing urban smart tourism information service platforms and also provide technical support for improving the quality of tourism city services.

Secondly, it can be used for the design and development of intelligent tourism electronic maps. For urban tourism, the development of smart tourism electronic maps dedicated to urban scenic spots and route recommendations is relatively weak. The proposed algorithm can be directly used to develop smart tourism electronic maps for cities, providing exclusive services for tourists.

Thirdly, it can be deeply integrated with technologies such as the Internet of Things and the Internet of Vehicles. The constructed scenic spot recommendation algorithm and route-planning algorithm can be integrated into the intelligent connected vehicle navigation system as a necessary embedded algorithm for developing intelligent navigation vehicles in smart tourism.

5.3. Limitations and Improvement Direction

Based on the algorithm objectives and the modeling process, it can be concluded that the proposed algorithm has certain limitations, and corresponding improvements could be made in the next research step. For example, the algorithm is mainly applicable to the main urban areas of tourism cities, with transportation methods such as urban buses, taxis, and shared bicycles, which does not involve cross-city scenic spot recommendations and route planning. When expanding the research area to counties around cities or adjacent cities, it is necessary to further construct algorithms with a larger scope based on cross-city scenic spots, cross-city roads, and cross-city transportation modes. In addition, the proposed algorithm mainly generates the global optimal route for the selection of a certain fixed transportation mode and is not suitable for randomly changing transportation modes. In the next step of research, for different route intervals of random transportation modes chosen by tourists, randomly choosing methods of transportation could be considered to design the optimal route algorithm, which can improve the flexibility and application scope of the algorithm.

Author Contributions: Conceptualization, X.Z., Z.Z., X.L. and M.S.; methodology, X.Z. and Z.Z.; formal analysis, X.Z., X.L. and M.S.; writing—original draft preparation, X.Z. and Z.Z.; writing—review and editing, X.Z., Z.Z., X.L. and M.S.; funding acquisition, X.Z., Z.Z. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Project of Henan Provincial Natural Science Foundation Project (No. 242300420624), the Project of the Key Research Base of Region and Country of Sichuan Province, Center for Southeast Asian Economic and Culture Studies (No. DNY2301), the National 13th Five-Year Plan Key R&D Projects (No. 2016YFB0502300), the National 14th Five-Year Plan Key R&D Projects (No. SQ2021YFB3900025), and Leshan Science and Technology Plan Project (No. 22ZRKX025).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arif, A.; Du, J.T. Understanding collaborative tourism information searching to support online travel planning. *Online Inf. Rev.* **2019**, *43*, 369–386. [\[CrossRef\]](#)
2. Zhang, Q.; Liu, Y.; Liu, L.; Lu, S.; Feng, Y.; Yu, X. Location identification and personalized recommendation of tourist attractions based on image processing. *Trait. Du Signal* **2021**, *38*, 197–205. [\[CrossRef\]](#)
3. Sarkar, J.L.; Majumder, A. A new point-of-interest approach based on multi-itinerary recommendation engine. *Expert Syst. Appl.* **2021**, *181*, 115026. [\[CrossRef\]](#)
4. Ahmad, S.; Ullah, I.; Mehmood, F.; Fayaz, M.; Kim, D. Astochastic approach towards travel route optimization and recommendation based on users constraints using Markov chain. *IEEE Access* **2019**, *7*, 90760–90776. [\[CrossRef\]](#)
5. Yun, L.Y.; Jiao, H.H.; Lu, K. Tourist Attraction Recommendation Method Based on Megadata and Artificial Intelligence Algorithm. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 4461165. [\[CrossRef\]](#)
6. Lin, S.Q. Implementation of Personalized Scenic Spot Recommendation Algorithm Based on Generalized Regression Neural Network for 5G Smart Tourism System. *Comput. Intell. Neurosci.* **2022**, *2022*, 3704494. [\[CrossRef\]](#)
7. Cepeda-Pacheco, J.; Domingo, M. Deep learning and Internet of Things for tourist attraction recommendations in smart cities. *Neural Comput. Appl.* **2022**, *34*, 7691–7709. [\[CrossRef\]](#)
8. Han, S.S.; Liu, C.M.; Chen, K.Y.; Gui, D.W.; Du, Q.Y. A Tourist Attraction Recommendation Model Fusing Spatial, Temporal, and Visual Embeddings for Flickr-Geotagged Photos. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 20. [\[CrossRef\]](#)
9. Zheng, X.; You, H.; Huang, H.; Sun, L.; Yu, Q.; Luo, Y. Two-stage greedy algorithm based on crowd sensing for tour route recommendation. *Appl. Soft Comput.* **2024**, *153*, 111260. [\[CrossRef\]](#)

10. Lyu, H.J.; Han, Z.K. Algorithm of Scenic Spot Tourism Route Planning Based on Convolution Neural Network. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5330611. [\[CrossRef\]](#)
11. Niu, H. The effect of intelligent tour guide system based on attraction positioning and recommendation to improve the experience of tourists visiting scenic spots. *Intell. Syst. Appl.* **2023**, 200263. [\[CrossRef\]](#)
12. Li, S.; Luo, T.Y.; Wang, L.; Xing, L.N.; Ren, T. Tourism route optimization based on improved knowledge ant colony algorithm. *Complex Intell. Syst.* **2022**, *8*, 3973–3988. [\[CrossRef\]](#)
13. Yuan, W.J.; Yang, L.; Yang, Q.; Sheng, Y.H.; Wang, Z.Y. Extracting Spatio-Temporal Information from Chinese Archaeological Site Text. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 175. [\[CrossRef\]](#)
14. Eldeeb, G.; Mohamed, M. Transit electrification state of the art: A machine-learning based text mining approach. *Transp. Res. Part D Transp. Environ.* **2011**, *111*, 103446. [\[CrossRef\]](#)
15. Ogryzek, M.; Podawca, K.; Cienciała, A. Geospatial tools in the analyses of land use in the perspective of the accessibility of selected educational services in Poland. *Land Use Policy* **2022**, *122*, 106373. [\[CrossRef\]](#)
16. Zhou, H.L.; Tian, Z.P. Population mobility network and spatial accessibility based on peer-to-peer interactive energy management. *Sustain. Comput. Inform. Syst.* **2022**, *36*, 100800. [\[CrossRef\]](#)
17. Luo, S.J.; Jiang, H.P.; Yi, D.S.; Liu, R.H.; Qin, J.H.; Liu, Y.S.; Zhang, J. PM2SFCA: Spatial Access to Urban Parks, Based on Park Perceptions and Multi-Travel Modes. A Case Study in Beijing. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 488. [\[CrossRef\]](#)
18. Hou, J.; Yuan, H.Q.; Pelillo, M. Towards Parameter-Free Clustering for Real-World Data. *Pattern Recognit.* **2023**, *134*, 109062. [\[CrossRef\]](#)
19. Hu, X. Geographic landscape space and tourism planning of Longtan Ancient Village. *Bol. Tec./Tech. Bull.* **2017**, *55*, 627–632.
20. Tlili, T.; Krichen, S. A simulated annealing-based recommender system for solving the tourist trip design problem. *Expert Syst. Appl.* **2021**, *186*, 115723. [\[CrossRef\]](#)
21. Kolahkaj, M.; Harounabadi, A.; Nikravanshalmani, A.; Chinipardaz, R. DBCACF: A multidimensional method for tourist recommendation based on users' demographic, context and feedback. *Recomm. Syst.* **2019**, *4*, 209–219.
22. Huang, C.D.; Goo, J.; Nam, K.; Yoo, C.W. Smart tourism technologies in travel planning: The role of exploration and exploitation. *Inf. Manag.* **2017**, *54*, 757–770. [\[CrossRef\]](#)
23. Joshi, S. Social network analysis in smart tourism driven service distribution channels: Evidence from tourism supply chain of Uttarakhand, India. *Int. J. Digit. Cult. Electron. Tour.* **2018**, *2*, 255–272. [\[CrossRef\]](#)
24. Zhang, T.; Cheung, C.; Law, R. Functionality evaluation for destination marketing websites in smart tourism cities. *J. China Tour. Res.* **2018**, *14*, 263–278. [\[CrossRef\]](#)
25. Nguyen, T.T.; Camacho, D.; Jung, J.E. Identifying and ranking cultural heritage resources on geotagged social media for smart cultural tourism services. *Pers. Ubiquitous Comput.* **2017**, *21*, 267–279. [\[CrossRef\]](#)
26. Lei, W.; Chen, G.; Bei, Z.; Yong, L. Deep learning based social-aware location recommendation. *J. Terahertz Sci. Electron. Inf. Technol.* **2019**, *17*, 502–508.
27. Bo, Y.; Fei, Z.P. Review of the art of recommendation algorithms. *J. Shanxi Univ. (Nat. Sci. Ed.)* **2011**, *34*, 337–350.
28. Gao, M.T.; Xu, B.Y. Recommendation algorithm based on circular neural network. *Comput. Eng.* **2019**, *45*, 198–202.
29. Yao, X.H.; Sun, Y.; Sun, B.W.; Huang, Y. The Impact of the Urban Forest Park Recreation Environment and Perceived Satisfaction on Post-Tour Behavioral Intention—Using Tongzhou Grand Canal Forest Park as an Example. *Forests* **2024**, *15*, 330. [\[CrossRef\]](#)
30. Telonis, G.; Panteli, A.; Boutsinas, B. A Point-of-Interest Recommender System for Tourist Groups Based on Cooperative Location Set Cover Problem. *Mathematics* **2023**, *11*, 3646. [\[CrossRef\]](#)
31. Yoon, J.H.; Choi, C. Real-Time Context-Aware Recommendation System for Tourism. *Sensors* **2023**, *23*, 3679. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.