

Article

POI Route Recommendation Model Based on Symmetrical Naive Bayes Classification Spatial Accessibility and Improved Cockroach Swarm Optimization Algorithm

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: The commonly used POI route recommendation methods usually ignore the effects of tourists' interests and transportation geographical conditions, and so may not output the optimal results. To solve the problems, we propose a POI route recommendation model based on symmetrical Naive Bayes classification spatial accessibility (NBCSA) and an improved cockroach swarm optimization algorithm (ICSOA), aiming to recommend POI routes that satisfy tourists' interests and have the lowest travel costs under tourism transportation geographical conditions. Using the historical POIs visited by tourists as the training set, we construct an improved symmetrical Naive Bayes classification algorithm (NBCA), and the POIs in the destination city are divided into categories by tourists' preferences. Then we propose an improved NBCSA model to calculate the spatial accessibility field strength (SAFS) for each category's POIs. Based on the recommended POIs, we propose the ICSOA to recommend optimal POI routes. The experiment verifies that the proposed algorithm can effectively classify the POIs and recommend POIs that best match the tourists' interests and produce the lowest travel costs. Compared with the TCA and GDA method, the proposed algorithm can output the POI routes with lower travel costs and has higher algorithm execution efficiency. Among the output optimal routes, the proposed algorithm can reduce costs by 5.62% at the lowest and 52.25% at the highest.

Keywords: symmetrical Naive Bayes classification; SAFS; ICSOA; tourism recommendation



Citation: Zhou, X.; Zhang, Z.; Liang, X.; Su, M. POI Route Recommendation Model Based on Symmetrical Naive Bayes Classification Spatial Accessibility and Improved Cockroach Swarm Optimization Algorithm. *Symmetry* **2024**, *16*, 424. <https://doi.org/10.3390/sym16040424>

Academic Editor: Tomohiro Inagaki

Received: 16 March 2024

Revised: 28 March 2024

Accepted: 1 April 2024

Published: 3 April 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

POI route recommendation methods include collaborative filtering, association rule, knowledge-based, etc., which focus on the operation efficiency and accuracy of the algorithms. For POI route recommendation, the commonly used methods include extracting route information through the locations of the travel photos, tracing tourists' GPS spatial trajectories, and recommending via travel logs and guides. The historical POI route information is directly recommended for tourists with similar interests. These methods have the following problems. First, the goal of recommending is not merely to improve the efficiency and accuracy of the algorithm, but also to satisfy tourists' interests; the commonly used methods lack studies on user interests, resulting in poor matching between the recommendation results and the tourists' interests. Second, the commonly used methods usually use evaluation scores as the criterion. It subjectively evaluates the feature attributes of POIs. Tourism activities are performed in a geospatial environment and are constrained by geospatial conditions; therefore, the spatial accessibility and distribution of POIs should

also be considered. Third, POI route recommendation cannot only be considered as the mining of historical information. It does not consider POIs that conform to the tourists' interests and the constraints of the geospatial environment, which causes deviations in route recommendations. Thus, POI route recommendation should combine the factors of tourists' interests, POI feature attributes and tourism geospatial constraints.

1.2. Related Works

Scholars' research on tourism and POI route recommendation mainly includes the following aspects.

- (1) Optimizing the recommendation algorithm. Chen et al. [1] studied the influence of power law distribution of long-tailed data on tourism recommendation results from a mathematical perspective. Zheng et al. [2] built a tourism recommendation model by using the neural network and matrix decomposition method to solve the information overload problem. Chen et al. [3] constructed a tourism recommendation model using the graph representation learning method and sequence mining, solving the problem of complex sequence semantics in tourism mining. Zheng et al. [4] proposed a recommendation algorithm based on user similarity, POI popularity and time context, which solved the problem of decreased user decision-making efficiency. Lin et al. [5] proposed a tourism recommendation method based on the constrained association rule algorithm. It obtained higher algorithm efficiency and made mining results more consistent with user needs. Cheng et al. [6] proposed a POI recommendation algorithm based on multidimensional feature clustering and user scoring. It solved the data sparsity problem.
- (2) Mining user interests based on historical data, browsing data and evaluation data, etc. Ahn Jinhyun et al. [7] proposed a tourism recommendation by users' behavior of refreshing and browsing tourism websites. Hong et al. [8] used a multi-scale tensor model to construct a tourism recommendation model based on spatiotemporal data. They analyzed the travel data with spatiotemporal changes and used spatiotemporal trajectories as the basis for recommendation. Abbasi Moud Zahra et al. [9] constructed a tourism recommendation model based on semantic clustering and sentiment analysis, and tourist social networks were used to extract tourists' interests. Jeong Chi Seo et al. [10] conducted data analysis based on tourist social networks, and used deep learning to mine data and set up a tourism recommendation system. Cui et al. [11] used sentiment intensity analysis and the TOPSIS sorting method to construct a recommendation algorithm based on user online comments. Cui et al. [12] proposed a tourism recommendation method based on user profiles. It obtained basic data and behavioral information of users and generated user profiles. On the aspect of big data application and optimization, Ahmed H. Almulih et al. [13] used intelligent computing technology to mine and analyze big data, which obtained the factors and problems that cause data security risks and improved the data security. By evaluating the correctness of the dataset, a dynamic digital healthcare data breach environment using a fuzzy based computational technique was established. This study helps to provide a new method for utilizing tourism big data and protecting the security of tourism big data, which can guarantee the safety for tourists' information.
- (3) Integrating tourism scene features such as POIs and geographic information for recommendation. Han Shanshan et al. [14] extracted spatiotemporal data from photos labeled with geographic information and recommended POIs based on geographic labels. Filipe Santos et al. [15] analyzed the functionality and accessibility of POIs, combining these with the physical and mental conditions of tourists, then constructed an individualized tourism recommendation system. Zhang et al. [16] set up a rural tourism recommendation model based on seasonal features and the geographic information extraction model. They analyzed the seasonal features of rural tourism and integrated geographic information factors.

1.3. Research Objectives and Algorithm Mechanism

1.3.1. Research Objectives

We propose a POI route recommendation model based on the symmetrical NBCSA and ICSOA. The aim is to recommend POIs and routes that not only satisfy tourists' interests but also achieve optimal spatial distribution and minimize the travel costs. It solves three problems. (1) Feature attribute mining is conducted on the POIs that the tourists have once visited to obtain interest data. Based on the data criterion, the symmetrical Naive Bayes classification algorithm (NBCA) is set up for the classification on the POIs in the destination city. Thus, the classification on the POIs is completely based on tourists' interests. (2) Under the geospatial constraints of the tourism destination, the model obtains the POIs with optimal spatial distributions and the lowest travel costs, and realizes a POI recommendation with the best-matched feature attributes and the lowest spatial costs. (3) Based on the recommended POIs, under the travel conditions and geospatial constraints, the improved cockroach swarm optimization algorithm (ICSOA) is designed to achieve the optimal solution, providing tourists with the lowest travel costs and the optimal routes.

1.3.2. Algorithm Mechanism

Figure 1 shows the modeling process of the proposed recommendation algorithm. The algorithm mechanism is as follows. Tourists firstly input the POIs they have visited before and classify their preferences. As to the classification of POIs and preferences, a training set X based on the historical POIs is established, which is used to construct a Naive Bayes classification algorithm (NBCA). In step ①, the text mining method is used to determine the feature vector L of the training set X , which stores the classification features of the NBCA. In step ②, in order to avoid the situation that a conditional probability gets to the value 0 when constructing the NBCA, a disturbance factor is introduced for the conditional probability, then an improved NBCA is established. The improved NBCA is used to classify the POIs in the destination city, and the classification results will match the interests and needs of tourists. In step ③, based on the classification results, a spatial accessibility field strength (SAFS) model is constructed to calculate the SAFS for each classified POI. Based on the calculation results, a classification POI recommendation algorithm based on the SAFS model is constructed to recommend the optimal POIs. In step ④, an improved cockroach swarm optimization algorithm (ICSOA) is constructed based on the recommended POIs and the urban geospatial environment. Finally, a POI route recommendation model based on the ICSOA is constructed.

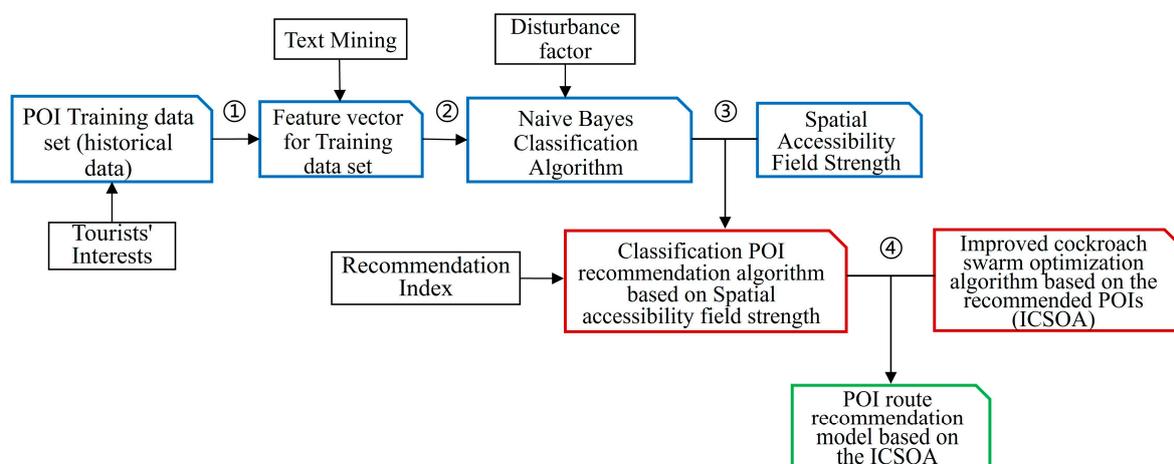


Figure 1. The modeling process of the proposed recommendation algorithm.

The proposed POI route recommendation model can accurately recommend POIs and POI routes for different cultural contexts and user groups. The algorithm mechanism that is adapted to the users from different cultural contexts is as follows.

Firstly, as to the process of the tourism recommendation model in Figure 1, the goal of this algorithm is to recommend POIs and POI routes for a single user. After inputting interest needs from any cultural contexts and user groups, this algorithm can accurately match POIs based on user interests, and search for the optimal POI route based on the recommended POIs and geospatial conditions. When the tourists' cultural contexts, the once-visited POIs and the preferences to the POIs change, the algorithm can still recommend relative POIs according to the dynamic input interest data, and then search for different optimal routes accordingly. That is, the recommendation results are always generated based on the tourists' interests, which can fully match the tourists' needs.

Secondly, the constructed NBCA in the recommendation model is a "filter" for different cultural contexts and user groups' interest needs. Its "filtering" function is mainly manifested in the following aspects: (1) User groups with different cultural contexts have different perceptions and interests in the once-visited POIs and the POIs in the destination city. Therefore, when collecting the training set, users will initially categorize the POIs they have once visited based on their own perceptions and preferences. This step is a raw collection of user interests. (2) Based on different user interests, the feature labels are designed to establish quantitative feature labels and label data ranges for the NBCA. The trained classification algorithm is based on the users' cultural contexts and interests, and thus, the classification on the POIs in the destination city is also based on the users' cultural contexts and interests. The classification results are highly accurate.

Thirdly, tourists with different cultural contexts have different preferences for the starting points of POI routes, the types of POIs they expect to visit and the specific quantities of POIs to be visited. They also have different preferences for the spatial accessibilities and feature attributes of POIs. Therefore, the recommended POIs output by the classification results and the SAFS model are also different. The filtering mechanism of the SAFS model can adapt to the needs of users from different cultural contexts and recommend precise POIs based on their interests. In terms of POI route recommendation, tourists from different cultural contexts have different choices of transportation tools, perceptions of geospatial environments, scheduled travel times, and planned travel costs. The recommendation algorithm ultimately searches for the optimal POI route based on these constraints, and the results are also different. Therefore, in terms of the POI route recommendation mechanism, this algorithm can adapt to the interests of tourists from different cultural contexts, and search and recommend POI routes based on the individualized interests.

2. POI Recommendation Algorithm Based on the Symmetrical NBCSA

Previous experiences are the best source of interests, and historical interests have a symmetrical relationship with the current preferences of tourists. Tourists who have visited certain POIs will exhibit certain degrees of preferences. Obtaining and classifying the once-visited POIs by preferences can clarify the tourists' interests, and then obtain the original training data set for constructing the classification model. The Naive Bayes algorithm has the advantages of high efficiency, accurate classification ability and suitability for small-scale data, making it suitable for POI classification [17,18]. Combined with POI feature attributes, the acquisition of classifier training data is improved. A classification algorithm incorporating a disturbance factor is proposed by using the preference classification and quantitative attributes of POIs as training data, which is used for the classification on the POIs in the destination city. Based on the classification results and geospatial constraints, the SAFS model is set up to search for the optimal POIs.

2.1. The Symmetrical NBCA Based on Tourists' Interests

The symmetrical NBCA includes two aspects: the acquisition of the training data set and the foundation of the model. The training data set X is composed of the N number of once-visited POIs with preference classifications provided by tourists. Each sample POI $x_{(i)}$ has feature attributes. The training data set X is grouped into $C_{(i)}$: $\{C_{(1)}$: favorite;

$C_{(2)}$: like; $C_{(3)}$: dislike}. The feature attributes are obtained by text mining in the tourism encyclopedia big data [19].

2.1.1. The Foundation of the Feature Vector on Sample POI

The feature vector L of the sample POI $x_{(i)}$ contains the feature attributes of the POI, which is the symbol to distinguish it from other POIs. The feature vector L is composed of k number of feature labels $l_{(i)}$, $0 < i \leq k, i, k \in \mathbf{N}$, in which the number of text feature label to express feature attribute is $k_{(1)}$, while the number of value labels to express tourism feature is $k_{(2)}$. Based on the $k_{(1)}$ number of text feature labels, the text feature label matrix $L_{(i,j)}$ is set up as Formula (1). In the formula, $l_{(i,j)}$ is the No. j sub-label of the text label $l_{(i)}$. Each sub-label $l_{(i,j)}$ represents a kind of text note for the text label group $l_{(i)}$. By calculating the weight value for the text label group $l_{(i)}$, the constraint ranges for the $k_{(1)}$ number of text feature labels in the vector L are confirmed.

$$L_{(i,j)} = \begin{bmatrix} l_{(1,1)} & l_{(1,2)} & \dots & l_{(1,p)} \\ l_{(2,1)} & l_{(2,2)} & \dots & l_{(2,p)} \\ \dots & \dots & \dots & \dots \\ l_{(k(1),2)} & \dots & \dots & l_{(k(1),p)} \end{bmatrix} \tag{1}$$

Search for the $|D|$ number of text documents of POI $x_{(i)}$ from the corpus in the encyclopedia and tourism service platform. Based on the feature label $l_{(i)}$, count the occurrence number of each sub-label $l_{(i,j)}$ of $l_{(i)}$ in the $|D|$ number of documents and get the initial value for the word frequency of $TF_{l_{(i)}}$. Calculate the weight values for all the labels $l_{(i)}$ by the TF-IDF. The process is shown as follows.

Step 1. Search and count the occurrence number of label $l_{(1)}$ in the matrix $L_{(i,j)}$. Search and count the occurrence number of sub-label $l_{(1,j)}$ in the $|D|$ number of documents, noted as $n_{\cdot l_{(1,j)}}$. Iterate until $j = p + 1$ and stop searching. Output $n_{\cdot l_{(1)}}$ as Formula (2) shows.

$$n_{\cdot l_{(1)}} = \sum_{j=1}^p n_{\cdot l_{(1,j)}} \tag{2}$$

Step 2. Search and count the occurrence number of label $l_{(2)}$ in the matrix $L_{(i,j)}$. Search and count the occurrence number of sub-label $l_{(2,j)}$ in the $|D|$ number of documents, noted as $n_{\cdot l_{(2,j)}}$. Iterate until $j = p + 1$ and stop searching. Output $n_{\cdot l_{(2)}}$.

Step 3. In line with the Step 1–Step 2, search and count the occurrence number $n_{\cdot l_{(i)}}$ of label $l_{(i)}$ in the matrix $L_{(i,j)}$. Iterate until $i = k_{(1)} + 1$ and stop searching. Output $n_{\cdot l_{(k(1))}}$, $0 < i \leq k_{(1)}, i, k_{(1)} \in \mathbf{N}$.

Step 4. Calculate the word frequency $TF_{l_{(i)}}$ of label $l_{(i)}$ in the $|D|$ number of documents as Formula (3).

$$TF_{l_{(i)}} = \frac{\sum_{j=1}^p n_{\cdot l_{(i,j)}}}{\sum_{i=1}^{k(1)} \sum_{j=1}^p n_{\cdot l_{(1,j)}}} \tag{3}$$

Step 5. Calculate the inverse document frequency $IDF_{l_{(i)}}$ and feature label weight $TFIDF_{l_{(i)}}$ of label $l_{(i)}$ in the $|D|$ number of documents, shown as Formulas (4) and (5). $\left| \left\{ s : l_{(i)} \in d_s \right\} \right|$ represents the occurrence number of documents with the label $l_{(i)}$: $\left\{ i, j \in (0, p] \mid l_{(i,j)} \right\}$ in the $|D|$ number of documents.

$$IDF_{l_{(i)}} = \log \frac{|D|}{\left| \left\{ s : l_{(i)} \in d_s \right\} \right| + 1} \tag{4}$$

$$\text{TFIDF}_{l(i)} = \text{TF}_{l(i)} \times \text{IDF}_{l(i)} \quad (5)$$

The weight value $\text{TFIDF}_{l(i)}$ of $l(i)$ is the quantization for the $k_{(1)}$ number of text labels of $x_{(i)}$. Through the weight value $\text{TFIDF}_{l(i)}$, the constraint ranges of $k_{(1)}$ number of text labels are confirmed. The quantization values for the other $k_{(2)}$ number of tourism attributes of $x_{(i)}$ are searched out or calculated from the tourism big data to confirm the constraint ranges. The $1 \times k$ dimension vector which is composed of the weight values $\text{TFIDF}_{l(i)}$ of $k_{(1)}$ number of text labels and the quantization values of the $k_{(2)}$ number of tourism attributes is defined as the feature vector L of the POI $x_{(i)}$. According to the modeling conditions of NBCA and the quantization values of feature vector L labels, convert each value into a value range and store them into the training data set.

2.1.2. The Symmetrical NBCA Based on Tourists' Interests

As to arbitrary $\forall x_{(i)}$ in the N number of POIs, it must relate to the label vector: $\{\text{TFIDF}_{l(i)}, \text{TEXT}_{l(i)}, C_{(i)}\}$, in which $\text{TFIDF}_{l(i)}$ stands for the $k_{(1)}$ number of text feature label quantization ranges, $\text{TEXT}_{l(i)}$ stands for the $k_{(2)}$ number of tourism attribute label quantization ranges and $C_{(i)}$ stands for the preference categories of $x_{(i)}$. The k number of feature labels in vector L are mutually independent, which conforms to the modeling conditions for the NBCA.

Set $y_{(i)}$ as the POI to be classified in the destination city. The classifying is to match the $y_{(i)}$ with the once-visited POIs $x_{(i)}$ as well as classifications $C_{(i)}$. Through the classification algorithm, $y_{(i)}$ will be absorbed into $C_{(i)}$ with the highest calculation conditional probability. In this process, the tourists' interests are set as the criterion to measure whether $y_{(i)}$ should be an interested POI. Suppose H is the assumption that the sample $y_{(i)}$ belongs to the category $C_{(i)}$. Then the basic idea of the Bayes theorem is to calculate the conditional probability $P(C_{(i)}|y_{(i)})$ which makes the assumption H be true. Formula (6) is the Bayes posterior classification model.

$$P(C_{(i)}|y_{(i)}) = \frac{P(y_{(i)}|C_{(i)})P(C_{(i)})}{P(y_{(i)})} \quad (6)$$

As to the interest preference category $C_{(i)}$ confirmed by tourists, $0 < i \leq t, i, t \in \mathbf{N}$, the Naive Bayes classifier will predict that the sample $y_{(i)}$ belongs to the category with the highest posterior probability. Namely, as to the t number of categories $C_{(i)}$ in the sample space, the condition that the sample $y_{(i)}$ belongs to the category $C_{(i)}$ is when and only when $P(C_{(i)}|y_{(i)}) > P(C_{(j)}|y_{(i)})$, $0 < i, j \leq t, i \neq j$. At this time, the category $C_{(i)}$ with the highest posterior probability $P(C_{(i)}|y_{(i)})$ is the maximum posterior assumption. The improved NBCA to classify the sample $y_{(i)}$ in the destination city is designed as follows.

Step 1. As to the Bayes posterior classification model, the probability $P(y_{(i)})$ of the sample $y_{(i)}$ is a constant value to each category $C_{(i)}$. Convert the issue of calculating $P(C_{(i)}|y_{(i)})$ in the Bayes posterior classification model to calculating the $P(y_{(i)}|C_{(i)})P(C_{(i)})$, which is defined as the conversion value $\delta_{(y_{(i)})}$.

Step 2. Based on the N number of $x_{(i)}$ and categories set by the tourists, the feature label vector for each sample $x_{(i)}$ is obtained as $L: \{\text{TFIDF}_{l(i)}, \text{TEXT}_{l(i)}, C_{(i)}\}$. The quantity of sample POI $x_{(i)}$ for each category $C_{(i)}$ is $s_{(i)}$ and the total quantity of the sample $x_{(i)}$ is S . The probability for each sample's classification is different, calculated by Formula (7).

$$P(C_{(i)}) = \frac{s_{(i)}}{S} \quad (7)$$

Step 3. Continue to calculate $P(y_{(i)}|C_{(i)})$. Since the attributes in the vector L are mutually independent, the calculation method of $P(y_{(i)}|C_{(i)})$ is designed as Formula (8). The $P(l_{(k)}|C_{(i)})$ is evaluated by the training samples. Since the sample's feature attributes $l_{(k)}$ are all discrete values, the calculation method of $P(l_{(k)}|C_{(i)})$ is designed as Formula (9), in which $s_{(i,k)}$ is the quantity of the training samples that contains $l_{(k)}$ in category $C_{(i)}$ and $s_{(i)}$ is the quantity of samples in category $C_{(i)}$. According to the attribute labels and actual tourism conditions, the situation of $s_{(i,k)} = 0$ may occur on the label $l_{(k)}$. Introduce the disturbance factor ε as the adjustment for the label $l_{(k)}$ when its quantity is 0, designed as Formula (10).

$$P(y_{(i)}|C_{(i)}) = \prod_{k=1}^n P(l_{(k)}|C_{(i)}) \quad (8)$$

$$P(l_{(k)}|C_{(i)}) = \frac{s_{(i,k)}}{s_{(i)}} \quad (9)$$

$$P(y_{(i)}|C_{(i)})_{\varepsilon}^{\Delta} = \prod_{k=1}^n \left(\frac{s_{(i,k)}}{s_{(i)}} + \varepsilon \right) \quad (10)$$

Step 4. Calculate the conversion value $\delta_{(y_{(i)})}$ of the Bayes posterior classification model. The $C_{(i)}$ with the maximum value $\max \delta_{(y_{(i)})}$ is the category that the sample $y_{(i)}$ belongs to. Store the probability value.

2.2. POI Recommendation Algorithm Based on NBCSA

As to the m number of POIs in the destination city, set the quantity of POIs that are grouped into $C_{(i)}$ as $m_{(i)}$. POIs are distributed in the geospatial environment, and not all the POIs are optimal in spatial distribution; thus, the spatial feature is the key to determining travel costs [20–22]. The SAFS model is constructed to calculate spatial accessibility and recommend the POIs with the best-matched attributes and the optimal spatial distribution.

2.2.1. SAFS Model of the Naive Bayes Classification

Tourists start trips at certain initial points S . POIs have spatial attributes such as latitude, longitude, spatial distance, etc. Confined by geospatial conditions, the process of traveling from the initial point S to the POI $y_{(i)}$ will be hindered by the space and the time. The weaker the hindrance is, the higher the spatial accessibility of $y_{(i)}$ will be, and the lower the travel costs will be. Constructing SAFS is the key to quantifying the spatial distributions for the POIs.

Definition 1. Research domain spatial grid G . As to the research scope, design a spatial range ϕ with length L and width W to cover the whole domain. Divide the ϕ into $u \times v$ number of unit grids $g_{(x,y)}$ and each grid represents the internal micro-region of the destination city. x and y represent the coordinates of the grid $g_{(x,y)}$. The spatial location for $y_{(i)}$ is the spatial coordinate $L_{y_{(i)}} = (x_{(y_{(i)})}, y_{(y_{(i)})})$.

Definition 2. SAFS of $A_{(y_{(i)},S)}$ and $A^*_{(y_{(i)},S)}$ based on the spatial grid G . The absolute SAFS of $A_{(y_{(i)},S)}$ represents the space and time hindrance between the initial point S and the POI $y_{(i)}$. Set the S coordinate as $L_S = (x_{(S)}, y_{(S)})$. $A_{(y_{(i)},S)}$ is constructed as Formula (11), and $\zeta_{(0)}$ is the normalization coefficient. Introduce the relative SAFS of $A^*_{(y_{(i)},S)}$, namely, the ratio between the absolute SAFS of $y_{(i)}$ and S to the sum of the absolute SAFS of other samples $y_{(i)}$ and S , shown as Formula (12).

$$A_{(y_{(i)},S)} = \zeta_{(0)} \times \frac{1}{\sqrt{|x_{(y_{(i)})} - x_{(S)}|^2 + |y_{(y_{(i)})} - y_{(S)}|^2}} \quad (11)$$

$$A^*_{(y(i),S)} = \frac{A_{(y(i),S)}}{\sum_{i=1}^m A_{(y(i),S)}} \tag{12}$$

Definition 3. POI SAFS matrix $\mathbf{A}_{(x,y)}$. Based on the $u \times v$ number of unit grids $g_{(x,y)}$ contained in the grid G of the spatial range ϕ , set a $u \times v$ -dimension matrix with the initial element of the unit grid $g_{(x,y)}$ at the origin point of the coordinates XY_G . The matrix is used to store $A^*_{(y(i),S)}$ of $y_{(i)}$ and is defined as the POI SAFS matrix $\mathbf{A}_{(x,y)}$. According to the definition, the matrix element meets the condition of Formula (13).

$$\mathbf{A}_{(x,y)} = \begin{cases} 0, & y_{(i)} \notin g_{(x,y)} \\ A^*_{(y(i),S)}, & y_{(i)} \in g_{(x,y)} \end{cases} \tag{13}$$

Definition 4. POI SAFS of Φ . Take the POI SAFS matrix $\mathbf{A}_{(x,y)}$ as the base and build a three-dimensional spatial field visualization model; this model is defined as the POI SAFS of Φ . Figure 2 is the modeling process of the POI SAFS. Figure 2a shows the research domain, Figure 2b shows the spatial range, Figure 2c shows the spatial grid, Figure 2d shows the spatial coordinates, Figure 2e shows the spatial accessibility coordinates and Figure 2f shows the spatial accessibility field strength.

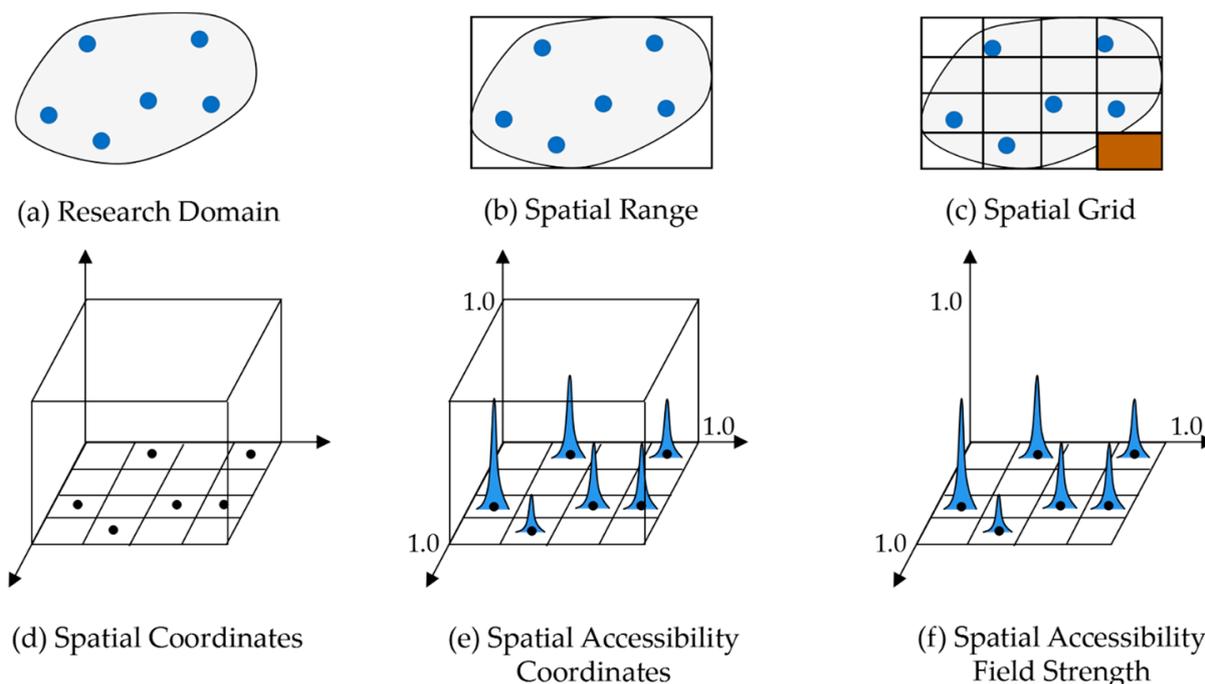


Figure 2. The modeling process of the POI SAFS. In (a–c), the blue dots represent the POIs, the brown range represents a spatial grid. In (e,f), the blue peak value represents the spatial accessibility field strength. The value 1.0 represents the measurement value of the spatial accessibility field strength.

2.2.2. POI Recommendation Algorithm Based on the SAFS Model

The recommendation of POI in real-world tourism scenarios should comprehensively consider the tourists’ interests and the spatial optimality of POIs, so that the POIs not only meet tourists’ interests, but also minimize the travel costs. The POI recommendation algorithm based on the SAFS model is established as follows, and Figure 3 shows the modeling process.

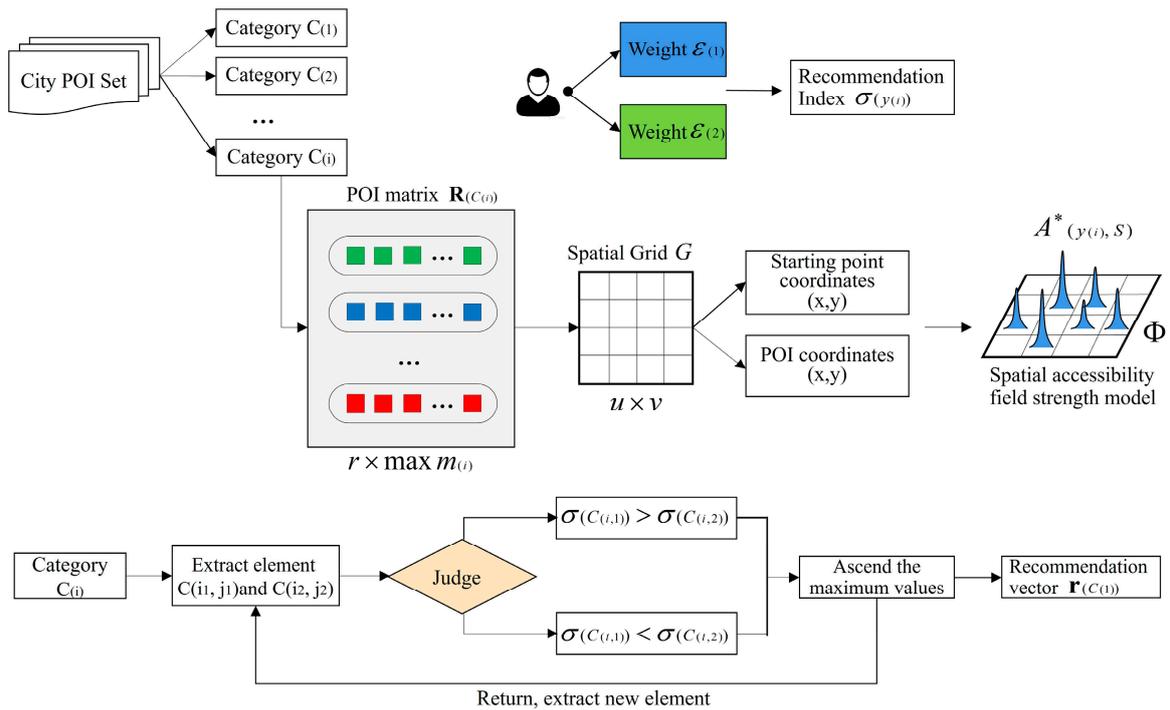


Figure 3. The modeling process of classification of POI recommendations. The different colors in POI matrix $\mathbf{R}_{(C(i))}$ mean the different rows of the matrix for different categories, and each color represents one category. The blue color in model Φ represents the spatial accessibility field strength.

Step 1. Set up the recommendation matrix $\mathbf{R}_{(C(i))}$ for the category $C(i)$.

In this step, the matrix $\mathbf{R}_{(C(i))}$ meets the following conditions:

- (1) The dimension of $\mathbf{R}_{(C(i))}$ is $r \times \max m(i)$.
- (2) The symbol r represents the number of the category $C(i)$.
- (3) The symbol $m(i)$ represents the quantity of POIs in the category $C(i)$ after classifying.
- (4) The No. i row stores POIs of $C(i)$. The No. j column stores the No. j POI in $C(i)$.
- (5) If the element in the matrix does not store POI, it is set as element number 0.
- (6) In the matrix, the arbitrary rows $\forall i$ and $\forall i'$, and the arbitrary columns $\forall j$ and $\forall j'$, are all nonlinear correlated.
- (7) The row rank meets the condition $\text{rank}(\mathbf{R}_{(C(i))})_{ro} = r$.
- (8) The column rank meets the condition $\text{rank}(\mathbf{R}_{(C(i))})_{co} = \max m(i)$.

Step 2. Set up the spatial grid G of the research domain ϕ .

- (1) Form the $u \times v$ dimension unit grids for the grid G .
- (2) Confirm the coordinates $L_{y(i)} = (x_{(y(i))}, y_{(y(i))})$ for each POI $y(i)$ in category $C(i)$ in the grid G , $0 < i \leq m$, $i, m \in \mathbf{N}$.
- (3) The coordinates of all the POIs in r number of categories are confirmed.

Step 3. Calculate the spatial accessibility field strength (SAFS) for POIs.

- (1) Confirm the initial point S , and set its coordinate $L_S = (x_{(S)}, y_{(S)})$ in the grid G .
- (2) Calculate the SAFS $A^*(y(i), S)$ for all the POIs $y(i)$ in categories $C(i)$.
- (3) Generate the SAFS matrix $\mathbf{A}_{(x,y)}$ and SAFS model Φ .

Step 4. Set up the recommendation vector $\mathbf{r}_{(C(i))}$ for category $C(i)$. The dimension of the vector is $1 \times m(i)$ and it contains $m(i)$ number of elements.

Step 5. Make a judgement on the attribute and SAFS for $C(i)$. $\delta_{(y(i))}$ is the Naive Bayes conversion value; $A^*(y(i), S)$ is the SAFS.

Step 6. Introduce the tourist need weight $\varepsilon_{(x)}$ and recommendation index $\sigma_{(y_{(i)})}$.

- (1) Define $\varepsilon_{(1)}$ as the factor to confine $\delta_{(y_{(i)})}$, which measures the preferences of the tourists on the POI attributes.
- (2) Define $\varepsilon_{(2)}$ as the factor to confine $A^*_{(y_{(i)},S)}$, which measures the preferences of the tourists on the POI spatial accessibilities.
- (3) For the two factors $\varepsilon_{(1)}$ and $\varepsilon_{(2)}$, the higher the value is, the more likely the tourists prefer the factor.
- (4) The constraint is set as $\varepsilon_{(1)} + \varepsilon_{(2)} = 1$.
- (5) The recommendation index model is designed as Formula (14).

$$\sigma_{(y_{(i)})} = \varepsilon_{(1)} \cdot \delta_{(y_{(i)})} + \varepsilon_{(2)} \cdot A^*_{(y_{(i)},S)} \quad (14)$$

Step 7. Tourists provide $\varepsilon_{(x)}$. Calculate $\sigma_{(y_{(i)})}$ of POIs in categories $C_{(i)}$, and form full-ranked $\mathbf{r}_{(C_{(i)})}$.

- (1) Extract from the No.1 row $C_{(1)}$ the No.1 element $C_{(1,1)}$ and No.2 element $C_{(1,2)}$. Make a judgement.
 - (i) If $\sigma_{(C_{(1,1)})} > \sigma_{(C_{(1,2)})}$:
The $\sigma_{(y_{(i)})}$ of $C_{(1,1)}$ is higher than $C_{(1,2)}$. Store $C_{(1,1)}$ and $C_{(1,2)}$ into the No.1 and No.2 elements $\mathbf{r}_{(1)}$ and $\mathbf{r}_{(2)}$ in $\mathbf{r}_{(C_{(i)})}$.
 - (ii) If $\sigma_{(C_{(1,1)})} \leq \sigma_{(C_{(1,2)})}$:
The $\sigma_{(y_{(i)})}$ of $C_{(1,2)}$ is higher than $C_{(1,1)}$. Store $C_{(1,2)}$ and $C_{(1,1)}$ into the No.1 and No.2 elements $\mathbf{r}_{(1)}$ and $\mathbf{r}_{(2)}$ in $\mathbf{r}_{(C_{(i)})}$.
- (2) Extract the No.3 element $C_{(1,3)}$ and make a judgement.
 - (i) If $\sigma_{(C_{(1,1)})} > \sigma_{(C_{(1,2)})}$:
 - ① If $\sigma_{(C_{(1,1)})} > \sigma_{(C_{(1,2)})} \geq \sigma_{(C_{(1,3)})}$, keep $C_{(1,1)}$ and $C_{(1,2)}$; store $C_{(1,3)}$ into $\mathbf{r}_{(3)}$.
 - ② If $\sigma_{(C_{(1,1)})} \geq \sigma_{(C_{(1,3)})} > \sigma_{(C_{(1,2)})}$, store $C_{(1,2)}$ into $\mathbf{r}_{(3)}$; store $C_{(1,3)}$ into $\mathbf{r}_{(2)}$.
 - ③ If $\sigma_{(C_{(1,3)})} > \sigma_{(C_{(1,1)})} > \sigma_{(C_{(1,2)})}$, store $C_{(1,3)}$, $C_{(1,1)}$ and $C_{(1,2)}$ into $\mathbf{r}_{(1)}$, $\mathbf{r}_{(2)}$ and $\mathbf{r}_{(3)}$.
 - (ii) If $\sigma_{(C_{(1,1)})} \leq \sigma_{(C_{(1,2)})}$:
 - ① If $\sigma_{(C_{(1,3)})} < \sigma_{(C_{(1,1)})} \leq \sigma_{(C_{(1,2)})}$, keep $C_{(1,1)}$ and $C_{(1,2)}$; store $C_{(1,3)}$ into element $\mathbf{r}_{(3)}$.
 - ② If $\sigma_{(C_{(1,1)})} \leq \sigma_{(C_{(1,3)})} \leq \sigma_{(C_{(1,2)})}$, store $C_{(1,1)}$ into $\mathbf{r}_{(3)}$; store $C_{(1,3)}$ into $\mathbf{r}_{(2)}$.
 - ③ If $\sigma_{(C_{(1,1)})} \leq \sigma_{(C_{(1,2)})} < \sigma_{(C_{(1,3)})}$, store $C_{(1,3)}$, $C_{(1,2)}$ and $C_{(1,1)}$ into $\mathbf{r}_{(1)}$, $\mathbf{r}_{(2)}$ and $\mathbf{r}_{(3)}$.
- (3) In line with Step (1)–Step (2), continue searching the No. j element ($C_{(1,j)}$).
 - (i) Calculate the $\sigma_{(C_{(1,j)})}$.
 - (ii) Descend to store the index $\sigma_{(C_{(1,j)})}$ into vector $\mathbf{r}_{(C_{(1)})}$.
 - (iii) The stop searching condition is $j = m_{(i)} + 1$.
 - (iv) When the searching stops, output the full-ranked vector $\mathbf{r}_{(C_{(1)})}$.
- (4) Continue searching the other category rows $C_{(i)}$. Output the full-ranked vector $\mathbf{r}_{(C_{(i)})}$. Traverse $i \sim (0, r]$, $i, r \in \mathbf{N}$.

Step 8. Output all vectors $\mathbf{r}_{(C_{(i)})}$ for categories $C_{(i)}$.

- (1) In vector $\mathbf{r}_{(C_{(i)})}$, the front-end elements have the higher recommendation index $\sigma_{(y_{(i)})}$, and will be preferentially recommended.
- (2) According to the actual needs of tourists, w number of optimal POIs in categories $C_{(i)}$ will be taken as the tour route nodes to be visited.

3. POI Route Recommendation Model Based on the ICSOA

3.1. Modeling Principle

The POI route algorithm should be constructed on the recommended POIs. Tourists start from the points S and visit the w number of POIs. The tour will be influenced by several factors: transportation mode, travel distance, travel time, travel fee. It is necessary to find out the POI routes with the lowest costs. Thus, the modeling problem converts to the searching for an optimal route conforming to the geospatial conditions and passing through all POIs [23–25]. The cockroach swarm optimization algorithm (CSOA) is an excellent algorithm that can search for the global optimal solution [26,27]. We analyze the need and necessity of using the CSOA to construct a POI route recommendation model and the idea of optimizing the tourism recommendation model from the aspects of the algorithm characteristics, the algorithm execution principles and the route searching scenarios based on geospatial environment.

(1) The need and necessity of the algorithm

Firstly, the route searching algorithm is divided into two parts: interval local searching and global searching. Local searching is to find out the optimal route between two POIs, passing through road nodes. Global searching is to find out the optimal route through all POIs based on the local searching. There are many road nodes and POIs in urban geographic space, so it is necessary to choose a certain intelligent algorithm with fast computation speed, rapid convergence speed, and the ability to search for the global optimal solutions. The CSOA can meet this requirement.

Secondly, as to the interval local searching, the CSOA achieves the step size transformation through a 2-opt approach, forming new cockroach individuals. The calculation process for the generation cost of each cockroach individual is a linear superposition on the spatial cost between two road nodes, and the calculation speed is fast. The process of searching for the optimal cockroach individual involves traversing a certain number of steps and combining them with the sorting algorithm. The maximum amount of its computation is the factorial calculation, with low time complexity and fast convergence. Therefore, the CSOA is very suitable for the interval shortest distance calculation involving multiple road nodes and is suitable for the interval local searching.

Thirdly, as to the global searching, when traversing all the POIs, the CSOA achieves step size transformation through a 2-opt approach, forming new cockroach individuals. The calculation process for the generation cost of each cockroach individual is a linear superposition on the spatial cost between two POIs, and the computational speed is fast. The process of searching for the optimal cockroach individual is still a combination of the factorial operation and sorting algorithm, with low time complexity and fast convergence. Therefore, the CSOA is suitable in searching for the global optimal solution that includes all POIs.

Fourthly, compared to the other optimization algorithms that are prone to falling into local optimal solutions, the CSOA has a homing strategy that can jump out of the local optimal solution and ultimately search for the global optimal solution. Therefore, it is very suitable in searching for the global optimal route passing through all POIs.

(2) The application to optimize POI route recommendation model

Firstly, traditional POI route recommendation methods use historical big data, and directly recommend routes previously visited by tourists with similar interests to current tourists, or determine tourists' preferences and recommend similar routes based on the GPS trajectories on POIs visited by tourists. These methods are not based on the interests of current tourists, and they directly recommend routes visited by historical tourists, which does not conform to the current tourism scenarios and tourists' interests. The constructed CSOA can solve this problem. The algorithm is based on the current tourists' tourism scenarios and geographic information conditions, aiming to match the tourist interests. Finally, all the POIs on the route meet the tourists' interests, which improves the accuracy of the recommended routes.

Secondly, the goal of the ICSOA is to find out the optimal route between two POIs and the global optimal route passing through all POIs. The algorithm is based on the current geospatial data and the constraints on the travel costs. It performs the route-searching steps under the condition of the selected transportation mode. The final searched POI route is the route with the lowest costs, which can save travel costs and improve tourists' satisfaction.

Thirdly, compared to the other POI route planning algorithms such as the Dijkstra algorithm and A* algorithm, the ICSOA has fast convergence speed and low time complexity. Its homing strategy can jump out of the local optimal solutions and ultimately find out the global optimal solution, relating to the optimal POI route.

Based on the above analysis, in this section, an ICSOA that fuses geospatial constraints is proposed. It can output POI routes that best match tourists' interests and have the lowest travel costs under the selected transportation mode. Cockroach crawling is used to search for the shortest path in the sub-interval and confirm the pass hindrance factors. The factors are used to establish the sub-interval hindrance function, based on which the interval hindrance function is designed. Then the cockroach crawling is used to search for the optimal solution of the interval hindrance function to recommend the optimal route with the lowest travel costs.

3.2. Modeling Process

3.2.1. Related Definition

Definition 5. Cockroach crawling interval $C_r(S, t_{(x)})$ and cockroach crawling sub-interval $C_r(t_{(i)}, t_{(j)})$. Set the w number of POIs as $t_{(i)}$, $0 < i \leq w$, $i, w \in \mathbf{N}$. The interval between the initial point S and the terminal POI $t_{(x)}$ forms a complete route, and this interval is defined as the cockroach crawling interval $C_r(S, t_{(x)})$, $0 < x \leq w$, $x, w \in \mathbf{N}$. Tourists traveling from point $t_{(i)}$ to $t_{(j)}$ will pass several road nodes. This traveling interval is defined as the cockroach crawling sub-interval $C_r(t_{(i)}, t_{(j)})$. In $C_r(S, t_{(x)})$, the cockroach is noted as $Z_{(a,h)}$, which represents a POI route. In $C_r(t_{(i)}, t_{(j)})$, the cockroach is noted as $Z_{(b,s)}$, which represents a path between the point $t_{(i)}$ and $t_{(j)}$.

Definition 6. Control node $V_{(i)}$ in $C_r(t_{(i)}, t_{(j)})$. In $C_r(t_{(i)}, t_{(j)})$, the road node that the tourists pass is defined as the control node $V_{(i)}$. The set $\{V_{(i)}\}$ determines the path in $C_r(t_{(i)}, t_{(j)})$. Under different transportation modes, the shortest path in the same $C_r(t_{(i)}, t_{(j)})$ may pass through different control nodes $V_{(i)}$.

Definition 7. Cockroach crawling sub-interval hindrance factor $\zeta_{(i)}$. Tourists traveling in $C_r(t_{(i)}, t_{(j)})$ will be influenced by the transportation mode, travel distance, travel time and travel fee. A factor that will influence the travel costs is defined as the cockroach crawling sub-interval hindrance factor $\zeta_{(i)}$. Set the travel fee in the sub-interval as $F(t_{(i)}, t_{(j)})$, the travel distance as $D(t_{(i)}, t_{(j)})$, the travel time as $T(t_{(i)}, t_{(j)})$; the calculation method for the sub-interval hindrance factor $\zeta_{(i)}$ is designed as Formula (15).

$$\zeta_{(1)} = \frac{1}{F(t_{(i)}, t_{(j)})}, \zeta_{(2)} = \frac{1}{D(t_{(i)}, t_{(j)})}, \zeta_{(3)} = \frac{1}{T(t_{(i)}, t_{(j)})} \quad (15)$$

Definition 8. Cockroach crawling sub-interval hindrance function $f(t_{(i)}, t_{(j)})_{(k)}$ and cockroach crawling interval hindrance function $f(S, t_{(x)})$. The travel cost function generated in the process of tourists' traveling from the point $t_{(i)}$ to $t_{(j)}$ in the sub-interval $C_r(t_{(i)}, t_{(j)})$ is defined as the cockroach crawling sub-interval hindrance function $f(t_{(i)}, t_{(j)})_{(k)}$. k represents the code for the current sub-interval, $0 < k \leq w$, $k, w \in \mathbf{N}$; w is the number of POIs. The travel cost function generated in the process of tourists' traveling from the initial point S to the terminal POI $t_{(x)}$ in

$C_r(S, t_{(x)})$ is defined as the cockroach crawling interval hindrance function $f(S, t_{(x)})$. The functions $f(t_{(i)}, t_{(j)})$ and $f(S, t_{(x)})$ are constructed as Formulas (16) and (17).

$$f(t_{(i)}, t_{(j)})_{(k)} = \sum_{i=1}^{\max i} \zeta_{(i)} \tag{16}$$

$$f(S, t_{(x)}) = \sum_{k=1}^w f(t_{(i)}, t_{(j)})_{(k)} \tag{17}$$

Definition 9. Cockroach crawling interval step size $S_{tep}(t_{(i)}, t_{(j)})$ and cockroach crawling sub-interval step size $S_{tep}(V_{(i)}, V_{(j)})$. In $C_r(S, t_{(x)})$, a cockroach crawling for one time is called the cockroach crawling interval step size $S_{tep}(t_{(i)}, t_{(j)})$. In $C_r(t_{(i)}, t_{(j)})$, a cockroach crawling for one time is called cockroach crawling sub-interval step size $S_{tep}(V_{(i)}, V_{(j)})$. A step size performs a 2-opt operation.

Definition 10. Cockroach crawling interval path $P_{ath}(S, t_{(x)})$ and cockroach crawling sub-interval path $P_{ath}(t_{(i)}, t_{(j)})$. The path formed by a cockroach crawling for p times of step size $S_{tep}(t_{(i)}, t_{(j)})$ is defined as the cockroach crawling interval path $P_{ath}(S, t_{(x)})$. The path formed by a cockroach crawling for q times of step size $S_{tep}(V_{(i)}, V_{(j)})$ is defined as the cockroach crawling sub-interval path $P_{ath}(t_{(i)}, t_{(j)})$. Formulas (18) and (19) are the calculation methods for $P_{ath}(S, t_{(x)})$ and $P_{ath}(t_{(i)}, t_{(j)})$.

$$P_{ath}(S, t_{(x)}) = \sum_{u=1}^p S_{tep}(t_{(i)}, t_{(j)})_{(u)} \tag{18}$$

$$P_{ath}(t_{(i)}, t_{(j)}) = \sum_{u=1}^q S_{tep}(V_{(i)}, V_{(j)})_{(u)} \tag{19}$$

3.2.2. Modeling Process

The ICSOA is set up as follows.

Step 1. As to the initial point S and the w number of POIs $t_{(i)}$, form w number of $C_r(t_{(i)}, t_{(j)})$.

Step 2. Confirm the g number of control nodes $V_{(i)}$ in the sub-interval $C_r(t_{(i)}, t_{(j)})$.

Step 3. Search for the optimal solution in the $C_r(t_{(i)}, t_{(j)})$ and output optimal $D(t_{(i)}, t_{(j)})$ in each $C_r(t_{(i)}, t_{(j)})$.

- (1) Initialize the cockroach individuals $Z_{(a,h)}$.
 - (i) Initialize the node distances $d = \{V_{(i)}, V_{(j)}\}, 0 < i, j \leq g, i, j, g \in \mathbf{N}$.
 - (ii) Search $g_{(i)}$ number of nodes $V_{(i)}$ from $t_{(i)}$ to $t_{(j)}$ in $C_r(t_{(i)}, t_{(j)})$, $0 < g_{(i)} \leq g$.
 - (iii) Initialize $Z_{(a,h)}, D(t_{(i)}, t_{(j)})_{(h)} (h = 1, 2, 3, \dots, m)$.
 - (iv) Randomly select cockroach $\text{Rand}(Z_{(a,h)}) \sim Z_{(a,h1)}$ as a target solution.
- (2) All the $Z_{(a,h)}$ crawl to $Z_{(a,h1)}$ by $S_{tep}(V_{(i)}, V_{(j)})$.
 - (i) The path is $P_{ath}(t_{(i)}, t_{(j)})_{(h)}$.
 - (ii) When the $m - 1$ number of cockroaches crawl to $Z_{(a,h1)}$, if there exists a better $D(t_{(i)}, t_{(j)})_{(h)}$ relating to $Z_{(a,h2)}$, replace $Z_{(a,h1)}$ by $Z_{(a,h2)}$.
- (3) All the cockroaches go back to nest, namely the initial status.
 - (i) Set $Z_{(a,h2)}$ as a new target solution.
 - (ii) All the $Z_{(a,h)}$ crawl to $Z_{(a,h2)}$ by $S_{tep}(V_{(i)}, V_{(j)})$.
 - (iii) The path is $P_{ath}(t_{(i)}, t_{(j)})_{(h)}$.

- (iv) When the $m - 1$ number of cockroaches crawl to $Z_{(a,h2)}$, if there exists a better $D(t_{(i)}, t_{(j)})_{(h)}$ relating to $Z_{(a,h3)}$, replace $Z_{(a,h2)}$ by $Z_{(a,h3)}$.
- (4) Repeat Step 3(1)–Step 3(3) until reach the status when cockroaches crawl to and reach the current optimal $Z_{(a,h)\text{opt}}$; the searched optimal solution $Z_{(a,h)\text{opt}}$ is always unchanging. The searching ends.
- (5) Output the stable optimal $Z_{(a,h)\text{opt}}$, and the related $D(t_{(i)}, t_{(j)})_{(h)}$ is the travel distance for $C_r(t_{(i)}, t_{(j)})$.

Step 4. Based on the $D(t_{(i)}, t_{(j)})_{(h)}$, calculate the hindrance factor $\zeta_{(i)}$ of $C_r(t_{(i)}, t_{(j)})$ and hindrance function value $f(t_{(i)}, t_{(j)})_{(k)}$, k is the code for the current sub-interval.

Step 5. Search the hindrance function value $f(S, t_{(x)})$ relating to the optimal solution for $C_r(S, t_{(x)})$.

- (1) Initialize the cockroach individuals $Z_{(b,s)}$.
 - (i) Initialize the hindrance function value $f(t_{(i)}, t_{(j)})_{(k)}$ for the sub-intervals, $0 < k \leq w$, $k, w \in \mathbf{N}$.
 - (ii) Initialize $Z_{(b,s)}(s = 1, 2, 3, \dots, n)$, $f(S, t_{(x)})_{(h)}(h = 1, 2, 3, \dots, n)$.
 - (iii) Randomly select cockroach $\text{Rand}(Z_{(b,s)}) \sim Z_{(b,s1)}$ as the target solution.
- (2) All the $Z_{(b,s)}$ crawl to $Z_{(b,s1)}$ by $S_{\text{tep}}(t_{(i)}, t_{(j)})$.
 - (i) The path is $P_{\text{ath}}(S, t_{(x)})_{(h)}$.
 - (ii) When the $n - 1$ number of cockroaches crawl to $Z_{(b,s1)}$, if there exists a better $f(S, t_{(x)})_{(h)}$ relating to $Z_{(b,s2)}$, replace the $Z_{(b,s1)}$ by $Z_{(b,s2)}$.
- (3) All the cockroaches go back to nest, namely the initial status.
 - (i) Set $Z_{(b,s2)}$ as a new target solution.
 - (ii) All the $Z_{(b,s)}$ crawl to $Z_{(b,s2)}$ by $S_{\text{tep}}(t_{(i)}, t_{(j)})$.
 - (iii) The path is $P_{\text{ath}}(S, t_{(x)})_{(h)}$.
 - (iv) When the $n - 1$ number of cockroaches crawl to $Z_{(b,s2)}$, if there exists a better $f(S, t_{(x)})_{(h)}$ relating to $Z_{(b,s3)}$, replace the $Z_{(b,s2)}$ by $Z_{(b,s3)}$.
- (4) Repeat Step 5(1)–Step 5(3) until get to the status when cockroaches crawl to and reach the current optimal $Z_{(b,s)\text{opt}}$, the searched optimal solution $Z_{(b,s)\text{opt}}$ is always unchanging. The searching process ends.
- (5) Output the optimal $Z_{(b,s)\text{opt}}$. The related $f(S, t_{(x)})_{(h)}$ stands for the lowest travel costs for $C_r(S, t_{(x)})$.

Step 6. Output the route with the optimal hindrance function value $f(S, t_{(x)})_{(h)}$ as the optimal POI route for recommendation.

Figure 4 is the constructed POI route recommendation algorithm based on ICSOA. Figure 4A shows the specific algorithm steps for the sub-interval, Figure 4B shows the specific algorithm steps for the interval, Figure 4C shows the flow chart of the constructed algorithm.

3.3. The Analysis on the Algorithm

The proposed POI route recommendation model based on ICSOA consists of the NBCA, the SAFS model and the CSOA model. Among them, the NBCA and the SAFS model are the foundation. Based on the modeling ideas, the key factors influencing algorithm accuracy and efficiency are analyzed as follows.

- (1) The improved NBCA aims to mine the tourists' interests based on the POIs they have visited before and classify the POIs in the destination city based on their interests. Therefore, the accuracy of classification depends on the selection of feature labels in the training set and the quantified range values of the feature labels. The closer the feature labels are to the tourists' interests, the more accurate the range values of

feature labels calculated by text mining, and the more accurate the final classification results will be, which will better match the tourists' interests. Meanwhile, the selected feature labels must be independent to ensure the stability of the algorithm. The NBCA has a very stable efficiency, simple execution mechanism and low time complexity, and is suitable for the small-scale data classification. In the process of algorithm modeling, the collected training sets and the POI sets to be classified are all small-scale datasets, so the algorithm has high computational efficiency.

- (2) The goal of the SAFS is to find out the optimal POIs from the already classified ones. The factors that determine the accuracy of the SAFS model are the coordinates of the starting points and the POIs. The spatial accessibility and feature attributes have a decisive impact on the accuracy of POI recommendation. We define weights $\varepsilon_{(1)}$ and $\varepsilon_{(2)}$, representing tourists' preferences for spatial accessibilities and feature attributes. The higher the weight is, the more emphasis the tourists will lay on certain conditions, which directly affects the final recommendation results.
- (3) The ICSOA generates new individuals by changing the step size. The algorithm searching for the optimal route includes two parts: interval searching and global searching. When searching between two POIs, the key factor affecting the optimal path is the distance between road nodes, which determines the travel time and cost, and ultimately determines the individual costs of cockroaches. The key factor affecting the optimal route in global searching is the travel costs between POIs. Therefore, determining road nodes and distances, as well as the cost of path between POIs, is the key to the algorithm accuracy. The ICSOA adopts the 2-opt idea when searching for individual costs of cockroaches. The time complexity of searching for the better solutions that replace previous cockroach individuals is not higher than that of factorial operation; thus, the computational speed is very fast. The algorithm constantly replaces cockroach individuals and searches for the global optimal solution; thus, it has a time complexity not higher than that of the sorting algorithm, and the computational speed is also very fast. Therefore, the key factors affecting algorithm efficiency are the quantity of road nodes and POIs. Since the set of road nodes between two POIs and the set of POIs to be visited are both small datasets, the algorithm convergence speed will be very fast. It can quickly output the global optimal solution and recommend the optimal POI route for tourists.

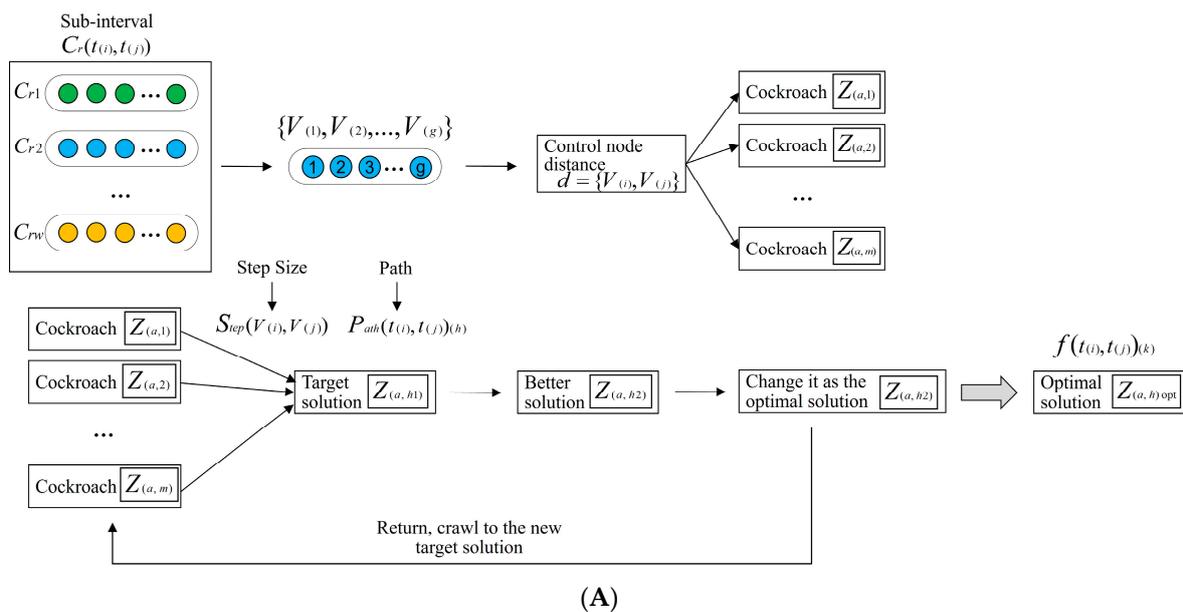
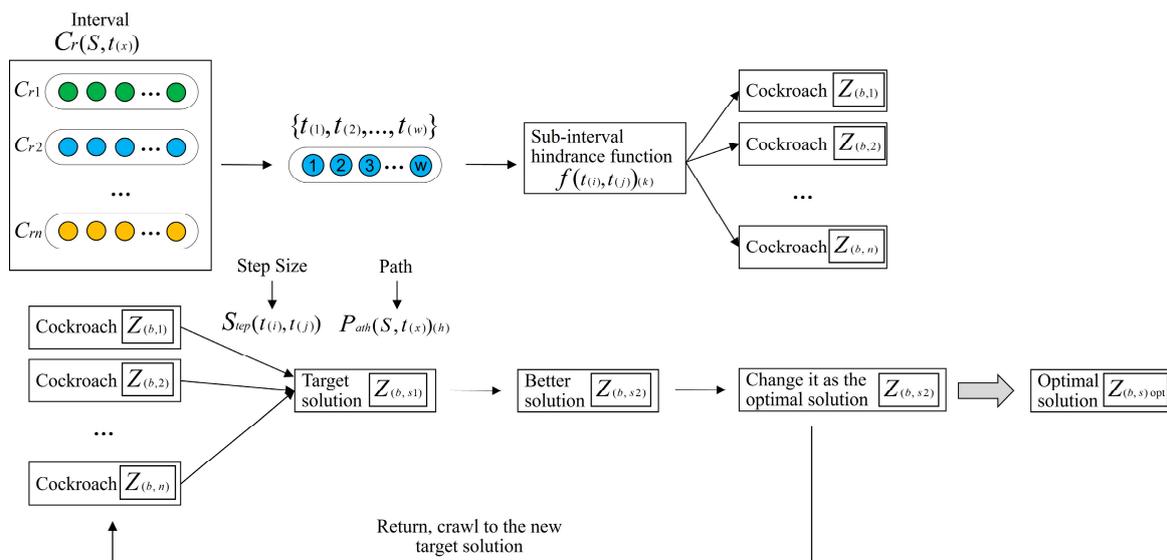
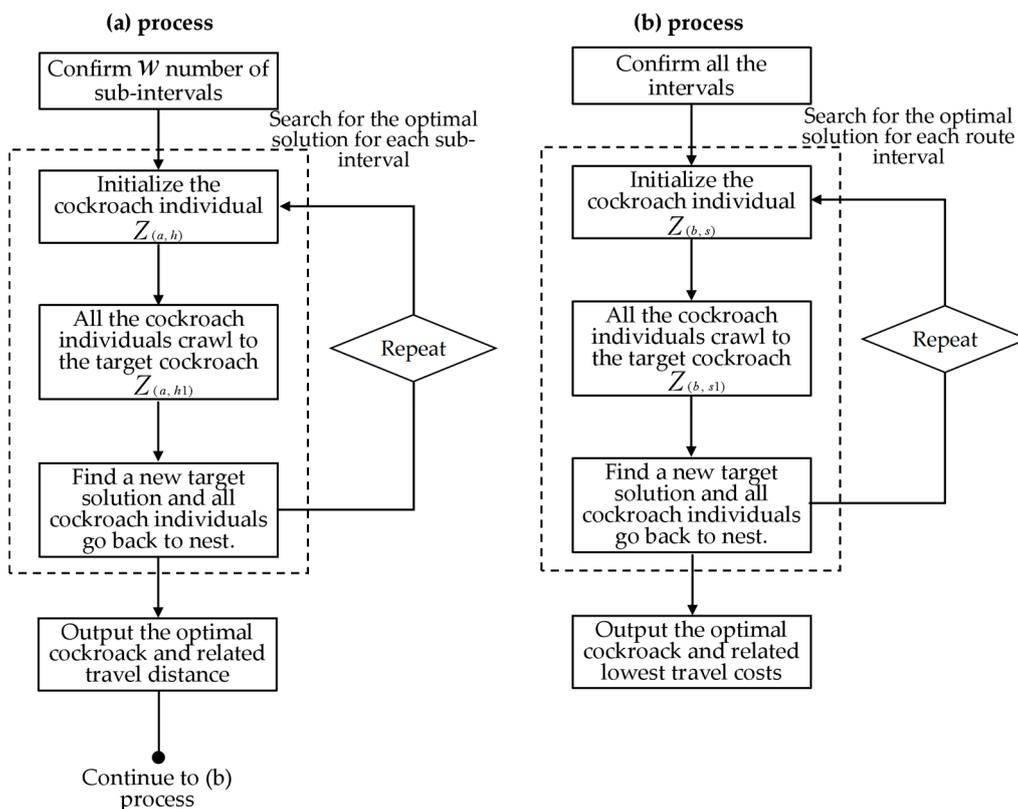


Figure 4. Cont.



(B)



(C)

Figure 4. The constructed POI route recommendation algorithm based on ICSOA. (A) shows the specific algorithm steps for the sub-interval, (B) shows the specific algorithm steps for the interval, (C) shows the flow chart of the constructed algorithm. In (A,B), the different colors in sub-interval and interval mean the cockroaches, and one color means one cockroach. The blue color in $V_{(i)}$ and $t_{(i)}$ means the control node of the sub-interval and interval.

4. Experiment and Result Analysis

We designed an experiment to testify the proposed algorithm. The experimental area is the downtown of Chengdu. The experimental subjects are two visitors who come to Chengdu. The experiment studies the algorithm in five aspects: (1) Naive Bayes classification results; (2) SAFS; (3) POI recommendation results; (4) POI route recommendation results; (5) the comparison with the commonly used route plan methods.

4.1. Experimental Conditions and Materials

- (1) Tourist A and Tourist B both provide 10 once-visited POIs and make judgements on preferences, $C_{(i)}$: $\{C_{(1)}$: favorite; $C_{(2)}$: like; $C_{(3)}$: dislike $\}$.
- (2) Tourist A starts the trip from "Tianfu Square"; Tourist B starts the trip from "Chengdu Railway Station".
- (3) Choose 20 POIs in Chengdu. The feature attributes are $l_{(1)}$: natural view; $l_{(2)}$: humanity and history; $l_{(3)}$: consumption and shopping; $l_{(4)}$: amusement and sports. From tourism big data, obtain sub-labels $l_{(i,j)}$ for each label $l_{(i)}$.
- (4) Take 50 text documents to calculate weight $TFIDF_{l_{(i)}}$. The tourism attributes are $l_{(5)}$: entrance ticket fee (¥ CNY); $l_{(6)}$: visiting time (hour); $l_{(7)}$: attraction index.

Table 1 shows the results of weights $TFIDF_{l_{(i)}}$ and tourism attribute weights on the POIs provided by Tourists A and B. Table 2 shows the designed value ranges for NBCA. In the table, $T_{A(1)}$: The Summer Palace, $T_{A(2)}$: Disneyland, $T_{A(3)}$: Hangzhou West Lake, $T_{A(4)}$: The Palace Museum, $T_{A(5)}$: Yellow Crane Tower, $T_{A(6)}$: Juzizhou Island, $T_{A(7)}$: Tang Paradise, $T_{A(8)}$: Dianchi Lake, $T_{A(9)}$: Pingyao Ancient City, $T_{A(10)}$: Gulangyu Islet, $T_{B(1)}$: Yuyan Garden, $T_{B(2)}$: Wuhan East Lake, $T_{B(3)}$: Yanta Square, $T_{B(4)}$: Huangguoshu Waterfall, $T_{B(5)}$: Shaanxi Museum, $T_{B(6)}$: The Imperial Palace of Shenyang, $T_{B(7)}$: Hangzhou West Lake, $T_{B(8)}$: The Longmen Grottoes, $T_{B(9)}$: Tengwang Pavilion, $T_{B(10)}$: Leshan Grand Buddha.

Table 1. The calculation results of $TFIDF_{l_{(i)}}$ and tourism attribute weights on the POIs.

T-A	$T_{A(1)}$	$T_{A(2)}$	$T_{A(3)}$	$T_{A(4)}$	$T_{A(5)}$	$T_{A(6)}$	$T_{A(7)}$	$T_{A(8)}$	$T_{A(9)}$	$T_{A(10)}$
$l_{(1)}$	0.013	0.047	0.011	0.028	0.038	0.026	0.075	0.009	0.071	0.013
$l_{(2)}$	0.030	0.069	0.014	0.018	0.050	0.059	0.103	0.099	0.025	0.075
$l_{(3)}$	0.079	0.048	0.056	0.076	0.056	0.052	0.056	0.067	0.039	0.070
$l_{(4)}$	0.040	0.020	0.037	0.033	0.048	0.041	0.050	0.031	0.029	0.047
$l_{(5)}$	30.00	475.00	0	60.00	80.00	0	120.00	0	130.00	90.00
$l_{(6)}$	3.00	8.00	5.00	4.00	3.00	3.00	3.00	2.00	3.00	3.00
$l_{(7)}$	0.94	0.94	0.94	0.96	0.92	0.92	0.88	0.92	0.90	0.88
$C_{(i)}$	$C_{(1)}$	$C_{(3)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$	$C_{(3)}$	$C_{(1)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$
T-B	$T_{B(1)}$	$T_{B(2)}$	$T_{B(3)}$	$T_{B(4)}$	$T_{B(5)}$	$T_{B(6)}$	$T_{B(7)}$	$T_{B(8)}$	$T_{B(9)}$	$T_{B(10)}$
$l_{(1)}$	0.033	0.026	0.030	0.018	0.043	0.058	0.011	0.030	0.036	0.021
$l_{(2)}$	0.042	0.046	0.026	0.078	0.013	0.024	0.014	0.021	0.023	0.013
$l_{(3)}$	0.138	0.065	0.063	0.054	0.064	0.068	0.056	0.037	0.036	0.051
$l_{(4)}$	0.042	0.086	0.037	0.030	0.021	0.051	0.037	0.033	0.039	0.032
$l_{(5)}$	40.00	0	0	160.00	0	50.00	0	120.00	50.00	80.00
$l_{(6)}$	2.00	4.00	2.00	4.00	3.00	2.00	5.00	4.00	2.00	3.00
$l_{(7)}$	0.94	0.94	0.90	0.90	0.92	0.92	0.94	0.94	0.92	0.90
$C_{(i)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$	$C_{(1)}$	$C_{(3)}$	$C_{(3)}$	$C_{(1)}$	$C_{(2)}$	$C_{(2)}$	$C_{(1)}$

Table 2. The designed value ranges for the proposed classification algorithm.

	Value Range 1	Value Range 2	Value Range 3	Value Range 4	Value Range 5
$l_{(1)} \sim l_{(4)}$	$0 < l_{(i)} \leq 0.025$	$0.025 < l_{(i)} \leq 0.05$	$0.05 < l_{(1)} \leq 0.075$	$0.075 < l_{(1)} \leq 0.100$	$l_{(1)} > 0.100$
$l_{(5)}$	$0 \leq l_{(1)} \leq 50$	$50 < l_{(1)} \leq 100$	$100 < l_{(1)} \leq 150$	$150 < l_{(1)} \leq 200$	$l_{(1)} > 200$
$l_{(6)}$	$0 < l_{(1)} \leq 1$	$1 < l_{(1)} \leq 3$	$3 < l_{(1)} \leq 5$	$l_{(1)} > 5$	
$l_{(7)}$	$l_{(1)} \leq 0.90$	$0.90 < l_{(1)} \leq 0.92$	$0.92 < l_{(1)} \leq 0.94$	$0.94 < l_{(1)} \leq 0.96$	$l_{(1)} > 0.96$

Chengdu POIs are: $T_{u(1)}$: Temple of Marquis, $T_{u(2)}$: Kuanzhai Alley, $T_{u(3)}$: The People’s Park $T_{u(4)}$: Du Fu Thatched Cottage, $T_{u(5)}$: Chunxi Road, $T_{u(6)}$: Donghu Park, $T_{u(7)}$: Chengdu Happy Valley, $T_{u(8)}$: Panda Base, $T_{u(9)}$: Eastern Suburb Memory, $T_{u(10)}$: Qinglong Lake Wetland, $T_{u(11)}$: Jincheng Lake, $T_{u(12)}$: San Sheng Flower Town, $T_{u(13)}$: Wenshu Temple, $T_{u(14)}$: Qingyang Palace, $T_{u(15)}$: Jinsha site, $T_{u(16)}$: Jinli Street, $T_{u(17)}$: Sichuan Museum, $T_{u(18)}$: Tazishan Park, $T_{u(19)}$: Fenghuang Mountain, $T_{u(20)}$: Chengdu Zoo.

4.2. Experimental Results and Analysis

4.2.1. Classification Results and Analysis

(1) POI attribute label weight and classification results

Calculated by the improved NBCA, Chengdu POIs are grouped into three categories by tourists’ interests. Table 3 shows the feature attribute weights $TFIDF_{l_{(i)}}$ and tourism attribute weights. Table 4 shows the calculated Bayes conditional probability and classification results.

Table 3. The calculated $TFIDF_{l_{(i)}}$ and tourism attribute weights of the experimental POIs.

	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$l_{(1)}$	0.036	0.026	0.048	0.030	0.011	0.050	0.063	0.173	0.041	0.036
$l_{(2)}$	0.034	0.054	0.080	0.025	0.056	0.115	0.068	0.150	0.069	0.083
$l_{(3)}$	0.073	0.101	0.071	0.049	0.041	0.098	0.055	0.143	0.075	0.052
$l_{(4)}$	0.036	0.064	0.061	0.043	0.069	0.053	0.032	0.095	0.049	0.050
$l_{(5)}$	50.00	0	0	50.00	0	0	230.00	55.00	0	0
$l_{(6)}$	3.00	3.00	3.00	3.00	2.00	1.00	8.00	3.00	2.00	2.00
$l_{(7)}$	0.90	0.90	0.92	0.92	0.92	0.94	0.92	0.92	0.90	0.92
	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$
$l_{(1)}$	0.051	0.029	0.071	0.093	0.079	0.053	0.034	0.036	0.053	0.121
$l_{(2)}$	0.058	0.089	0.049	0.041	0.051	0.039	0.032	0.035	0.054	0.088
$l_{(3)}$	0.057	0.074	0.058	0.041	0.066	0.036	0.036	0.064	0.048	0.089
$l_{(4)}$	0.080	0.075	0.027	0.037	0.023	0.059	0.034	0.089	0.064	0.101
$l_{(5)}$	0	0	0	10.00	70.00	0	0	0	0	20.00
$l_{(6)}$	3.00	6.00	2.00	2.00	2.00	3.00	2.00	3.00	3.00	4.00
$l_{(7)}$	0.90	0.86	0.92	0.90	0.94	0.90	0.92	0.90	0.90	0.94

Table 4. Bayes conditional probability and classification results of the experimental POIs.

T-A	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$\delta_{(C_{(1)})}$	0.2373	0.0352	0.0264	0.1055	0.0176	0.0396	0.0281	0.0176	0.1582	0.0791
$\delta_{(C_{(2)})}$	0.4746	0.0527	0.2109	0.2109	0.1172	0.0141	0.0844	0.0313	0.4746	0.6328
$\delta_{(C_{(3)})}$	0.0750	0.0563	0.0750	0.1500	0.0750	0.0188	0.1500	0.0063	0.2250	0.1500
$C_{(i)}$	$C_{(2)}$	$C_{(3)}$	$C_{(2)}$	$C_{(2)}$	$C_{(2)}$	$C_{(1)}$	$C_{(3)}$	$C_{(2)}$	$C_{(2)}$	$C_{(2)}$
T-A	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$

Table 4. Cont.

$\delta_{(C_{(1)})}$	0.0703	0.0211	0.1582	0.1055	0.0264	0.0703	0.0791	0.0791	0.0469	0.0264
$\delta_{(C_{(2)})}$	0.1055	0.0633	0.4219	0.1055	0.1055	0.0352	0.2109	0.1582	0.0352	0.0211
$\delta_{(C_{(3)})}$	0.0375	0.0375	0.0500	0.0250	0.0750	0.0125	0.1500	0.0375	0.0375	0.0063
$C_{(i)}$	$C_{(2)}$	$C_{(2)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$	$C_{(2)}$	$C_{(1)}$	$C_{(1)}$
T-B	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$\delta_{(C_{(1)})}$	0.3164	0.0281	0.0633	0.0703	0.0234	0.0188	0.0234	0.0105	0.2109	0.1582
$\delta_{(C_{(2)})}$	0.7910	0.1318	0.1318	1.0547	0.0586	0.0469	0.0211	0.0178	0.3955	0.3955
$\delta_{(C_{(3)})}$	0.0844	0.0563	0.5063	0.2531	0.0844	0.0188	0.0108	0.0141	0.0970	0.2531
$C_{(i)}$	$C_{(2)}$	$C_{(2)}$	$C_{(3)}$	$C_{(2)}$	$C_{(3)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$	$C_{(2)}$	$C_{(2)}$
T-B	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$
$\delta_{(C_{(1)})}$	0.0844	0.0844	0.1055	0.0703	0.0422	0.0281	0.0527	0.1898	0.0188	0.0469
$\delta_{(C_{(2)})}$	0.0439	0.0527	0.2637	0.3516	0.0234	0.1172	1.0547	0.2637	0.0586	0.0234
$\delta_{(C_{(3)})}$	0.0844	0.0647	0.2531	0.0141	0.0323	0.0647	0.0844	0.0970	0.0563	0.0054
$C_{(i)}$	$C_{(1)}$	$C_{(1)}$	$C_{(2)}$	$C_{(2)}$	$C_{(1)}$	$C_{(2)}$	$C_{(2)}$	$C_{(2)}$	$C_{(2)}$	$C_{(1)}$

If the probability meets $\delta_{(y_{(i)})} = P(y_{(i)} | C_{(i)})P(C_{(i)})$, the disturbance factor is taken as $\varepsilon = 0.5$.

(2) The analysis on the classification results

(i) Analyze the Table 4 results.

For Tourist A, POIs are divided into:

- ① $\{C_{(1)}$: favorite}, including the Donghu Park, the Qingyang Palace, the Jinli Street, the Fenghuang Mountain and the Chengdu Zoo;
- ② $\{C_{(2)}$: like}, including the Wuhou Temple, the People’s Park, the Du Fu Thatched Cottage, the Chunxi Road, the Panda Base, the Eastern Suburb Memory, the Qinglong Lake Wetland, the Jincheng Lake, the Sansheng Flower Town, the Wenshu Temple site, the Sichuan Museum and the Tazishan Park;
- ③ $\{C_{(3)}$: dislike}, including the Kuanzhai Alley and the Chengdu Happy Valley.

For Tourist B, POIs are divided into:

- ① $\{C_{(1)}$: favorite}, including the Chengdu Happy Valley, the Jincheng Lake, the Sansheng Flower Town, the Jinsha Site and the Chengdu Zoo;
- ② $\{C_{(2)}$: like}, including the Wuhou Temple, the Kuanzhai Alley, the Du Fu Thatched Cottage, the East Lake Park, the Panda Base, the Eastern Suburb Memory, the Qinglong Lake Wetland, the Wenshu Temple, the Qingyang Palace, the Jinli Street, the Sichuan Museum, the Tazishan Park and the Fenghuang Mountain;
- ③ $\{C_{(3)}$: dislike}, including the People’s Park and the Chunxi Road.

It demonstrates that the proposed algorithm has good classification performance.

- (ii) For each POI to be classified, the category corresponding to the maximum Bayes conditional probability is taken as the category of the POI, which indicates that the feature attributes and tourism attributes of the POI are closest to the attributes of this category provided by the tourists. It proves that the proposed algorithm obtains the tourists’ interests and divides the POIs into the categories provided by the tourists based on individualized interests. It excludes the uninteresting POIs while recommend the interesting ones.
- (iii) When different tourists provide different interests, it will produce discrepant results. It proves that the proposed algorithm has the feature of universality, and the classification results are based on the individualized interests of tourists, which can recommend exclusive POIs for tourists.

4.2.2. Spatial Accessibility Calculation Results and Analysis

(1) Spatial accessibility calculation results

The initial points for Tourists A and B are: S_A : Tianfu Square; S_B : Chengdu Railway Station. The categories are: $C_{(1)}$ "favorite", $C_{(2)}$ "like" and $C_{(3)}$ "dislike".

The absolute and relative SAFS of $A_{(T_{u(i)},S)}$ and $A^*_{(T_{u(i)},S)}$ are shown in Table 5 and visualized in Figure 5. Figure 5a–c are the SAFS classified in $C_{(1)}, C_{(2)}$ and $C_{(3)}$ for Tourist A. Figure 5d–f are the SAFS classified in $C_{(1)}, C_{(2)}$ and $C_{(3)}$ for Tourist B. In the curves, the category $C_{(1)}, C_{(2)}$ and $C_{(3)}$ are noted in red, orange and blue circles.

Table 5. The calculated $A_{(T_{u(i)},S)}$ and $A^*_{(T_{u(i)},S)}$ for each category POI.

T-A	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$A_{(T_{u(i)},S_A)}$	0.4762	0.7143	1.2658	0.2778	0.6667	0.1961	0.1298	0.0862	0.1786	0.0806
$A^*_{(T_{u(i)},S_A)}$	0.0726	0.1089	0.1930	0.0424	0.1016	0.0299	0.0198	0.0131	0.0272	0.0123
T-A	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$
$A_{(T_{u(i)},S_A)}$	0.1031	0.0862	0.4762	0.4348	0.1724	0.4762	0.3226	0.1639	0.1087	0.1429
$A^*_{(T_{u(i)},S_A)}$	0.0157	0.0131	0.0726	0.0663	0.0263	0.0726	0.0492	0.0250	0.0166	0.0218
T-B	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$A_{(T_{u(i)},S_B)}$	0.1667	0.2439	0.2222	0.1695	0.2128	0.1124	0.2128	0.1235	0.1786	0.0741
$A^*_{(T_{u(i)},S_B)}$	0.0459	0.0671	0.0611	0.0466	0.0585	0.0309	0.0585	0.0340	0.0491	0.0204
T-B	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$
$A_{(T_{u(i)},S_B)}$	0.0714	0.0662	0.4348	0.2083	0.1667	0.1667	0.1852	0.1174	0.2083	0.2941
$A^*_{(T_{u(i)},S_B)}$	0.0196	0.0182	0.1196	0.0573	0.0459	0.0459	0.0509	0.0323	0.0573	0.0809

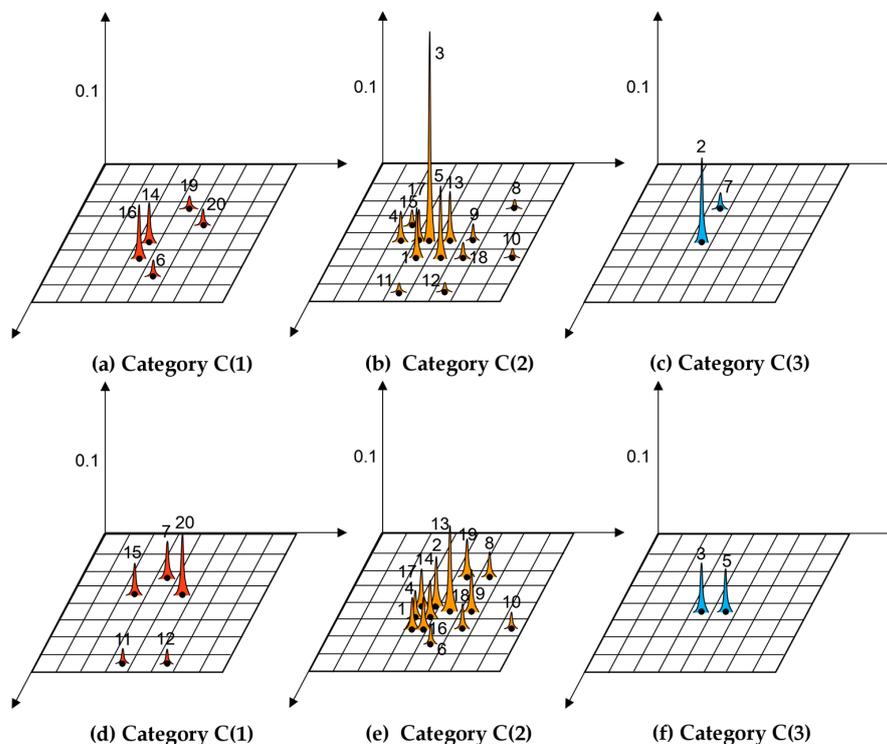


Figure 5. The visualized SAFS for the three categories. In the figure, different colors mean different categories. Red represents the category $C_{(1)}$, Orange represents the category $C_{(2)}$, Blue represents the category $C_{(3)}$.

- (2) Analysis on the spatial accessibility
 - (i) As to Table 5, different spatial attributes of POIs cause different travel costs: the weaker the SAFS is, the higher the travel costs will be for the tourists to reach the POI.
 - ① For Tourist A, in the “favorite” category, Jinli Street has the highest spatial field strengths (SFS) of 0.4762 and 0.0726, and Fenghuang Mountain has the lowest SFSs of 0.1087 and 0.0166. Under the same conditions of interests, the travel costs to Jinli Street are the lowest; in the “like” category, The People’s Park has the highest SFSs of 1.2658 and 0.1930, the Qinglong Lake Wetland has the lowest SFS of 0.0806 and 0.0123. In the same interests, the travel costs to the People’s Park is the lowest.
 - ② For Tourist B, in the “favorite” category, the Chengdu Zoo has the highest SFSs of 0.2941 and 0.0809, the Sansheng flower Town has the lowest SFSs of 0.0662 and 0.0182. In the same interests, the travel costs to Chengdu Zoo are the lowest; in the “like” category, the Wenshu Temple has the highest SFSs of 0.4348 and 0.1196, the Qinglong Lake Wetland has the lowest SFSs of 0.0741 and 0.0204. In the same interests, the travel costs to Wenshu Temple are the lowest. POIs in the “dislike” category are not recommended.
 - (ii) Analyze Figure 5. According to the SAFS model, each POI is distributed in the unit grid $g_{(x,y)}$ of the spatial grid G formed by the urban geographical space. The position (x, y) in the grid G represents the coordinate position in the urban geographical space. The SFS distribution of POIs relative to the initial points of Tourist A and B, Tianfu Square and Chengdu Railway Station, are totally discrepant. The higher the peak value is, the stronger the spatial accessibility will be.
 - (iii) The experimental results verify that the proposed algorithm can not only output POIs with attributes matching tourists’ interests, but also output POIs with the best spatial distribution and the lowest travel costs.

4.2.3. POI Recommendation Results and Analysis

- (1) POI recommendation results
 - (i) Based on data in Tables 4 and 5, the recommendation index $\sigma_{(T_{u(i)})}$ is calculated. Tourist A lays the same importance on POI attribute and spatial accessibility; set $\epsilon_{(1)} = 0.5, \epsilon_{(2)} = 0.5$. Tourist B lays more emphasis on the travel costs; set $\epsilon_{(1)} = 0.3, \epsilon_{(2)} = 0.7$.
 - ① Table 6 shows the calculation results of the recommendation index $\sigma_{(T_{u(i)})}$.

Table 6. Recommendation index calculation results for the POIs.

T-A	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$\sigma_{(T_{u(i)})}$	0.2736	0.0826	0.2020	0.1267	0.1094	0.0348	0.0849	0.0222	0.2509	0.3226
T-A	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$
$\sigma_{(T_{u(i)})}$	0.0606	0.0382	0.2473	0.0859	0.0659	0.0715	0.1301	0.0916	0.0318	0.0241
T-B	$T_{u(1)}$	$T_{u(2)}$	$T_{u(3)}$	$T_{u(4)}$	$T_{u(5)}$	$T_{u(6)}$	$T_{u(7)}$	$T_{u(8)}$	$T_{u(9)}$	$T_{u(10)}$
$\sigma_{(T_{u(i)})}$	0.2694	0.0865	0.1947	0.3490	0.0663	0.0357	0.0480	0.0291	0.1530	0.1329
T-B	$T_{u(11)}$	$T_{u(12)}$	$T_{u(13)}$	$T_{u(14)}$	$T_{u(15)}$	$T_{u(16)}$	$T_{u(17)}$	$T_{u(18)}$	$T_{u(19)}$	$T_{u(20)}$
$\sigma_{(T_{u(i)})}$	0.0390	0.0381	0.1628	0.1456	0.0448	0.0673	0.3520	0.1017	0.0577	0.0707

② Figure 6 shows the curves of the index $\sigma_{(T_{u(i)})}$. Different colors represent different categories. The red points relate to $C_{(1)}$ “favorite”, the orange points relate to $C_{(2)}$

“like” and the blue points relate to $C_{(3)}$ “dislike”. Figure 6a,b are for Tourists A and B, respectively.

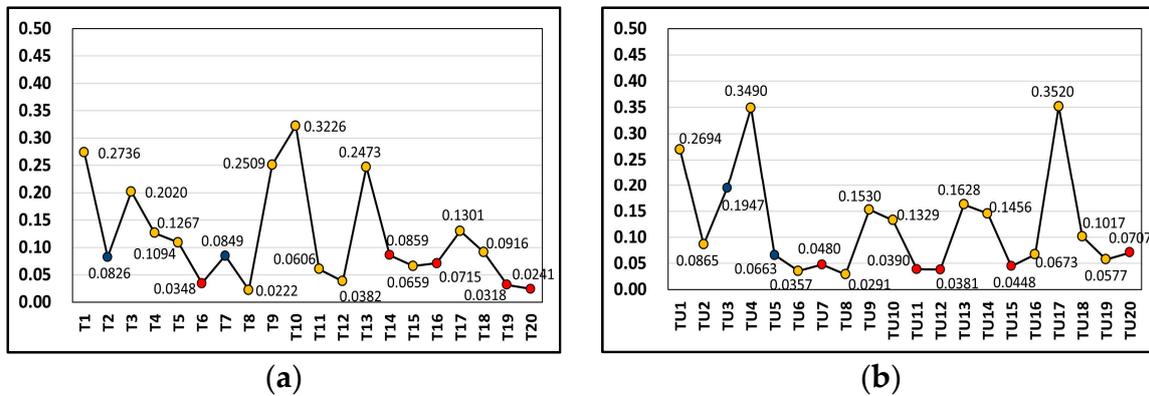


Figure 6. The tendency curve of the recommendation indexes for POIs. (a,b) are Tourists A and B, respectively. Different colors represent different categories. The red points relate to $C_{(1)}$ “favorite”, the orange points relate to $C_{(2)}$ “like” and the blue points relate to $C_{(3)}$ “dislike”.

③ Figure 7 shows the index $\sigma_{(T_{u(i)})}$ distribution for each category, and Figure 7a–c relate to Tourist A and Figure 7d–f relate to Tourist B.

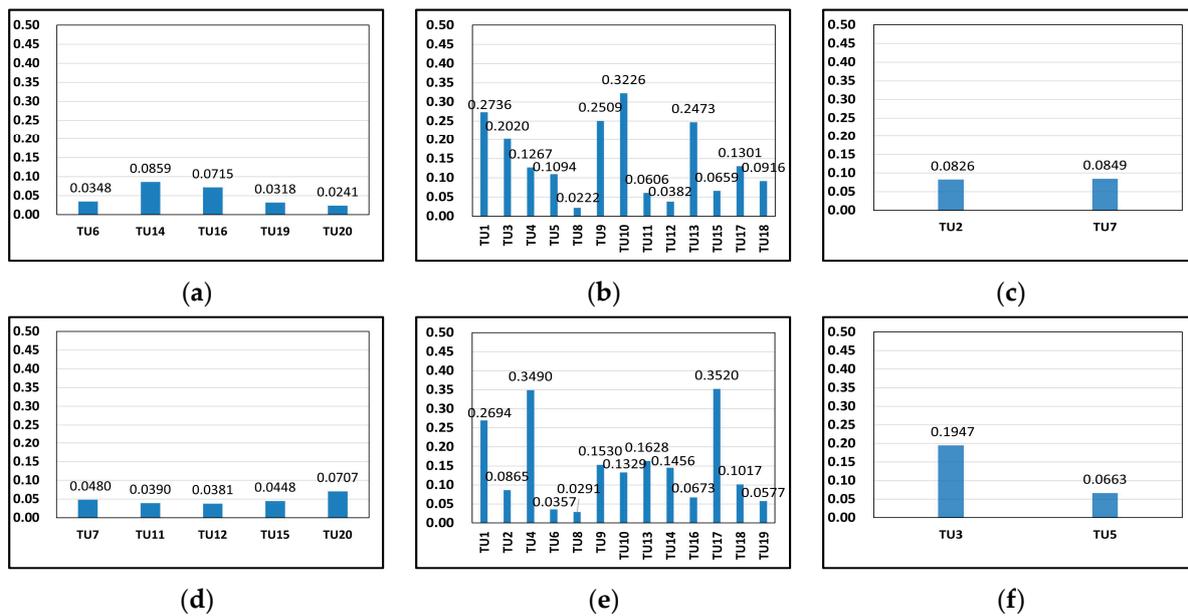


Figure 7. The recommendation index distributions for each category. (a–c) relate to Tourist A and (d–f) relate to Tourist B.

(ii) From Table 6 and the Figures 6 and 7 results, the four recommended POIs for Tourist A are:

- $T_{u(1)}$: Temple of Marquis;
- $T_{u(9)}$: Eastern Suburb Memory;
- $T_{u(10)}$: Qinglong Lake Wetland;
- $T_{u(13)}$: Wenshu Temple.

The four recommended POIs for the Tourist B are:

- $T_{u(1)}$: Temple of Marquis;
- $T_{u(3)}$: The People’s Park;
- $T_{u(4)}$: Du Fu Thatched Cottage;

$T_{u(7)}$: Chengdu Happy Valley.

(2) Analysis on the recommendation results

(i) The demand weight $\varepsilon_{(x)}$ directly affects the recommendation results.

- ① When $\varepsilon_{(1)} > \varepsilon_{(2)}$, tourists lay more emphasis on the attributes;
- ② When $\varepsilon_{(1)} < \varepsilon_{(2)}$, tourists lay more emphasis on the spatial accessibility;
- ③ When $\varepsilon_{(1)} = \varepsilon_{(2)}$, tourists lay equal importance on the two constraint factors.

In this experiment, Tourist A lays equal emphasis on both constraint factors, while Tourist B lays more emphasis on the spatial accessibility. The recommendation results are completely different with different weights $\varepsilon_{(x)}$. It proves that the proposed algorithm is greatly influenced by the subjective needs of individual tourists, and it can recommend POIs that match tourists' interests with the lowest travel costs.

(ii) Analyze Table 6 and Figure 6. The POIs with the highest index $\sigma_{(y_{(i)})}$ are classified into the category "favorite" or "like" and are preferentially recommended. The same POI has different recommendation indexes influenced by the weights $\varepsilon_{(x)}$, and the recommendation degree shows a fluctuating trend.

① As to the weight $\varepsilon_{(x)}$ of Tourist A, the average recommendation index of POIs is 0.1178, with a variance of 0.0079.

② As to the weight $\varepsilon_{(x)}$ of Tourist B, the average recommendation index of POIs is 0.1222, with a variance of 0.0095.

This indicates that the average recommendation index for Tourist B is higher. For the fluctuating trend, the smaller variance of Tourist A makes the curve fluctuation closer to the average recommendation value. It indicates that under the given weight of Tourist A, the recommendation degree is more stable, and the recommendation probability for each POI is more balanced, but for Tourist B, the recommendation degree stability is lower, and the recommendation probability for each POI is more discrepant.

(iii) Analyzing Figure 7, the POIs are grouped into different categories, the recommendation degree for each category is different.

① For Tourist A, the highest recommendation degree in the category "favorite" is 0.0859, while the highest recommendation degree in the category "like" is 0.3226, which is generally higher than that in the category "favorite". Therefore, the POIs in the category "favorite" are recommended for Tourist A.

② For Tourist B, the highest recommendation degree in the category "favorite" is 0.0707, while the highest recommendation degree in the category "like" is 0.3520, which is also generally higher than that in the category "favorite". Therefore, the POIs in the category "favorite" are recommended for Tourist B. The POIs in the category "dislike" will not be recommended.

4.2.4. POI Route Recommendation Results and Analysis

(1) POI route recommendation results

(i) Tourist A chooses electric bicycle; Tourist B chooses taxi. The average speed of the electric bicycle in the downtown area is set as 10 km/h, and the average speed of the taxi in the downtown area is set as 20 km/h. The charging rule for the electric bicycle is: starting price 2 CNY/15 min; exceeding 15 min, 1 CNY/10 min. The charging rule for taxi is: starting price 8 CNY/2 km, exceeding 2 km, 1.9 CNY/km. Initial points are S_A : Tianfu Square and S_B : Chengdu Railway Station.

(ii) Based on the recommended POIs, combined with the geospatial and transportation conditions of Chengdu, the cockroach crawling sub-intervals, cockroach crawling intervals and control nodes within the sub-intervals are confirmed. The ICSOA is used to search for the shortest path within each sub-interval, by which the hindrance factor $\zeta_{(i)}$ and hindrance function $f^{(t_{(i)}, t_{(j)})_{(k)}}$ of the cockroach crawling sub-interval are calculated.

The recommended POIs for Tourist A are noted as:

- $T_{u(1)}$: Temple of Marquis (a_1);
- $T_{u(9)}$: Eastern Suburb Memory (a_2);
- $T_{u(10)}$: Qinglong Lake Wetland (a_3);
- $T_{u(13)}$: Wenshu Temple (a_4).

The recommended POIs for Tourist B are noted as:

- $T_{u(1)}$: Temple of Marquis (b_1);
- $T_{u(3)}$: The People’s Park (b_2);
- $T_{u(4)}$: Du Fu Thatched Cottage (b_3);
- $T_{u(7)}$: Chengdu Happy Valley (b_4).

(iii) Table 7 shows the sub-interval hindrance factors $\zeta_{(i)}$ and hindrance function values $f(t_{(i)}, t_{(j)})_{(k)}$ output by ICSOA. Table 8 shows the cockroach crawling intervals $C_r(S, t_{(x)})$ and hindrance function values $f(S, t_{(x)})$ output by ICSOA. a_{1234} represents the cockroach crawling interval formed by Tourist A starting from S_A , followed by Temple of Marquis (a_1), Eastern Suburb Memory (a_2), Qinglong Lake Wetland (a_3) and Wen Shu Temple (a_4).

Table 7. Sub-interval hindrance factors $\zeta_{(i)}$ and hindrance function values $f(t_{(i)}, t_{(j)})_{(k)}$ output by ICSOA.

T-A	$C_r(S_A, a_{(1)})$	$C_r(S_A, a_{(2)})$	$C_r(S_A, a_{(3)})$	$C_r(S_A, a_{(4)})$		
$\zeta_{(1)}$	0.5319	0.2273	0.1319	0.5495		
$\zeta_{(2)}$	0.4348	0.1538	0.0847	0.4545		
$\zeta_{(3)}$	4.3478	1.5385	0.8475	4.5455		
$f(t_{(i)}, t_{(j)})_{(k)}$	5.3145	1.9196	1.0641	5.5495		
T-A	$C_r(a_{(1)}, a_{(2)})$	$C_r(a_{(1)}, a_{(3)})$	$C_r(a_{(1)}, a_{(4)})$	$C_r(a_{(2)}, a_{(3)})$	$C_r(a_{(2)}, a_{(4)})$	$C_r(a_{(3)}, a_{(4)})$
$\zeta_{(1)}$	0.1712	0.1109	0.3012	0.2075	0.2591	0.1188
$\zeta_{(2)}$	0.1124	0.0704	0.2128	0.1389	0.1786	0.0758
$\zeta_{(3)}$	1.1236	0.7042	2.1277	1.3889	1.7857	0.7576
$f(t_{(i)}, t_{(j)})_{(k)}$	1.4072	0.8855	2.6416	1.7352	2.2234	0.9521
T-B	$C_r(S_B, b_{(1)})$	$C_r(S_B, b_{(2)})$	$C_r(S_B, b_{(3)})$	$C_r(S_B, b_{(4)})$		
$\zeta_{(1)}$	0.0482	0.0674	0.0510	0.0571		
$\zeta_{(2)}$	0.1149	0.1786	0.1235	0.1429		
$\zeta_{(3)}$	2.2989	3.5714	2.4691	2.8571		
$f(t_{(i)}, t_{(j)})_{(k)}$	2.4620	3.8174	2.6436	3.0571		
T-B	$C_r(b_{(1)}, b_{(2)})$	$C_r(b_{(1)}, b_{(3)})$	$C_r(b_{(1)}, b_{(4)})$	$C_r(b_{(2)}, b_{(3)})$	$C_r(b_{(2)}, b_{(4)})$	$C_r(b_{(3)}, b_{(4)})$
$\zeta_{(1)}$	0.0991	0.0808	0.0390	0.0938	0.0449	0.0453
$\zeta_{(2)}$	0.3226	0.2326	0.0885	0.2941	0.1053	0.1064
$\zeta_{(3)}$	6.4516	4.6512	1.7699	5.8824	2.1053	2.1277
$f(t_{(i)}, t_{(j)})_{(k)}$	6.8733	4.9646	1.8974	6.2703	2.2555	2.2794

Table 8. Function values $f(t_{(i)}, t_{(j)})$ and $f(S, t_{(x)})$ output by ICSOA.

T-A	$f(t_{(i)}, t_{(j)})_{(1)}$	$f(t_{(i)}, t_{(j)})_{(2)}$	$f(t_{(i)}, t_{(j)})_{(3)}$	$f(t_{(i)}, t_{(j)})_{(4)}$	$f(S, t_{(x)})$	T-B	$f(t_{(i)}, t_{(j)})_{(1)}$	$f(t_{(i)}, t_{(j)})_{(2)}$	$f(t_{(i)}, t_{(j)})_{(3)}$	$f(t_{(i)}, t_{(j)})_{(4)}$	$f(S, t_{(x)})$
a_{1234}	5.3145	1.4072	1.7352	0.9521	2.5254	b_{1234}	2.4620	6.8733	6.2703	2.2794	1.1499
a_{1243}	5.3145	1.4072	2.2234	0.9521	2.3989	b_{1243}	2.4620	6.8733	2.2555	2.2794	1.4337
a_{1324}	5.3145	0.8855	1.7352	2.2234	2.3435	b_{1324}	2.4620	4.9646	6.2703	2.2555	1.2104
a_{1342}	5.3145	0.8855	0.9521	2.2234	2.8175	b_{1342}	2.4620	4.9646	2.2794	2.2555	1.4897
a_{1423}	5.3145	2.6416	2.2234	1.7352	1.5928	b_{1423}	2.4620	1.8974	2.2555	6.2703	1.5361
a_{1432}	5.3145	2.6416	0.9521	1.7352	2.1933	b_{1432}	2.4620	1.8974	2.2794	6.2703	1.5314

Table 8. Cont.

T-A	$f(t_{(i)}, t_{(j)})_{(1)}$	$f(t_{(i)}, t_{(j)})_{(2)}$	$f(t_{(i)}, t_{(j)})_{(3)}$	$f(t_{(i)}, t_{(j)})_{(4)}$	$f(S, t_{(x)})$	T-B	$f(t_{(i)}, t_{(j)})_{(1)}$	$f(t_{(i)}, t_{(j)})_{(2)}$	$f(t_{(i)}, t_{(j)})_{(3)}$	$f(t_{(i)}, t_{(j)})_{(4)}$	$f(S, t_{(x)})$
a_{2134}	1.9196	1.4072	0.8855	0.9521	3.4112	b_{2134}	3.8174	6.8733	4.9646	2.2794	1.0476
a_{2143}	1.9196	1.4072	2.6416	0.9521	2.6604	b_{2143}	3.8174	6.8733	1.8974	2.2794	1.3732
a_{2314}	1.9196	1.7352	0.8855	2.6416	2.6051	b_{2314}	3.8174	6.2703	4.9646	1.8974	1.1499
a_{2341}	1.9196	1.7352	0.9521	2.6416	2.5261	b_{2341}	3.8174	6.2703	2.2794	1.8974	1.3872
a_{2413}	1.9196	2.2234	2.6416	0.8855	2.4786	b_{2413}	3.8174	2.2555	1.8974	4.9646	1.4338
a_{2431}	1.9196	2.2234	0.9521	0.8855	3.1503	b_{2431}	3.8174	2.2555	2.2794	4.9646	1.3455
a_{3124}	1.0641	0.8855	1.4072	2.2234	3.2295	b_{3124}	2.6436	4.9646	6.8733	2.2555	1.1685
a_{3142}	1.0641	0.8855	2.6416	2.2234	2.8974	b_{3142}	2.6436	4.9646	1.8974	2.2555	1.5501
a_{3214}	1.0641	1.7352	1.4072	2.6416	2.6053	b_{3214}	2.6436	6.2703	6.8733	1.8974	1.2103
a_{3241}	1.0641	1.7352	2.2234	2.6416	2.3444	b_{3241}	2.6436	6.2703	2.2555	1.8974	1.5082
a_{3412}	1.0641	0.9521	2.6416	1.4072	3.0793	b_{3412}	2.6436	2.2794	1.8974	6.8733	1.4895
a_{3421}	1.0641	0.9521	2.2234	1.4072	3.1505	b_{3421}	3.0571	2.2794	2.2555	6.8733	1.3547
a_{4123}	5.5495	2.6416	1.4072	1.7352	1.8457	b_{4123}	3.0571	1.8974	6.8733	6.2703	1.1591
a_{4132}	5.5495	2.6416	0.8855	1.7352	2.2644	b_{4132}	3.0571	1.8974	4.9646	6.2703	1.2151
a_{4213}	5.5495	2.2234	1.4072	0.8855	2.4699	b_{4213}	3.0571	2.2555	6.8733	4.9646	1.1174
a_{4231}	5.5495	2.2234	1.7352	0.8855	2.3356	b_{4231}	3.0571	2.2555	6.2703	4.9646	1.1314
a_{4312}	5.5495	0.9521	0.8855	1.4072	3.0704	b_{4312}	3.0571	2.2794	4.9646	6.8733	1.1127
a_{4321}	5.5495	0.9521	1.7352	1.4072	2.5174	b_{4321}	3.0571	2.2794	6.2703	6.8733	1.0708

(iv) Figure 8a,b show the fluctuating curves of hindrance function values in sub-intervals of Tourist A and B. Figure 8c,d show the fluctuating curves of hindrance function values in intervals searched by the ICSSOA of Tourist A and B.

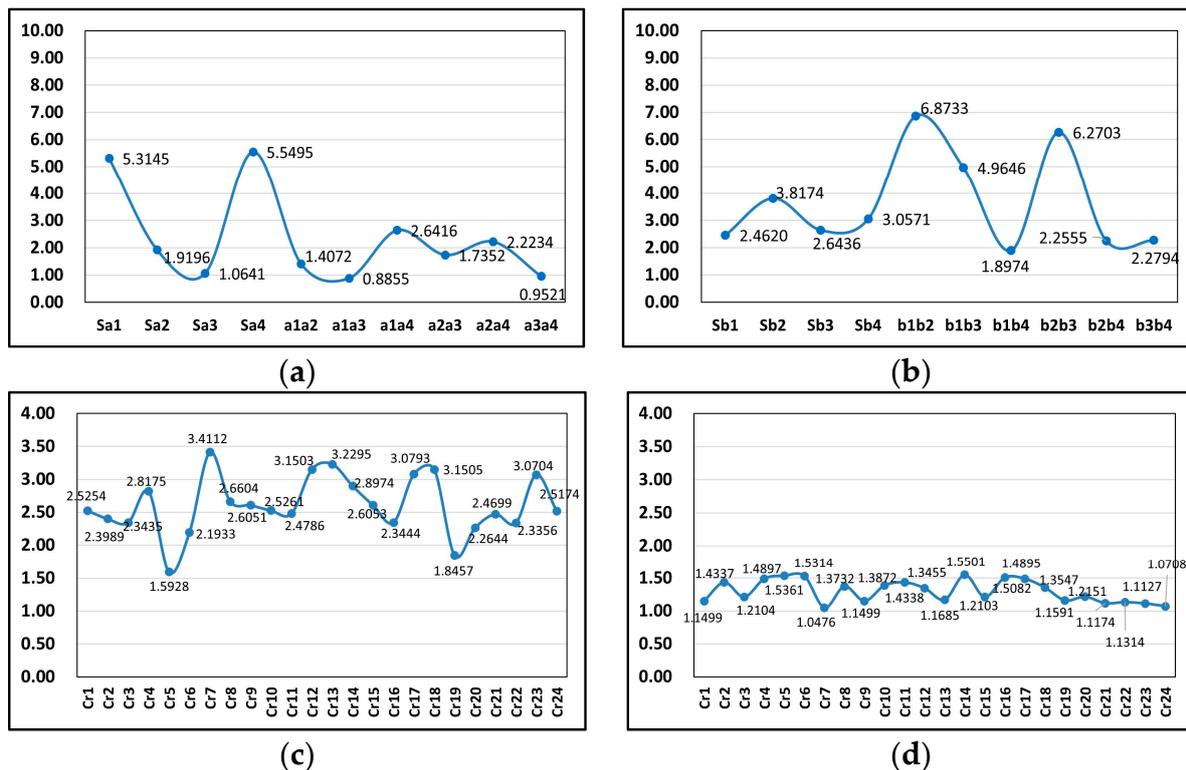


Figure 8. Trends of the hindrance function values in sub-intervals and the hindrance function values in intervals. (a,b) show the fluctuating curves of hindrance function values in sub-intervals of TouristS A and B. (c,d) show the fluctuating curves of hindrance function values in intervals searched by the ICSSOA of TouristS A and B.

As to Tourist A, interval a_{1423} relates to the minimum interval hindrance function value. Tourist A taking the route “ S_A : Tianfu Square—Temple of Marquis (a_1)—Wen shu Temple (a_4)—Eastern Suburb Memory (a_2)—Qinglong Lake Wetland (a_3)” for traveling will bring the lowest costs.

As to Tourist B, interval b_{2134} relates to the minimum interval hindrance function value. Tourist B taking the route “ S_B : Chengdu railway station—the People’s park (b_2)-Temple of Marquis (b_1)—Du Fu Thatched Cottage (b_3)—Chengdu Happy Valley (b_4)” for traveling will bring the lowest costs.

(2) Analysis on the POI route recommendation results

(i) Analyze Table 7. The hindrance factor of the sub-interval is determined by the ICSOA within the sub-interval. It can find the shortest path within the sub-interval, resulting in the lowest travel distance, travel fee and travel time.

① The sub-interval hindrance factor is inversely proportional to the travel distance, travel fee and travel time.

② The sub-interval hindrance function value is also positively correlated with the sub-interval hindrance factor, and negatively correlated with the travel distance, travel fee and travel time.

③ The higher the values of the travel distance, travel fee and travel time are, the smaller the sub-interval hindrance factors will be, and the higher the sub-interval hindrance function values will be, the higher the travel costs will be in the sub-interval; tourists traveling through all sub-intervals will cause higher travel costs.

(ii) Analyze Figure 8a,b. The hindrance function values generated by tourists in different sub-intervals all show fluctuating trends.

① As to Tourist A, the minimum sub-interval hindrance value is 0.8855, at $C_r(a_{(1)}, a_{(3)})$, indicating that the travel costs between Temple of Marquis and Qinglong Lake Wetland is the highest; The maximum sub-interval hindrance value is 5.5495, at $C_r(S_A, a_{(4)})$, indicating that the travel costs between Tianfu Square and Wenshu Temple are the lowest.

② As to Tourist B, the minimum sub-interval hindrance value is 1.8974, indicating that the travel costs between Temple of Marquis and Chengdu Happy Valley are the highest; the maximum sub-interval hindrance value is 6.8733, indicating that the travel costs between Temple of Marquis and the People’s Park are the lowest.

(iii) Analyze Table 8. The interval hindrance function value is determined by the ICSOA within the interval. The algorithm can find out the shortest path within the interval, making the travel costs lowest.

The interval hindrance function value is negatively correlated with the sub-interval hindrance function value. The higher the sub-interval hindrance function value is, the smaller the interval hindrance function value will be and the lower travel costs will be; the route will be more likely recommended. Conversely, the smaller the sub-interval hindrance function value is, the larger the interval hindrance function value will be and higher travel costs will be produced; the route’s probability of being recommended will be lowered.

(iv) Analyze Figure 8c,d. The hindrance function values corresponding to different intervals all show fluctuating trends.

① As to Tourist A, the minimum interval hindrance value is 1.5928 for route a_{1423} ; the maximum interval hindrance value is 3.4112 for route a_{2134} . This indicates that the travel costs of route a_{1423} are the lowest, while those for the route a_{2134} are the highest.

② As to Tourist B, the minimum interval hindrance value is 1.0476 for route b_{2134} ; the maximum interval hindrance value is 1.5501 for route b_{3142} . This indicates that the travel costs of route b_{2134} are the lowest, while those for the route b_{3142} are the highest.

(v) Tables 7 and 8 and Figure 8 prove that ICSOA can find out the global optimal solution. When cockroach swarm crawling towards the current optimal solution, each step can find out a feasible solution until all solutions are traversed. It traverses all control nodes within the sub-interval and all POIs within the interval, allowing the algorithm to jump out of the constraints of local optimal solutions and finally find out the global optimal solution.

4.2.5. Route Method Comparison and Analysis

(1) Selection of comparative algorithms and design of comparative experiment

To verify the advantages of the proposed algorithm, a comparative experiment is designed in this section. Referring to the comparative experimental methods used in the literature figure [28,29], combined with the actual travel experiences of tourists, the algorithms commonly used for map route searching are selected as the control group in the experiment. The proposed algorithm is set as the experimental group. The experiment compares the algorithms in the following aspects:

- (i) The comparison of the optimal output route and cost function of each algorithm;
- (ii) The advantages of the experimental group algorithm in reducing travel costs;
- (iii) The advantages of the experimental group algorithm in terms of computational complexity.

The principle and method for selecting control group algorithms in the comparative experiment are as follows.

- (i) In real tourism scenarios, tourists generally search for POI routes through electronic maps, which plan routes that meet travel requirements for tourists based on urban geographical and transportation conditions. Due to the tendency of map algorithms to overlook certain road nodes between two POIs, it is possible that the routes searched in sub-intervals $C_r(t_{(i)}, t_{(j)})$ may not be the global optimal solutions. Therefore, the traditional map route searching methods have limitations. Our proposed algorithm is based on ICSOA. Each possible road node is a node included in the individual cockroach, and the searching objective is the global optimal solution. In terms of algorithm design and logic, our proposed algorithm has advantages over the control group algorithms.
- (ii) The commonly used algorithms for map route searching include the Dijkstra algorithm and A* algorithm, both of which are shortest-path-searching algorithms. They themselves have certain drawbacks. For example, Dijkstra is a local greedy searching algorithm that does not traverse all feasible solutions, resulting in low computational efficiency when there are too many nodes. The A* algorithm also does not traverse all feasible solutions, resulting in low computational efficiency when there are too many nodes. Our proposed algorithm essentially compares the cost functions of cockroach individuals. Firstly, the algorithm of searching for the quantity of cockroach individuals is the factorial calculation, while the process of calculating the cost of each cockroach individual is linear, with fast computational efficiency. The process of comparing individual costs of cockroaches uses the sorting algorithm, which has high computational efficiency. The comparative experiment sets Tencent Maps (TCA) and GaoDe Maps (GDA) as the control group, using Dijkstra Dijkstra algorithm and A* algorithm, respectively. The proposed algorithm ICSOA is the experimental group.
- (iii) The control group algorithms are used for map route searching and they are the most classical methods among route searching algorithms. Based on the conventional behaviors of tourists in tourism activities and the universality and convenience of the current electronic maps on planning routes, the map searching algorithm is determined as the control group algorithm. The comparison between the proposed algorithm and the classical map route searching algorithms not only has academic research value but also practical application value, which can provide new ideas for improving map route searching algorithms.

According to the principle and method in determining the control group algorithms, the specific content of the comparative experiment is designed as follows. The control group algorithms are TCA and GDA, while the experimental group algorithm is ICSOA.

- (i) Set the same POIs as the algorithm nodes.

Tourist A will visit: $T_{u(1)}$: Wu hou Temple (a_1), $T_{u(9)}$: Eastern Suburb Memory (a_2), $T_{u(10)}$: Qinglong Lake Wetland (a_3), $T_{u(13)}$: Wenshu Academy (a_4);

Tourist B will visit: $T_{u(1)}$: Wuhou Temple (b_1), $T_{u(3)}$: The People’s Park (b_2), $T_{u(4)}$: Du Fu Thatched Cottage (b_3), and $T_{u(7)}$: Happy Valley (b_4).

- (ii) Set the same transportation mode. Tourist A chooses the electric bicycle, while Tourist B chooses the taxi. The average speed of a taxi in the urban area is set as 20 km/h. The charging rules for the electric bicycle are: the initial time range, 2 CNY/15 min; exceeding the initial time, 1 CNY/10 min. The charging rules for a taxi are: the initial time range, 8 CNY/2 km; exceeding the initial time, 1.9 CNY/km. Tourist A’s starting point is the Tianfu Square, Tourist B’s starting point is the Chengdu Railway Station.
- (iii) Take the travel conditions provided by Tourists A and B as the comparative experimental conditions, and each algorithm searches for the optimal route and two suboptimal routes that pass through the starting point and all POIs without repeating any node. Make comparisons on the output optimal routes and costs, route cost differences and algorithm time complexity on the three algorithms.

(2) Comparison results on the route algorithms

- (i) Table 9 shows the three optimal routes and relative data searched by TCA, GDA and ICSOA under the same experimental conditions. The number 1 is the optimal route and the numbers 2 and 3 are the suboptimal routes.

Table 9. The three optimal POI routes and route data searched by ICSOA, TCA and GDA.

T-A		$C_r(S, t(x))$	$f(t(i), t(j))_{(1)}$	$f(t(i), t(j))_{(2)}$	$f(t(i), t(j))_{(3)}$	$f(t(i), t(j))_{(4)}$	$f(S, t(x))$
ICSOA	Route-1	a_{1423}	5.3145	2.6416	2.2234	1.7352	1.5928
	Route-2	a_{4123}	5.5495	2.6416	1.4072	1.7352	1.8457
	Route-3	a_{1432}	5.3145	2.6416	0.9521	1.7352	2.1933
TCA	Route-1	a_{1423}	5.3145	2.6416	1.8630	1.7118	1.6877
	Route-2	a_{4123}	5.3145	2.6416	1.3766	1.7118	1.8773
	Route-3	a_{1432}	5.3145	2.6416	0.9380	1.7118	2.2170
GDA	Route-1	a_{1423}	4.5458	2.4857	2.1119	1.1618	1.9565
	Route-2	a_{4123}	4.7162	2.4857	1.3473	1.1618	2.2173
	Route-3	a_{1243}	4.5458	1.3473	2.1119	0.8223	2.6519
T-B		$C_r(S, t(x))$	$f(t(i), t(j))_{(1)}$	$f(t(i), t(j))_{(2)}$	$f(t(i), t(j))_{(3)}$	$f(t(i), t(j))_{(4)}$	$f(S, t(x))$
ICSOA	Route-1	b_{2134}	3.8174	6.8733	4.9646	2.2794	1.0476
	Route-2	b_{4321}	3.0571	2.2794	6.2703	6.8733	1.0708
	Route-3	b_{4312}	3.0571	2.2794	4.9646	6.8733	1.1127
TCA	Route-1	b_{2134}	1.9623	3.6475	2.5796	0.9780	2.1939
	Route-2	b_{4321}	1.3920	0.9780	3.3291	3.6475	2.3154
	Route-3	b_{1234}	1.3126	3.6475	3.3291	0.9780	2.3589
GDA	Route-1	b_{2134}	1.8974	3.1461	2.5796	1.1319	2.1160
	Route-2	b_{2314}	1.8974	3.3291	2.5796	1.0124	2.2028
	Route-3	b_{4321}	1.3920	1.1319	3.3291	3.1461	2.2201

- (ii) Figure 9 shows the comparison data curves between the optimal route 1 (Route 1), suboptimal route 2 (Route 2) and suboptimal route 3 (Route 3) of the experimental and control group. Figure 9a–c represent the 1–3 routes of Tourist A, red corresponding to ICSOA, green corresponding to TCA and blue corresponding to GDA. Figure 9d–f represent the 1–3 routes of Tourist B, red corresponding to ICSOA, green corresponding to TCA and blue corresponding to GDA.

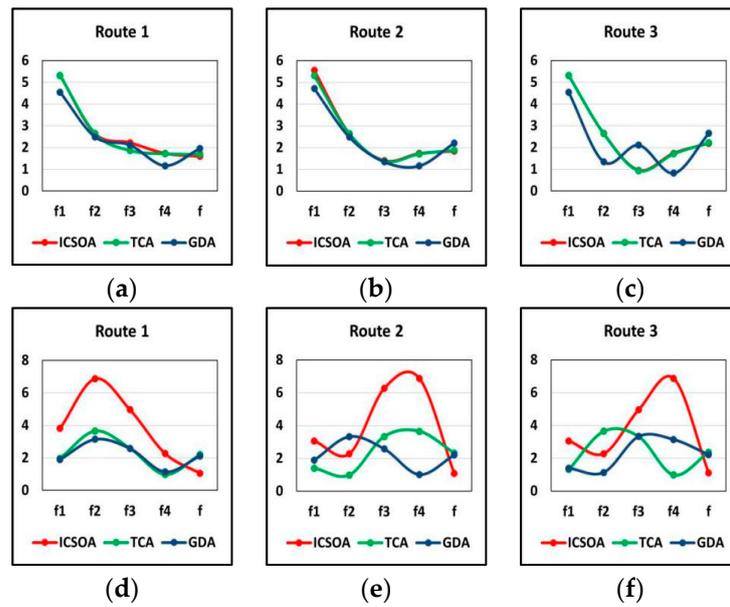


Figure 9. The comparison curves between the Route 1~3 by the three methods ICSOA, TCA and GDA. (a–c) represent the 1–3 routes of Tourist A, red corresponding to ICSOA, green corresponding to TCA and blue corresponding to GDA. (d–f) represent the 1–3 routes of Tourist B, red corresponding to ICSOA, green corresponding to TCA and blue corresponding to GDA.

(iii) Based on the experimental results in Table 9, extract the total cost $f(S, t_{(x)})$ of each route output by each algorithm. Taking a tourist as an example, for the optimal Route 1, calculate the difference in route cost between ICSOA and TCA, as well as between ICSOA and GDA, and calculate the ratio of cost reduction of ICSOA compared to TCA and ICSOA compared to GDA. Similarly, for Route 2 and Route 3, the same calculations are also performed and the experimental results in Table 10 are obtained. It proves that the proposed algorithm ICSOA has advantages over the control group algorithms and can effectively reduce travel costs.

Table 10. The reduced ratio on the travel costs for ICSOA comparing to TCA and GDA.

		Route 1 (Optimal)	Route 2 (Sub-Optimal)	Route 3 (Sub-Optimal)
T-A	ICSOA-TCA	5.62%	1.68%	1.07%
	ICSOA-GDA	18.60%	16.76%	17.29%
T-B	ICSOA-TCA	52.25%	53.75%	52.83%
	ICSOA-GDA	50.49%	51.39%	49.88%

(iv) Table 10 shows the ratio of cost reduction for ICSOA compared to TCA and GDA. “ICSOA-TCA” in the table represents the ratio of cost reduction in the experimental group’s ICSOA compared to the control group’s TCA, and “ICSOA-GDA” represents the ratio of cost reduction in the experimental group’s ICSOA compared to the control group’s GDA. T-A represents Tourist A, T-B represents Tourist B.

(v) Analyze the time complexity on the experimental group algorithm and the control group algorithms. In the route map graph composed by POIs, the number of nodes is n , which is the number of POIs.

① The time complexity of the Dijkstra algorithm used in the control group TCA is determined by the number of nodes, and it is a kind of single-source shortest path searching method.

- ② The time complexity of the A* algorithm used by the control group GDA is also determined by the number of nodes, and it is a kind of heuristic searching algorithm and has a faster computational efficiency than the Dijkstra algorithm.
- ③ The proposed algorithm uses factorials to calculate the cost function of individual cockroach. When searching for the optimal individual cockroach, the maximum searching algorithm is used, and its convergence speed is very fast. According to the analysis, the time complexities of the three algorithms are shown in Table 11.

Table 11. The comparison on time complexities of ICSOA, TCA and GDA.

	TCA	GDA	ICSOA
Time Complexity	$O(n^2)$	$O(n \log n)$	$O(n)$

- (3) Analysis on the route comparison results
- (i) Analyze the Table 9 and Figure 9. For each route, the sub-interval hindrance function values of the three algorithms all show fluctuating trends.
- ① For Tourist A, the TCA, GDA and ICSOA curves of route 1 are close, but overall, the hindrance function values of the ICSOA sub-intervals are relatively higher. For route 3, the TCA and ICSOA curves are relatively close, while the GDA and ICSOA curves are significantly different. Overall, the hindrance function values of the ICSOA sub-intervals are higher. Therefore, the overall performance of ICSOA is the best, and its output interval hindrance function values are the lowest in routes 1, 2 and 3; the travel costs are also the lowest. TCA performance is close to ICSOA, while the interval hindrance function value of GDA is the highest, so its route costs are also the highest.
- ② For Tourist B, the TCA and GDA curves of route 1 are relatively close, and the sub-interval hindrance function values of ICSOA are higher than those of TCA and GDA. The sub-interval hindrance function values of ICSOA in routes 2 and 3 are generally higher than those of TCA and GDA. Therefore, ICSOA has the lowest hindrance function values in the interval of routes 1, 2 and 3, and its optimal and suboptimal routes all have lower costs than TCA and GDA.
- (ii) Analyze Table 10, under the conditions of the two tourist samples:
- ① For the optimal route 1, the lowest cost reduction of ICSOA is 5.62% compared to the control group methods, and the highest cost reduction is 52.25%;
- ② For the suboptimal route 2, the lowest cost reduction of ICSOA is 1.68% compared to the control group methods, and the highest cost reduction is 53.75%;
- ③ For the suboptimal route 3, the lowest cost reduction of ICSOA is 1.07% compared to the control group methods, and the highest cost reduction is 52.83%.
- The ICSOA generates lower travel costs on both optimal and suboptimal routes compared to TCA and GDA, which can help tourists effectively save travel expenses.
- (iii) Analyze the time complexity of each algorithm in Table 11.
- ① The control group uses the Dijkstra algorithm and A* algorithm to search for the shortest route, with the time complexities of $O(n^2)$ and $O(n \log n)$. As to the time complexity $O(n)$ of the experimental group algorithm, when the number n reaches a certain level, the time complexity of the experimental group algorithm is always lower than that of the control group algorithms. The higher the value of n is, the higher efficiency of the experimental algorithm will be, and the lower time complexity will be, compared to the control group algorithms.
- ② In actual tourism scenarios, the value of n is usually $n \geq 3$, that is, a complete POI route is composed of at least three POIs. Therefore, when the number of POIs to be visited is larger than 3, the proposed algorithm has the advantage of higher computational efficiency. When the number of POIs to be visited is less than 3, the computational efficiency of the three algorithms is equivalent, but the route planning algorithm has no practical significance.

Based on the above analysis, the time complexity of the proposed algorithm is lower than that of the control group algorithms, and it converges the fastest and has the highest efficiency when searching for the optimal route.

- (iv) This proves that the proposed ICSOA can realize the optimal POI route recommendation based on the selected optimal POIs under the constraints of tourists' traveling conditions and geospatial conditions. It can provide tourists with POI routes that best match their interests and have the lowest costs. Compared with the commonly used TCA and GDA methods, ICSOA has better performance in the aspect of reducing travel costs.

5. Conclusions and Prospects

5.1. The Conclusions on the Research

On the basis of analyzing the existing problems in the current tourism recommendation research, this paper proposes a tourism recommendation model based on improved Naive Bayes classification space accessibility (NBCSA) and an improved cockroach swarm optimization algorithm (ICSOA). It mainly solves three problems. First, an improved Naive Bayes classification algorithm (NBCA) fused with a disturbance factor is constructed by using historical POIs provided by tourists as training data set. It realizes the POIs' classification in the destination city under the constraints of tourists' requirements, which makes the POIs match tourists' interests. Second, based on the geographical and transportation constraints, a spatial accessibility field strength (SAFS) model based on Naive Bayes classification is established. The tourist demand weights are introduced into the model. It obtains the POIs with the best spatial distribution and the lowest travel costs from the categories. It realizes that the recommended POIs simultaneously satisfy interests and produce lowest spatial costs. Third, based on the recommended POIs, the geographical constraint factors are introduced to propose the ICSOA to realize the optimal POI route recommendation. An experiment is performed to compare the TCA and GDA methods with the ICSOA. The experimental results show that the ICSOA can effectively classify POIs based on tourists' interests, and visually output SAFS of the POIs. The ICSOA can output POI routes with lower travel costs. In the optimal routes, compared with the control group algorithms, the ICSOA can reduce costs by 5.62% at the lowest and 52.25% at the highest, and effectively reduce the travel expenses.

5.2. The Potential Applications of the Recommendation Model

The proposed tourism recommendation model can provide new methods for smart tourism research. It plays a positive role in promoting the development of tourism informatization and intelligence. The improved NBCA, SAFS model and the ICSOA in the model have broad application prospects.

Firstly, the improved NBCA could be used to develop an intelligent system for tourist classification, which can classify tourists with different cultural contexts, interest conditions and tourism requirements, providing reference for tourism decision-making, marketing and planning. An intelligent system for POI classification could also be developed, which can classify POIs in the destination city based on user interests, and can also manage, plan and recommend POIs. In addition, the text mining algorithm constructed in the classification algorithm could be used to develop the tourism text analysis and mining system, which can mine and analyze the tourism encyclopedia big data, tourism evaluation data, etc., and provide decision-making support for tourism official departments.

Secondly, the SAFS model could be used to construct a spatial accessibility distribution model for the urban, municipal and provincial POIs. A dynamic distribution map could be established from multiple dimensions such as spatial distances, functional attributes, popularity levels and travel costs, providing tourists with visualized three-dimensional maps and decision-making support for planning their itineraries. In addition, the SAFS model could be used for decision-making in tourism transportation deployment, and in the planning of tourism transportation projects based on the spatial field strength

of different POIs, determining the locations of tourism service centers, and conducting buffer zone analysis of tourism service centers, which will provide convenient services for tourists' traveling.

Thirdly, based on the ICSOA, the functions of existing electronic maps could be optimized to improve the accuracy and efficiency of searching for the optimal routes. The algorithm could be used to develop a POI route planning system, providing decision support for tourists to make POI route schedules. Combining with the intelligent connected vehicle technology, it could be used to develop a tourism intelligent connected vehicle route guidance system, which includes buffer zone POI searching, POI recommendation and intelligent vehicle route guidance as the system functions, providing convenient services for future tourists' traveling. Combined with tourism emergency management, the algorithm could be used to construct the spatial distribution relationship between emergency shelters and POIs in tourism cities, plan emergency escape routes and provide decision-making solutions for urban emergencies.

5.3. Future Research Directions and Challenges

Based on this study, the following future research directions and challenges are proposed. Firstly, we introduce the concept of tourist demand weight when constructing the POI recommendation algorithm based on Naive Bayes classification space accessibility, representing the importance evaluation on the two factors of attribute matching and spatial accessibility by tourists. In future studies, further quantitative research on the tourist demand weight is needed to study the differences in recommendation results under different tourist demand conditions on the relationship between attribute demand and spatial accessibility weight. Secondly, the establishment of the SAFS model requires further in-depth research on the multidimensional attributes of POIs. Based on attribute characteristics, a multidimensional field strength distribution model could be established, and the SAFS model could be established from multiple dimensions such as cultural attributes, service attributes, popularity levels and tourism costs to further improve the accuracy of recommended POIs matching tourists' interests. Thirdly, the constraints introduced to the construction of the ICSOA are the key factors that affect the travel costs. In future studies, other factors that affect the recommendation results of the ICSOA could be studied based on real-world tourism scenarios. On the above research directions, some challenges might be concluded. Since more factors could be merged into the modeling process, a more complicated algorithm design will be developed. The accuracy and efficiency are the two factors that must be considered. A more complicated design might cause higher time complexity. Thus, how to create an algorithm with suitable computational complexity is a challenge. Furthermore, the demand on the large quantity of big data is also a challenge. Data collection, data processing and data mining should conform to the high-dimensional algorithm modeling to guarantee accurate recommendation results. In future research, the challenges will be the motivation to improve the work.

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) and the National 14th Five-Year Plan Key R&D Projects (No. SQ2021YFB3900025).

Data Availability Statement: Data are contained within the article.

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

References

1. Chen, X.; Pan, Y.; Luo, B. Research on power-law distribution of long-tail data and its application to tourism recommendation. *Ind. Manag. Data Syst.* **2020**, *121*, 1268–1286. [[CrossRef](#)]
2. Zheng, Y.; Wang, H.; Cheng, T.; Chen, J. Tourism attraction recommendation algorithm based on deep neural network matrix factorization. *J. Hubei Univ. Technol.* **2021**, *36*, 29–33.
3. Chen, Y.; Gu, T.; Bin, C.; Liang, C.; Wang, W.; Li, Y. Attractions recommendation method incorporating graph representation learning with sequence mining. *Comput. Eng. Des.* **2020**, *41*, 3563–3569.
4. Zheng, S.; Tan, G.; Shi, Z. Recommending Tourism Attractions Based on Segmented User Groups and Time Contexts. *Data Anal. Knowl. Discov.* **2020**, *4*, 92–104.
5. Lin, J.; Gao, M.; Chen, C. Constrained association rules mining and its application on personalized tourist spots recommendation. *J. Fuzhou Univ.* **2019**, *47*, 320–326.
6. Cheng, P.; Liu, L.; Liu, X.; Xu, C.; Guo, H. Scenic spots recommendation algorithm based on multi-dimensional feature clustering and user score. *Comput. Eng. Des.* **2019**, *40*, 1322–1327.
7. Ahn, J.; Kim, E.; Kim, H. Big Data based POIs Recommendation—Focus on Korean Tourism Organization Linked Open Data. *Manag. Inf. Syst. Rev.* **2017**, *36*, 129–148.
8. Hong, M.; Jung, J. Multi-criteria tensor model consolidating spatial and temporal information for tourism recommendation. *J. Ambient. Intell. Smart Environ.* **2021**, *13*, 5–19. [[CrossRef](#)]
9. Abbasi-Moud, Z.; Vahdat-Nejad, H.; Sadri, J. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst. Appl.* **2021**, *167*, 114324. [[CrossRef](#)]
10. Jeong, C.-S.; Ryu, K.-H.; Lee, J.-Y.; Jung, K.-D. Deep Learning-based Tourism Recommendation System using Social Network Analysis. *Int. J. Internet Broadcast. Commun.* **2020**, *12*, 113–119.
11. Cui, C.; Wang, X.; Li, W. Research on the Algorithm of POI Recommendation Based on User's Online Comments. *J. Syst. Sci. Math. Sci.* **2020**, *40*, 1103–1116.
12. Cui, C.; Wang, X.; Li, W. Research on the Recommendation Algorithm of Tourism Products Based on User Portrait in Situational Environment. *J. Math. Pract. Theory* **2019**, *49*, 122–131.
13. Almulihi, A.; Alassery, F.; Khan, A.; Shukla, S.; Gupta, B.K.; Kumar, R. Analyzing the Implications of Healthcare Data Breaches through Computational Technique. *Intell. Autom. Soft Comput.* **2022**, *32*, 1763–1779. [[CrossRef](#)]
14. Han, S.; Liu, C.; Chen, K.; Gui, D.; Du, Q. A POI Recommendation Model Fusing Spatial, Temporal, and Visual Embeddings for Flickr-Geotagged Photos. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 20. [[CrossRef](#)]
15. Filipe, S.; Almeida, A.; Martins, C.; Gonçalves, R.; Martins, J. Using POI functionality and accessibility levels for delivering personalized tourism recommendations. *Comput. Environ. Urban Syst.* **2019**, *77*, 101173.
16. Zhang, X.; Yu, L.; Wang, M.; Gao, W. FM-based: Algorithm research on rural tourism recommendation combining seasonal and distribution features. *Pattern Recognit. Lett.* **2018**, *150*, 297–305. [[CrossRef](#)]
17. Ou, G.; He, Y.; Fournier Viger, P.; Huang, J.Z. A Novel Mixed-Attribute Fusion-Based Naive Bayesian Classifier. *Appl. Sci.* **2022**, *12*, 10443. [[CrossRef](#)]
18. Yu, S. A Study on Recommendation Method for Real Estate Using Naive Bayes Classification. *J. Korean Inst. Inf. Technol.* **2019**, *17*, 115–120.
19. Lee, J.-Y.; Choi, J.-W.; Choi, J.-H.; Lee, B.-H. Text-mining analysis using national R&D project data of South Korea to investigate innovation in graphene environment technology. *Int. J. Innov. Stud.* **2023**, *7*, 87–99.
20. Ayesha, S.; Sadia, I. A GIS Based Measurement of Accessibility of Urban Parks in Faisalabad City, Pakistan. *Acad. Res. Int.* **2014**, *5*, 94–99.
21. Kiran, K.C.; Corcoran, J.; Chhetri, P. Measuring the spatial accessibility to fire stations using enhanced floating catchment method. *Socio-Econ. Plan. Sci.* **2020**, *69*, 100673.
22. Wu, X.; Chen, C. Spatial distribution and accessibility of high level scenic spots in inner Mongolia. *Sustainability* **2022**, *14*, 7329. [[CrossRef](#)]
23. Ding, M. Research on Tourism Route Planning Based on Artificial Intelligence Technology. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 2227798. [[CrossRef](#)]
24. Gao, Y.; Wang, J.; Wu, W.; Sangaiah, A.K.; Lim, S.-J. Travel Route Planning with Optimal Coverage in Difficult Wireless Sensor Network Environment. *Sensors* **2019**, *19*, 1838. [[CrossRef](#)] [[PubMed](#)]
25. Du, P.; Hu, H. Optimization of tourism route planning algorithm for forest wetland based on GIS. *J. Discret. Math. Sci. Cryptogr.* **2018**, *21*, 283–288. [[CrossRef](#)]
26. Cheng, L.; Song, Y.; Bian, Y. Cockroach Swarm Optimization Using A Neighborhood-Based Strategy. *Int. Arab J. Inf. Technol.* **2019**, *16*, 784–790.
27. Joanna, K.; Marek, P. Cockroach Swarm Optimization Algorithm for Travel Planning. *Entropy* **2017**, *19*, 213. [[CrossRef](#)]

28. Akash, S. Optimized fractional overhead power term polynomial grey model (OFOPGM) for market clearing price prediction. *Electr. Power Syst. Res.* **2023**, *214*, 108800.
29. Kavita, J.; Akash, S. Simulation on supplier side bidding strategy at day-ahead electricity market using ant lion optimizer. *J. Comput. Cogn. Eng.* **2023**, *2*, 17–27.

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.