Received December 3, 2020, accepted December 19, 2020, date of publication December 24, 2020, date of current version January 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047264

Intelligent Tourism Recommendation Algorithm based on Text Mining and MP Nerve Cell Model of Multivariate Transportation Modes

XIAO ZHOU[1,2], MINGZHAN SU[3], GUANGHUI FENG[4,5], AND XINGHAN ZHOU[2]
[1]Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041, China
[2]Tourism Department, Leshan Vocational and Technical College, Leshan 614000, China
[3]Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China
[4]Institute of Information Engineering, Zhengzhou University of Industrial Technology, Zhengzhou 451159, China
[5]School of Computer Science, Guangzhou University, Guangzhou 510006, China

Corresponding author: Mingzhan Su (smzh625@infu.ac.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0503500, in part by the Fund of Leshan Science and Technology Project under Grant 20RKX0007 and Grant 20ZRKX006, in part by the Fund of Leshan Social Sciences Federation Project under Grant SKL2020D56, and in part by the Training Program for Young Backbone Teachers of Henan Higher Education Institute under Grant 2017GGJS196.

ABSTRACT Currently, the recommendation method on tourist sight and tour route lacks of the mechanism of tourists' interests mining and the precise tourist sights recommending, and the planned tour routes cannot properly and adequately combine with the real world environment. Meanwhile, the research on the multivariate transportation modes in tour route recommendation is not sufficient. Aim at the problems of the tourist sight and tour route recommendation in intelligent tourism recommendation system of the intelligent tourism construction, this paper brings forward a tourism recommendation algorithm based on text mining and MP nerve cell model of multivariate transportation modes. The research specially focuses on the optimal tourist sight matching algorithm based on tourists' interests mining and the optimal tour route chain algorithm based on the multivariate transportation modes. First, it analyzes the problems on tourism recommendation, based on which, the tourist sight clustering algorithm on feature attribute label and the tourist sight text mining algorithm on interest label are developed. The mined tourist sights will approach tourists' interests to the maximum extent. Secondly, Considering the critical impact of the selected transportation mode on motive benefit satisfaction in the tour route chain, the tour route chain algorithm based on the nerve cell model of multivariate transportation modes is developed. This algorithm combines with geographic information element and transportation element, and it simulates the bionic principle of input and output information process on MP nerve cell, then the tour route chain model based on the nerve cell of multivariate transportation modes is set up. Through the iteration of multiple layer nerve cell motive weight values and accommodation coefficients, the algorithm finally outputs the signal information flow motive values, in which the tour route chain with the maximum information flow motive value is generated. Thirdly, to testify the feasibility and practicalness of the algorithm, an experimental example in real-world environment is designed and performed. The feasible matched tourist sights and tour route chains are output, meanwhile, the three commonly used optimal route searching algorithms are set as the control group, and along with the developed algorithm, they are compared with each other on the aspect of optimal tour route chain. The experiment testifies that the developed algorithm is feasible and practical, and has advantages on the tourism recommendation. Through the algorithm design and the experiment, it finds that the mined tourist sights by the objective function in the algorithm can best match tourists' interest labels. The algorithm adequately combines with the real world tourism data of the geographic information, traffic information and tourist sight information and outputs the tour routes that best match tourists' interests. Compare with the

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Shariq Imran.

control group algorithms, the tour routes output by the developed algorithm have the highest motive satisfaction, lowest time complexity and space complexity. The developed algorithm is mainly used as the embedded algorithm for the intelligent tourism recommendation system, whose direct aim is to provide service for the tourists. Meanwhile, it can also provide service for the tourism administrations to collect, manage and mine the interest data as well as discover knowledge, and help the government to optimize the urban transportation system, launch the public vehicles and optimize the transportation guarantee strategy.

INDEX TERMS Intelligent tourism, text mining, multivariate transportation modes, MP nerve cell model, tour route chain.

I. INTRODUCTION

In the construction of intelligent tourism, the design and development of intelligent recommendation system has strategic importance. Its core functions include tourist sight recommendation and tour route recommendation. Before traveling, the tourists should firstly confirm the destination tourism city, and then plan the tour route and schedule. In tourism city, urban tourist sights and leisure areas are the first choices. However, not all tourist sights and leisure areas conform to tourists' individual interests. Usually, the tour route schedule depends on tourists' knowledge and travel strategy searching on the Internet, which can hardly obtain valuable information in mass data, especially the information conforming to tourists' interests. Aim at the construction of intelligent recommendation system, many researches on tourist sight recommendation and tour route recommendation have been performed. Most of these researches are based on the collaborative filtering algorithm, content recommendation, knowledge recommendation, user characteristics recommendation and association rules, etc [1]. For these researches, the quantitative studies on tourists' individual interests and tourist sight feature attributes are insufficient. The recommended tourist sights cannot precisely match the tourists' interests. On the module of tour route recommendation, the combination of geographic information data and transportation data in the existing researches is also insufficient. When tourists traveling in a tourism city, they will be influenced by city geographic information service and transportation service [2], [3]. Especially when they choose different transportation modes, the same tour route may brings different motive satisfaction, that is, different transportation modes may generate different optimal tour routes. In the research area of tour route planning, there are commonly used methods such as Dijkstra algorithm, Floyd-Warshall algorithm, Bellman-Ford algorithm [56]–[60], etc. These algorithms mainly solve the TSP problem on the aspect of the shortest path and the least time consumption. They absorb and connect all the tourist sights in the city to form the tour route to find out the optimal one with the shortest path and the least time consumption.

This research mode is in the means of theoretical study and simulation study, which has certain difference with the true tourism activities in the real world environment. Firstly, the most important criterion to judge whether a tour route is optimal or not depends on the extent that the tour route could meet the tourists' interests. Secondly, tourists cannot visit all of the tourist sights in a city for one time, in the real world tourism activity, the tourist sights should be optional. Thirdly, the factors of geographic information and traffic information that influence the tourism activity in the travelling process are not considered, including the tourists' interests. Thus, the aim of the developed algorithm is to solve the problems in the tour route planning, focusing on the research of tourists' interests mining, precise tourist sights mining, geographic information and traffic information integrating, and brings forward the new method on the tour route planning, which can best match the real world travelling process.

In accordance with the contextualization, this research is very significant on the aspect of solving the problems on tourist sight and tour route recommendation, optimizing the embedded algorithm in the recommendation system and providing service for the tourists and tourism administration. On the research of tourists' interests mining, precise tourist sights matching and mining and recommending the optimal tour route based on the multivariate transportation modes, this research has technological innovation, and it can solve the problems, including that the over-dependence of tour route planning on computer simulation, neglecting the real world environment for tourism activity and specific conditions for tourists' travelling process, not considering the factors of multivariate transportation modes, etc. Also, compared with the experimental algorithms, it has the advantages on higher motive benefit satisfaction, lower time complexity and space complexity. The developed algorithm in the research is mainly used as the embedded algorithm for the intelligent tourism recommendation system, whose direct aim is to serve for tourists, and provide service for the tourism administration to collect, manage and mine the interest data as well as discover knowledge, and help the government to optimize the urban transportation system, launch the public vehicles and optimize the transportation guarantee strategy.

Based on the above analysis, the problems that the proposed method aims to solve include the following three aspects. First, research on how to obtain the tourists' individualized interests and how to mine the precise tourist sights that best match the interests. Second, the process of recommending tour route should not only consider the shortest path in the computer simulation environment, but also should consider how to combine with the real world environment and the geographic information data and traffic information data that tourism activities rely on. Third, research on the mechanism of the intelligent recommendation system on planning tour route under the condition of multivariate transportation modes. According to the three aim issues, the interest data mining, precise tourist sight matching and the multivariate transportation modes is the critical content on tourism recommendation, which will directly influence tourists' motive satisfaction [15], [16].

Aim at the analyzed problems, this paper brings forward the tourism recommendation algorithm based on text mining and MP nerve cell model of the multivariate transportation modes [4], [5], [7]. First, as to the problem of tourist sight recommendation, the tourist sight clustering algorithm based on tourists' interest label and tourist sight feature attribute label is set up, and its function is to generate tourist sight clusters for the research domain. The clusters are the precondition for the further mining tourist sights to match tourists' interests. The matching degree between the tourist sights and interest label is generated by the text mining algorithm. Combining with the other quantitative needs of tourists, they are set as correlation factors to set up tourist sight mining objective function model. Through the objective function model, the clusters are mined to output the tourist sights that are closest to the quantitative needs of tourists. It has relatively high efficiency and accuracy. The output tourist sights are arranged from the highest matching degree to the lowest one, which is convenient for the smart recommendation system to extract the most suitable tourist sights. Second, to solve the problem of tour route recommendation, the tour route chain algorithm based on the multivariate transportation modes is set up [46], [47]. The algorithm sets the mined optimal tourist sights as critical nodes, and it simulates the mode of MP (McCulloch-Pitts) nerve cell to input and output information stream to design the tour route chain model. It also sets the tourist sight nerve cell motive weight $\omega_{(k)}$ and accommodation coefficient $\theta_{(k)}$ determined by geographic information and traffic information as upper layer positive factors and under layer negative factors [6], [13], [14]. The information stream goes through multiple layers' iteration, and finally outputs the maximum information stream motive value of the tour route chain. The global optimal searching algorithm is used to output the maximum value. The algorithm is set up under the precondition of tourists' multivariate transportation modes, and combines with multiple true environment factors to quantify the real-world tour process into mathematics model and then outputs the tour route chain that can best meet the tourists' interests [22], [23]. Finally, an experiment is designed and performed to do quantitative research on the developed algorithm. Three commonly used route planning algorithms are set as the control group to make comparison with the developed algorithm on basic experimental data, the generated clusters, the mined tourist sight results, the output tour route chains, etc [30], [31]. The experiment testifies that the developed algorithm is feasible and practical, it also has advantages as to the control group algorithms. The proposed algorithm brings forward the new thought and method on recommending tourist sights and planning tour routes, and it can solve the problem on the over-dependence of tour route planning on computer simulation, neglecting the real world environment for tourism activity and specific conditions for tourists' travelling process, not considering the factors of multivariate transportation modes, etc. Also, compared with the experimental algorithms, it has the advantages on higher motive benefit satisfaction, lower time complexity and space complexity. The developed algorithm in the research is mainly used as the embedded algorithm for the intelligent tourism recommendation system. As one of the most important contents in tourism GIS development, the intelligent tourism recommendation has the function to provide tourists with tourist sights and tour routes, meanwhile, it can provide tourism administration with the reference service on monitoring and mining interest data, researching on popular and hot tourist sights and tour routes, and predicting of tourist volume. The function of recommending tourist sights and tour routes based on multivariate transportation modes can obtain the mass data of tourists' choice on the transportation methods according to the actual conditions of the tourism city. Via the big data mining, relative knowledge can be obtained, and it can be used as the basis for the government to optimize the public transportation system, launch the public vehicle and optimize the tourism transportation guarantee strategy.

By analyzing the previous research, the developed algorithm in the research is supported and enlightened by the former studies. The literature [1] surveys and analyzes the building methods of the tourism recommendation system, and discusses the issues on the recommendation function, features and the problems to be solved. The literature [2] studies the new method and framework for thee tourism plan, which provides for the theoretical basis for the research. Literature [6] provides a research method on tourism system based on GIS and network analysis. It studies on the spatial structure of the tourist sights. Based on the analysis. Literature [15], [22], [23], [35] and [39] study on the important factors that influence the tourism recommendation system. By analyzing these factors, this research brings forward the thought to set up the influence factors for the tourist sights mining and tour route planning algorithm. of the research, the research brings forward the thought to classify the tourist sights according to feature attribute labels. Literature [15] analyzes the method to choose residence combining with low-carbon traveling on the aspect of tourists' choice. Literature [22] sets up the tourists' needs predicting model by the neural network and provides a method to get tourists' interests. Literature [23] combines the neural network with Grey-Markov model, and brings forward a method to predict tourists' interests. The two predicting methods provide reference for this research on the aspect of tourists' interest mining. Literature [35] studies the influence of the geographically weighted regression on the usage frequency of the city transportation modes. Literature [39] studies the perception on the transportation mode and the traveling time on the aspect of socioeconomic characteristics, trip characteristics and facility usage. The two literature have important value for this research's study on the influence of multivariate transportation modes on tourism activities. Literature [7], [13], [14], [18] and [28] study the tourist sight and tour route recommending methods on the aspect of specific technique and algorithm developing. By analyzing these methods, the problems to be solved and optimized on the tourism recommendation system are concluded, and the tourist interest data mining, precise tourist sight mining and the tour route planning algorithm based on the multivariate transportation modes. Literature [7] studies on the lane-level road map for intelligent vehicle systems, and analyzes the factors that influence the city vehicle systems and the key method to set up the model. Literature [13] studies on the shortest tour route by the method of genetic algorithm and the oriented spanning tree. The genetic algorithm is the basic framework, and combines with certain factors. Literature [14] studies the influence of the transportation modes on the aspect of traveling time and condition. The two factors are the two important factors to set up the developed algorithm. Literature [18] studies the condition and method for tourists to choose vehicle for traveling by tourists' preference, which mainly discusses the factors that influence tourists' decision on the aspect of interests, and then set up the algorithm model. Literature [28] studies on how to design the individualized tour route by the heuristic algorithm aiming at different types of tourist groups, mainly emphasizes on tourists' interests and individualized service. And it is also the emphasis of this research.

It is significant to analyze the previous researches in the literature that support our research work. Meanwhile, compare the similar researches with our work and find out the differences. Literature [40] studies the tourism recommendation and tour route planning based on the location social media, and mainly focuses on the social media big data. The social media big data is a kind of dynamic data, which is different from the static text data in our research, also, the interest mining method is different. In literature [40], tourists' interests are mined from the social media and used to recommend tourist sights and routes. Similarly, the literature [51] sets up the spatial network for the tourist sights based on geographic label and social media data. The algorithm method to recommend the tourist sights and tour routes is different from our work, which emphasizes the influence of the social media data and the spatial network, while our work emphasizes the process of text data mining to match tourists' interests and the impact of the multivariate transportation modes on the tourism recommendation. Literature [41] studies the tour route plan method by using real time traffic data, which is also a dynamic method. In our research work, the used basic data include geographic information data, traffic information data and tourist sight information data, and they are all static data, which is different from the literature [41]. Literature [49] uses the GPS big data that is generated by tourists' cellphones in the tour to set up the model and studies the tourists' activities and traces. This is a GPS big data research method. Our research work emphasizes the study on urban tour route in the range of the downtown area of a tourism city. The tourists' activities and traces are both in the city, not considering the GPS data. The tour route algorithm in the literature [52] is based on the improved heuristic searching. It judges whether the searched location conforms to the optimal condition or not, and it is set as the principle to do further searching. This core of the algorithm is the confirming the heuristic searching function and avoiding the unnecessary searching process. Thus, in this method, the optimal solution will be obtained by the sub-optimal ones will be neglected. In the recommendation system, both the optimal ans sub-optimal tour routes should be provided for tourists, thus, this method has certain disadvantage. Our research work is different from the literature [52], and it is a global searching process combining with geographic information data and traffic information data to get the globally optimal solution, thus, it is suitable for the embedded algorithm in the recommendation system.

Compared with the other important methods such as the methods mentioned in literature [29], [55], the proposed method has different principle, process and purpose. LSTM is a long short-term memory network which could be used in dealing with sequential data, and it is especially useful and effective to predict the important issues that have long time interval and delay in time series. The LSTM method is used in natural language processing such as text mining and processing. By training plentiful and abundant data, it sets up the network model with long short-term memory function, which could overcome the problem of traditional RNN on the aspect of long-term dependencies. Also, it needs relatively complex training process and time. When LSTM is used in text mining, its main purpose is to obtain the deep semantic, for example, text sentiment analysis and knowledge acquisition. The proposed text mining algorithm is based on encyclopedia text word frequency statistical analyzing algorithm, which figures out the matching degree between tourists' interest words and the tourist sight feature encyclopedia big data. The matching degree is the critical element on confirming the tendency degree of the tourists to the tourist sight clusters. On the aspect of setting up the algorithm model, the proposed method of the research doesn't need complex training data to form the neural network model. Based on the statistical mining, the interest tendency matching degree could be obtained. On the aspect of the purpose on text mining, LSTM tends to mine the deep semantics of the text, while the proposed method faces the general public tourists. To obtain the direct demands and interest data are the main purpose of the proposed method, while the deep semantics or emotional semantics of the text itself are not the target. On the aspect of setting up the recommender system, the proposed method has relatively smaller computational cost than the LSTM, it doesn't need complex training process to get the matching degree, which could save the calculating time and space.The literature [29] sets up the SQL injection attack detection method based on LSTM neural network. By combining with the LSTM, it could rise the accuracy of the SQL injection attack detection while decrease the error rate in the intelligent transportation system. It has more advantages than the traditional learning method. The literature [55] brings forward a fusion model based on LSTM network and CNN deep learning methods, and it could greatly improve the precision and accuracy of text sentiment classification. The literature [29] and [55] are both based on LSTM neural network, and use plentiful and abundant training data to set up and train the model. Different from the two methods, the proposed method in the research uses the encyclopedia text big data to set up the statistical model to mine the matching degree between the tourists' interest labels and the tourist sight features. The proposed algorithm is relatively easy to implement, there is no need for large quantity of training data and adjusting parameters for the neural network, and it costs smaller calculating time and space. As the tourism recommender system should be highly efficient, effective, convenient and easy to be realized, according to the above analysis, this research did not consider using the LSTM method.

The structure, main content and the research methods include the following aspects.

The first section: Introduction. The research background, the significance of the research, the research questions, the objectives, the literature review and the paper structure are discussed. By comparing with the previous researches, the problems of the current tourist sight and tour route recommendation are analyzed. And the basic thought, principle, the used technology and the experiment design of the research are discussed.

The second section: The tourist interest data mining algorithm and the interest tourist sight mining algorithm. The tourist interest data mining algorithm is brought forward. By setting up the tourist sight clustering algorithm model, the clusters of the tourist sights in the research domain are generated, which provides the basic data for the interest mining and the precise tourist sight matching. The interest tourist sight mining algorithm is set up to obtain the tourist sights that best match tourists' interests.

The third section: The tour route algorithm based on the interest tourist sights and the multivariate transportation modes.

The motive nerve cell model based on the multivariate transportation mode matching keys is set up. And based on the nerve cell model, the tour route chain algorithm based on the tourist sight motive nerve cell is set up, which could output tour route chains combing with the geographic information data and the traffic information data under the condition of the multivariate transportation modes, and then the optimal tour route chains are obtained.

The fourth section: The sample experiment and analysis. Take the tourist sights and the leisure areas in the downtown area of the tourism city Leshan as the research range and the objects to get the basic experiment data. The developed algorithm is used to output the precisely matched tourist sights and the tour route chains under the multivariate transportation modes. The experiment sets the three commonly used tour route planning algorithms as the control group, and under the same experimental conditions, the results of the developed algorithm and the control group algorithms are obtained.

The fifth section: The analysis and discussion of the experimental results. The experimental results are discussed on the aspect of the experimental basic data, the tourist sight clustering results, the tourist sight matching and mining results, the tour route chain results, the comparison of the algorithms and the practicalness of the algorithm, and then get the conclusion.

The sixth section: The conclusions and the future work. The research work is concluded. The features and the advantages of the algorithm are analyzed, meanwhile, the limitation of the algorithm and the future work are brought forward.

II. TOURIST INTEREST TEXT MINING ALGORITHM AND INTEREST TOURIST SIGHT MINING ALGORITHM

Tourists need to make travel schedule before they visit a tourism city. The schedule's contents mainly include the knowledge and confirmation on interests, the knowledge and confirmation on the service provided by the tourism city and tourist sights, the time schedule of the tour, the cost of the tour, the transportation mode used in the tour,etc [17], [18]. Firstly, the knowledge and confirmation on interests is the core because tourists should primarily confirm the destination and the psychological satisfaction tendency and extent, and this is determined by the objective law that tourists are the center for tourism activity. The satisfaction extent of tourists directly reflects their motive satisfaction and will influence the evaluations of tourists on the tourism city and tourist sights, and will indirectly influence the decision of visiting the tourism city of the subsequent tourists. Thus, the tourist sight and tour route recommendation system should meet tourists' interests. Secondly, the provided services and function attributes of the tourism city and tourist sights should be quantified, including the geographic location and distribution, feature attributes, function localization, geographic information service, traffic information service, etc. Tourism city and tourist sights' function attributes have direct influence on tourists' motive benefits. Different tourist sights have different function attributes, which will bring discrepancy on tourists' satisfaction. Then, the time schedule, travel cost and the choice of transportation mode will also influence the tourist sight and tour route recommendation [21]. The classification and quantity of the recommended tourist sight are not only determined by tourists' interests, but also by tourist sight distribution, time schedule and travel cost. Moreover, the choice of transportation mode will determine the travel mode. As it will be influenced by geographic information service and traffic information service, it will finally cause discrepancy on the recommendation of the optimal tour route [10], [19], [20]. According to the analysis, based on one tourism city's geographic space environment, the tourist interest text mining algorithm and interest tourist sight mining algorithm are set up.

A. TOURIST INTEREST TEXT MINING ALGORITHM

Tourist interest Text mining is based on the knowledge and confirmation on interests, the knowledge and confirmation on the service provided by the tourism city and tourist sights, the time schedule of the tour, the cost of the tour. The knowledge and confirmation on interests include the tourist sight function attributes, time schedule of visiting the tourist sight, the basic cost of visiting the tourist sight and tourist sight attraction [8], [9], [11]. Since different tourist sights provide discrepant services, their matching degrees with tourists' interests are also discrepant. By setting up the matching degree algorithm of tourists' interests and tourist sight quantitative service, the affinity degree of tourists' interests and tourist sight feature attributes could be confirmed.

1) THE FOUNDATION OF THE IMPROVED TOURIST SIGHT CLUSTERING ALGORITHM

Tourist sights in the tourism city have the features on diversification, homogeneity and heterogeneity, spatial discreteness, etc [48]. To cluster the city tourist sights is the basis of setting up the tourist sight clustering algorithm model.

Def 1.1: Tourist sight domain C. Choose certain quantity of city tourist sights which are typical and valuable and group them into one set to be clustered, and this set is called the tourist sight domain C. The domain C is used to store the selected typical tourist sights for the research. In order to confirm that the tourists could get the best travel experience, the quantity of tourist sights in the tourist sight domain C should be smaller than the total quantity of all the city's tourist sights. Set the quantity of the tourist sights in set C is $n, n > 0, n \in Z^+$. Each tourist sight in set C is noted as $c_{(i)}$, and then $i \in (0, n] \in Z^+$.

Def 1.2: Tourist sight cluster $\mathbf{C}_{(u)}$ and tourist sight meta data $c_{(u,v)}$. Cluster n quantity of typical tourist sights $c_{(i)}$ in the set C according to tourist sights' feature attributes into p quantity of clusters $\mathbf{C}_{(u)}$. Each cluster is noted as $c_{(i)} \sim c_{(u,v)}$. And then, set the quantity of each cluster $\mathbf{C}_{(u)}$ as $n_{(u)}$,and $u \in (0,p] \in Z^+, v \in (0, n_{(u)}] \in Z^+$.The tourist sight cluster $\mathbf{C}_{(u)}$ represents a sort of tourist sights gathered and constrained by certain criterion. It is the basis for the tourist interest data mining to get the precisely matched tourist sights.

The criterion for setting up the cluster $\mathbf{C}_{(u)}$ and the tourist sight clustering algorithm are as follows.

Step 1 Set up the tourist sight feature attribute label vector $\mathbf{c}_{(i)}$. The label vector that is composed of s quantity of keywords $w_{(i,t)}$ determined by one tourist sight $c_{(i)}$ text definition representing its feature attributes is called the tourist sight feature attribute label vector $\mathbf{c}_{(i)}$, and $t \in (0, s] \in Z^+$. The vector $\mathbf{c}_{(i)}$ is used to store the tourist sight text labels. Its elements are words and phrases. Two elements are different with each other and they are equal. The elements are used to describe the tourist sights' features.

According to the definition, $\mathbf{c}_{(i)} = (w_{(i,1)}, w_{(i,2)}, \ldots, w_{(i,t(i))}, \ldots, w(i, s))$, an it relates to one point of the tourist sight domain C. The keyword $w_{(i,t)}$ is the No. t element value in the vector $\mathbf{c}_{(i)}$. As to each tourist sight $c_{(i)}$, define the dimension of label vector $\mathbf{c}_{(i)}$ as $1 \times s$ and it is full rank $rank(\mathbf{c}(i)) = s$. And arbitrary two vectors meet the condition $\forall w_{(i,t)} \neq \forall w(i, \neg t)$.

Step 2 Set up the criterion $D(c_{(i)}, c_{(j)})$ for the clustering on tourist sight feature attribute. Set the feature attribute matching threshold value as $T_{(c_{(i)},c_{(j)})}$. Set up the criterion $D(c_{(i)}, c_{(j)})$ algorithm as follows.

Sub-step 1 Randomly search arbitrary one tourist sight $\forall c_{(i)}$ of n elements in the set C and obtain the tourist sight feature attribute label vector $\mathbf{c}_{(i)} = (w_{(j,1)}, w_{(j,2)}, \ldots, w_{(j,t_{(2)})}, \ldots, w_{(j,s)})$.

Sub-step 2 Randomly search arbitrary one tourist sight $\forall c_{(j)}$ other than the element $c_{(i)}$ of $n-1$ elements in the set C and obtain the tourist sight feature attribute label vector $\mathbf{c}_{(j)} = (w_{(j,1)}, w_{(j,2)}, \ldots, w_{(j,t_{(2)})}, \ldots, w_{(j,s)})$.

Sub-step 3 Set up the extreme matching value δ of label vector. Define that the maximum quantity of the matching labels of the feature attribute label vector elements between the tourist sight $\forall c_{(i)}$ and $\forall c_{(j)}$ as the extreme matching value δ of the two label vectors. The extreme matching value δ of label vector reflects the similarity of the two tourist sights, and it is an important parameter to set up tourist sight clusters. Here is the process of setting up the extreme matching value δ algorithm of label vector.

(1) Set the initial value $\delta_{(1)} = 0$. Take the first element $w_{(i,1)}$ in the label vector $\mathbf{c}_{(i)}$ of the tourist sight $c_{(i)}$.

① Compare $w_{(i,1)}$ with $w_{(j,1)}$.

i) If $w_{(i,1)} = w_{(j,1)}$, this feature attributes of $c_{(i)}$ and $c_{(j)}$ are the same, $\delta_{(1)} = \delta_{(1)} + 1$, stop searching, jump to step (2);

ii) If $w_{(i,1)} \neq w_{(j,1)}$, this feature attributes of $c_{(i)}$ and $c_{(j)}$ are different, $\delta_{(1)}$ remain unchanged, jump to step②.

② Compare $w_{(i,1)}$ with $w_{(j,2)}$.

i) If $w_{(i,1)} = w_{(j,2)}$, this feature attributes of $c_{(i)}$ and $c_{(j)}$ are the same, $\delta_{(1)} = \delta_{(1)} + 1$, stop searching, jump to step (2);

ii) If $w_{(i,1)} \neq w_{(j,2)}$, this feature attributes of $c_{(i)}$ and $c_{(j)}$ are different, $\delta_{(1)}$ remain unchanged, jump to step③.

③ Compare $w_{(i,1)}$ with $w_{(j,t_{(2)})}$ by the same method from step ① ∼② until $t_{(2)} = s$, and $t_{(2)} \sim (0, s] \in Z^+$. And obtain the related extreme matching value $\delta_{(1)}$ for the first element $w_{(i,1)}$ in tourist sight $c_{(i)}$ label vector $\mathbf{c}_{(i)}$.

(2) Set the initial value $\delta_{(2)} = 0$. Take the second element $w_{(i,2)}$ in the label vector $\mathbf{c}_{(i)}$ of the tourist sight $c_{(i)}$. Compare $w_{(i,2)}$ with $w_{(j,t_{(2)})}$ by the same method from step ① ∼ ③ until $t_{(2)} = s$, and $t_{(2)} \sim (0, s] \in Z^+$. And obtain the related extreme matching value $\delta_{(2)}$ for the second element $w_{(i,2)}$ in tourist sight $c_{(i)}$ label vector $\mathbf{c}_{(i)}$.

(3) Set the initial value $\delta_{(t_{(1)})} = 0$. Take the No. $t_{(1)}$ element $w_{(i,t_{(1)})}$ in the label vector $\mathbf{c}_{(i)}$ of the tourist sight $c_{(i)}$. Compare $w_{(i,t_{(1)})}$ with $w_{(j,t_{(2)})}$ by the same method from step ① ∼ ③ until $t_{(2)} = s$, and $t_{(2)} \sim (0, s] \in Z^+$. And obtain the related extreme matching value $\delta_{(t_{(1)})}$ for the No. $t_{(1)}$ element $w_{(i,t_{(1)})}$ in tourist sight $c_{(i)}$ label vector $\mathbf{c}_{(i)}$.

(4) Traverse $t_{(1)} \sim (0, s] \in Z^+$ until $t_{(1)} = s$ by the same method of step(15)~(3). And obtain all the extreme matching values $\delta_{(t_{(1)})}$ for all the elements in tourist sight $c_{(i)}$ label vector $\mathbf{c}_{(i)}$.

(5) Accumulate and calculate the extreme matching value δ for the tourist sight $c_{(i)}$ label vector $\mathbf{c}_{(i)}$, and the value conforms to the result $\delta = \sum_{t_{(1)}=1}^{s} \delta_{(t_{(1)})}$. Set the tourist sight feature attribute clustering criterion as $D(c_{(i)}, c_{(j)}) = \delta$.

Sub-step 4 Set up the tourist sight clustering algorithm criterion. When arbitrary tourist sight $\forall c_{(i)}$ and $\forall c_{(j)}$ in set C meet arbitrary one of the following conditions, the tourist sight $c_{(i)}$ and $c_{(j)}$ could be clustered into one $\mathbf{C}_{(u)}$. In order to set up the algorithm, if the two tourist sights meet arbitrary one condition, note $+c_{(i)}c_{(j)}$, or note $-c_{(i)}c_{(j)}$.

(1) As to arbitrary tourist sight $\forall c_{(i)},\ c_{(j)} \in C_{(u)}$, there is: $D_{(c_{(i)},c_{(j)})^{-1}} \leq T_{(c_{(i)},c_{(j)})}$;

(2) As to arbitrary tourist sight $\forall c_{(i)} \in C_{(u)}$, there is: $[(k-1)^{-1} \cdot \sum_j D(c_{(i)}, c_{(j)})]^{-1} \leq T_{(c_{(i)},c_{(j)})}$;

(3) As to arbitrary tourist sight $\forall c_{(i)},\ c_{(j)} \in C_{(u)}$, there is: $[k^{-1}(k-1)^{-1} \cdot \sum_i \sum_j D(c_{(i)}, c_{(j)})]^{-1} \leq T_{(c_{(i)},c_{(j)})}$, and there must be $D_{(c_{(i)},c_{(j)})^{-1}} \leq T_{(c_{(i)},c_{(j)})}$;

(4) As to arbitrary tourist sight $\forall c_{(i)} \in C_{(u)}$, there exists a sample tourist sight $c_{(j)}$ in the cluster $\mathbf{C}_{(u)}$ that meets the condition $D_{(c_{(i)},c_{(j)})^{-1}} \leq T_{(c_{(i)},c_{(j)})}$.

According to the definition, tourist sight cluster $\mathbf{C}_{(u)}$ and tourist sight $c_{(i)}$ should meet the following conditions:

(1) Arbitrary cluster is nonempty set: $\forall C_{(u)} \neq \emptyset$;

(2) The union set with the p amount of clusters $\mathbf{C}_{(u)}$ is C:

$$C_{(1)} \cup C_{(2)} \cup \ldots \cup C(p) = C;$$

(3) The intersection set of arbitrary two clusters is the null set: $\forall C(u(15)) \cap \forall C(u(2)) = \emptyset, 0 < u(15) \neq u(2) \leq p, u(15), u(2) \in Z^+$;

(4) Arbitrary tourist sight $c_{(i)}$ belongs to and only belongs to one tourist sight cluster $\mathbf{C}_{(u)}$, and it has the exclusive code $c_{(i)} \sim c_{(u,v)}$; Set up the subordinate degree function $\xi(u, i) = \xi C_{(u)}(c_{(i)})$, and it represents the subordinate relationship between tourist sight $c_{(i)}$ label vector $\mathbf{c}_{(i)}$ and tourist sight cluster $\mathbf{C}_{(u)}, 0 < u \leq p$, and there is the quantitative relationship between tourist sight and its cluster in Formula (15). According to the definition, if the subordinate degree of one tourist sight $c_{(i)}$ to the cluster $\mathbf{C}_{(u)}$ is 1, and its subordinate degree to other clusters $\neg C_{(u)}$ is 0. The subordinate degree function is used to set up the tourist sight cluster subordinate degree matrix $\boldsymbol{\xi}_{(p\times \max n_{(i)})}$. The subordinate degree represents whether the single tourist sight $c_{(i)}$ belongs to the cluster $\mathbf{C}_{(u)}$ or not. If the degree is 1, the tourist sight $c_{(i)}$ belongs to the cluster $\mathbf{C}_{(u)}$. If the degree is 0, the tourist sight $c_{(i)}$ doesn't belong to the cluster $\mathbf{C}_{(u)}$.

$$\xi C_{(u)}(c_{(i)}) = \xi(u, i) = \begin{cases} 0, & c_{(i)} \in C_{(u)}, \\ 1, & c_{(i)} \notin C_{(u)}. \end{cases} \quad (1)$$

Def 1.3: Cluster generation point $\Delta c_{(i)}$, homogeneous cluster and heterogeneous cluster. In the process of dividing tourist sight domain C into p quantity of clusters $\mathbf{C}_{(u)}$, the p quantity of tourist sights which have different feature attributes and each can represent one cluster are confirmed by the algorithm, and each of the tourist sight is called the cluster generation point, noted as $\Delta c_{(i)}$. The generation point $\Delta c_{(i)}$ is the starting point of the cluster $\mathbf{C}_{(u)}$. Starting from the point $\Delta c_{(i)}$, the other points $\neg\Delta c_{(i)}$ are judged by the clustering criterion $D(c_{(i)}, c_{(j)})$ on whether they are in the same cluster $\mathbf{C}_{(u)}$ with the point $\Delta c_{(i)}$. Start from the point $\Delta c_{(i)}$ and search another tourist sight $c_{(j)}$ and then make the following judgment.

(1) If the tourist sight $c_{(i)}$ and the tourist sight $c_{(j)}$ meets the condition of $+c_{(i)}c_{(j)}$, then the two tourist sights $c_{(i)}$ and $c_{(j)}$ are defined in the homogeneous cluster, the subordinate degree of the tourist sight $c_{(j)}$ to the cluster $\mathbf{C}_{(u)}$ with the point $\Delta c_{(i)}$ is 1;

(2) If the tourist sight $c_{(i)}$ and the tourist sight $c_{(j)}$ meets the condition of $-c_{(i)}c_{(j)}$, then the two tourist sights $c_{(i)}$ and $c_{(j)}$ are defined in the heterogeneous cluster, the subordinate degree of the tourist sight $c_{(j)}$ to the cluster $\mathbf{C}_{(u)}$ with the point $\Delta c_{(i)}$ is 0.

Def 1.4: Subordinate degree matrix $\boldsymbol{\xi}_{(p\times \max n_{(i)})}$. Store the n quantity of tourist sights in domain C into the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$, and the storage principle meet the following conditions.

(1) The dimension of the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$ is $p \times \max n_{(i)}$, that is, the matrix row is p and the matrix column is $\max n_{(i)}$;

(2) The quantity of the tourist sight meets the condition that $0 < n \leq p \times \max n_{(i)}$;

(3) The minimum element code of the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$ is 1, and the maximum element code is $p \times \max n_{(i)}$. The code sequence is arranged from the minimum code to the maximum code of each row and column.

(4) The storage location of the tourist sight $c_{(i)}$ is in the row $[(i-1)/\max n_{(i)}]+1$ and column $[(i-1) \bmod (\max n_{(i)})]+1$;

(5) The row and column of the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$ are all full rank.

In the process of searching other tourist sights $c_{(j)}$ from the cluster generation point $\Delta c_{(i)}$ of cluster $\mathbf{C}_{(u)}$, each tourist sight $c_{(j)}$ subordinate degree to the cluster $\mathbf{C}_{(u)}$ in the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$ is confirmed, $0 < j \neq i \leq n$, which forms the matrix with the subordinate degree value and $p \times \max n_{(i)}$ dimension. This matrix is called the subordinate degree matrix, noted as $\boldsymbol{\xi}_{(p\times \max n_{(i)})}$. The subordinate degree matrix reflects the storage and distribution condition for one cluster of tourist sights, with the basic element 1 and 0. The element 1 represents the distribution for the tourist sights of cluster $\mathbf{C}_{(u)}$ in the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$, that is, the element location of 1. Oppositely, the tourist sights of the data 0 element locations are not in the cluster $\mathbf{C}_{(u)}$.

Step 3 Set up the tourist sight clustering algorithm. Based on tourist sight feature attribute, label vector and clustering criterion, the tourist sight clustering algorithm is set up, which is used to form clusters on the n quantity of tourist sights in the domain C. The algorithm is as follows.

Sub-step 1 Set up the tourist sight storage matrix $\mathbf{C}_{(p\times \max n_{(i)})}$.

Store each tourist sight $c_{(j)}$ into the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$;

Sub-step 2 Confirm the initial generation point $\Delta c_{(1)}$. And randomly choose one tourist sight $c_{(i)}$ in the matrix $\mathbf{C}_{(p\times \max n_{(i)})}$ as the initial generation point $\Delta c_{(1)}$, relating to the cluster $C_{(1)}$;

Sub-step 3 Traverse to search the tourist sight $c(j_{(1)})$, and $0 < j_{(1)} \neq i \leq n$. The row coordinate of the searched tourist sight is $[(j_{(1)} - 1)/\max n_{(i)}] + 1$, and the column coordinate of the searched tourist sight is $[(j_{(1)} - 1) \bmod (\max n_{(i)})] + 1$. The initial searching value meets the condition $j_{(1)} \neq i$.

(1) Set the initial searching value as $j_{(1)} = 1$. Calculate the subordinate degree according to the tourist sight clustering criterion.

① If the tourist sight $c(j_{(1)})$ and $\Delta c_{(1)}$ are in the homogeneous cluster, then $+c(j_{(1)})\Delta c_{(1)}$. The subordinate degree of $c(j_{(1)})$ to the cluster $C_{(1)}$ is 1, jump to the step (2);

② If the tourist sight $c(j_{(1)})$ and $\Delta c_{(1)}$ are in the heterogeneous cluster, then $-c(j_{(1)})\Delta c_{(1)}$. The subordinate degree of $c(j_{(1)})$ to the cluster $C_{(1)}$ is 0, jump to Sub-step 4.

(2) Set $j_{(1)} = 2, 3, \ldots, i - 1, i + 1, \ldots n$. Search other tourist sight $c(j_{(1)})$ by the same method in step (15). The homogeneous tourist sight is absorbed into the cluster $C_{(1)}$ while the heterogeneous tourist sight is used to form a new cluster.

(3) Generate the subordinate degree matrix $\boldsymbol{\xi}_1(p\times \max n_{(i)})$ with the initial generation point $\Delta c_{(1)}$.

Sub-step 4 Define this heterogeneous tourist sight $c(j_{(1)})$ as the second cluster generation point $\Delta c_{(2)}$, relating to the cluster $C_{(2)}$. Set up the new tourist sight matrix $\mathbf{C}_{(p\times \max n_{(i)})}$ and traverse to search tourist sight $c(j_{(2)})$. The row coordinate of the searched tourist sight is $[(j_{(2)} - 1)/\max n_{(i)}] + 1$, and the column coordinate of the searched tourist sight is $[(j_{(2)} - 1) \bmod (\max n_{(i)})] + 1$.

(1) Set the initial searching value as $j_{(2)} = 1$. Calculate the subordinate degree according to the tourist sight clustering criterion. The confirmed generation tourist sights will not be searched any more.

① If the tourist sight $c(j_{(2)})$ and $\Delta c_{(2)}$ are in the homogeneous cluster, then $+c(j_{(2)})\Delta c_{(2)}$. The subordinate degree of $c(j_{(2)})$ to the cluster $C_{(2)}$ is 1, jump to the step (2);

② If the tourist sight $c(j_{(2)})$ and $\Delta c_{(2)}$ are in the heterogeneous cluster, then $-c(j_{(2)})\Delta c_{(2)}$. The subordinate degree of $c(j_{(2)})$ to the cluster $C_{(2)}$ is 0, jump to the Sub-step 5.

(2) Set $j_{(2)} = 2, 3, \ldots, i - 1, i + 1, \ldots n$. Search other tourist sight $c(j_{(2)})$ by the same method in step (15). The homogeneous tourist sight is absorbed into the cluster $C_{(2)}$ while the heterogeneous tourist sight is used to form a new cluster.

(3) Generate the subordinate degree matrix $\boldsymbol{\xi}_2(p\times \max n_{(i)})$ with the initial generation point $\Delta c_{(2)}$.

Sub-step 5 Continue searching. Search other cluster generation point $\Delta c_{(u)}$, related cluster $\mathbf{C}_{(u)}$ and subordinate degree matrix $\boldsymbol{\xi}_u(p \times \max n_{(i)})$. Traverse the $u \sim (0, p]$ and confirm p quantity of clusters and p quantity of subordinate degree matrix. The quantity of tourist sight in each cluster $\mathbf{C}_{(u)}$ is $n_{(u)}$.

The following pseudo-code is for the tourist sight clustering algorithm.

Algorithm 1 The Tourist Sight Clustering Algorithm

Step 1: Input label vector $\mathbf{c}_{(i)}$.
Step 2: Calculate the criterion $D(c_{(i)}, c_{(j)})$.
  Sub-step 1: Search arbitrary $\forall c_{(i)}$ and get its $\mathbf{c}_{(i)}$.
  Sub-step 2: Search another arbitrary $\forall c_{(j)}$ and get its $\mathbf{c}_{(j)}$.
  Sub-step 3: Calculate δ.
    For each $t_{(1)} \sim (0, s] \in Z^+$ and $t_{(2)} \sim (0, s] \in Z^+$ **until** $t_{(1)} = s$ and $t_{(2)} = s$:
    For each $\delta_{(t_{(1)})} = 0$: Compare $w_{(i,2)}$ and $w_{(j,t_{(2)})}$.
      Accumulate $\delta_{(t_{(1)})} = \delta_{(t_{(1)})} + 1$
  Sub-step 4: Set clustering criterion.
Step 3: The clustering process.
  Sub-step 1: Set up $\mathbf{C}_{(p\times \max n_{(i)})}$.
  Sub-step 2∼5: **For each** $\Delta c_{(u)}$: $j(u) = 2, 3, \ldots, i - 1, i + 1, \ldots n$
    Confirm $\Delta c_{(u)}$, judge whether $c(j(u))$ and $\Delta c_{(u)}$ are in the same cluster.
    Confirm $\boldsymbol{\xi}_u(p \times \max n_{(i)})$ and $\mathbf{C}_{(u)}$.

2) INTEREST TEXT MINING ALGORITHM

The tourist interest Text mining algorithm is set up based on tourist sight clusters. The principles for tourists to choose tourist sights include their interests and the feature attributes that tourist sights can provide for tourists. Thus, the basis of the tourist interest Text mining algorithm is to confirm tourists' interests and tourist sights' feature attributes.

Def 2.1: Feature attribute matching factor $k_{(r)}$. The factors that are determined by the knowledge and cognition on tourist sight function, tourist sight traveling time schedule, basic traveling cost and tourist sight attraction index that influence the matching relationship of tourists' interests and tourist sights' feature attributes are called feature attribute matching factors $k_{(r)}, 0 < r \leq 4, r \in Z^+$. The factor $k_{(r)}$ is the key factor to mine the tourists' interests and obtain the precisely matched tourist sights. Also, it is the critical factor for the intelligent recommendation system to set the input conditions for the tourists. By comparing the demand conditions of the tourists with the conditions that the tourist sights could provides, the optimal value for the matching function is obtained, and the tourist sights to be visited are also obtained. And each factor has its own quantization algorithm.

Def 2.2: Interest feature data vector $\mathbf{W}_{(t)}$. The union that is composed by s quantity of keywords $w_{(i,t)}$ representing tourist sights' feature attributes in the tourist sight feature attribute label vector $\mathbf{c}_{(i)}$ is called the interest feature data vector $\mathbf{W}_{(t)}$. Its element is noted as $W_{(t)}$, and then $\mathbf{W}_{(t)} = (W_{(1)}, W_{(2)}, \ldots, W_{(t)}, \ldots, W_{(ns)})$. Since the labels in the different vectors $\mathbf{c}_{(i)}$

could be repetitive, thus, the capacity of the union set of the n quantity of the vectors $\mathbf{c}_{(i)}$ could be smaller than the total sum of the n quantity of the vectors $\mathbf{c}_{(i)}$ capacities. According to the clustering algorithm, the union capacity of n quantity of tourist sight feature attribute label vectors $\mathbf{c}_{(i)}$ is far smaller than the theoretical capacity, that is $0 < t \ll ns, t, ns \in Z^+$. Arbitrary element of the interest feature data vector $\mathbf{W}_{(t)}$ is nonzero, that is $W_{(t)} \neq 0$, and the vector $\mathbf{W}_{(t)}$ is full rank. Set the vector $\mathbf{W}_{(t)}$ contains q quantity of elements, that is $rank(\mathbf{W}_{(t)}) = q$.The vector $\mathbf{W}_{(t)}$ is used to store all the feature labels of tourist sights contained in the domain C. It is the basic data set to mine tourists' interest data and obtain the precisely matched tourist sights. According to the definition, the interest feature data vector $\mathbf{W}_{(t)}$ meets the Formula (2). Formula (2) represents the union set of the n quantity of the vectors $\mathbf{c}_{(i)}$, which meets the condition that the arbitrary label vector $\mathbf{c}_{(i)}$ is the non-empty set.

$$\begin{cases} \mathbf{W}_{(t)} = \mathbf{c}_{(1)} \cup \mathbf{c}_{(2)} \ldots \cup \mathbf{c}_{(i)} \cup \mathbf{c}_{(n)}, \\ s.t. \ \mathbf{c}_{(i)} \neq \emptyset. \end{cases} \tag{2}$$

The generation algorithm of the interest feature data vector $\mathbf{W}_{(t)}$ set up by the definition 2.2 and Formula (2) is as follows.

Step 1 Initialize the empty set $\mathbf{W}^0_{(t)}$, and its capacity is ns.

Step 2 Store elements into the feature attribute label vector $\mathbf{c}_{(1)}$. Store the s quantity of labels in vector $\mathbf{c}_{(1)}$ into the first s elements in the empty vector $\mathbf{W}^0_{(t)}$ in the sequence of i element, and form the initial vector $\mathbf{W}_{(t)}$.

Step 3 Search the feature attribute label vector $\mathbf{c}_{(2)}$ and judge the relationship of the vector $\mathbf{c}_{(2)}$ elements and $\mathbf{W}^0_{(t)}$.

Sub-step 1 Compare the first element $w_{(2,1)}$ of the vector $\mathbf{c}_{(2)}$ with the $W_{(t)}$ element of the vector $\mathbf{W}_{(t)}$.

(1) Compare with the first element of the vector $\mathbf{W}_{(t)}$.

① If $w_{(2,1)} = W_{(1)}$, delete the element $w_{(2,1)}$;

② If $w_{(2,1)} \neq W_{(1)}$, jump to the step (2).

(2) Compare with the second element of the vector $\mathbf{W}_{(t)}$.

① If $w_{(2,1)} = W_{(2)}$, delete the element $w_{(2,1)}$;

② If $w_{(2,1)} \neq W_{(2)}$, jump to the step (3).

(3) Compare with the No. t element of the vector $\mathbf{W}_{(t)}$, $t = 3, 4, \ldots, s$. Search until $t = s$:

① If $w_{(2,1)} = W_{(s)}$, delete the element $w_{(2,1)}$, and the element $w_{(2,1)}$ is not the newly added one for the vector $\mathbf{W}_{(t)}$;

② If $w_{(2,1)} \neq W_{(2)}$, store $w_{(2,1)}$ into the No. $s+1$ element $W_{(s+1)}$ in the vector $\mathbf{W}_{(t)}$, and the element $w_{(2,1)}$ is the newly added one for the vector $\mathbf{W}_{(t)}$.

Sub-step 2 Compare the second element $w_{(2,2)}$ in the vector $\mathbf{c}_{(2)}$ with the element $W_{(t)}$ in the vector $\mathbf{W}_{(t)}$, and judge whether the element $w_{(2,2)}$ is the newly added one for the vector $\mathbf{W}_{(t)}$. If it is not the newly added one for the vector $\mathbf{W}_{(t)}$, delete $w_{(2,2)}$; If it is the newly added one for the vector $\mathbf{W}_{(t)}$, store the element $w_{(2,2)}$ into the No. $s+2$ element $W(s+2)$ in the vector $\mathbf{W}_{(t)}$.

Sub-step 3 Compare the No. t element $w(2, t)$ in the vector $\mathbf{c}_{(2)}$ with the element $W_{(t)}$ in the vector $\mathbf{W}_{(t)}$, and judge whether the element $w(2, t)$ is the newly added one for the vector $\mathbf{W}_{(t)}$. If it is not the newly added one for the vector $\mathbf{W}_{(t)}$, delete $w(2, t)$; If it is the newly added one for the vector $\mathbf{W}_{(t)}$, then store the element $w(2, t)$ into the No. $s+t$ element $W(s+t)$ in the vector $\mathbf{W}_{(t)}$.

Sub-step 4 Traverse $t \sim (1, s] \in Z^+$ until all of the elements of the vector $\mathbf{c}_{(2)}$ are searched and compared. Update the vector $\mathbf{W}_{(t)}$.

Step 4 Search the feature attribute label vector $\mathbf{c}_{(3)}$ via the Step 3, and then judge the vector $\mathbf{c}_{(3)}$ elements with the $\mathbf{W}^0_{(t)}$ elements. And then store the $\mathbf{c}_{(3)}$ elements that conform to the condition into the vector $\mathbf{W}_{(t)}$. After the searching process, update the vector $\mathbf{W}_{(t)}$.

Step 5 Search the feature attribute label vector $\mathbf{c}_{(i)}$ via the Step 3, and judge the vector $\mathbf{c}_{(i)}$ elements with $\mathbf{W}^0_{(t)}$ elements. Store the $\mathbf{c}_{(i)}$ elements that conform to the condition into the vector $\mathbf{W}_{(t)}$. After the searching process, update the vector $\mathbf{W}_{(t)}$.

Step 6 Traverse $i \sim (0, n] \in Z^+$ until all the vectors $\mathbf{c}_{(i)}$, and finally obtain the full rank vector $\mathbf{W}_{(t)}$ which contains q elements, $rank(\mathbf{W}_{(t)}) = q$.

Step 7 Traverse to search all elements $W_{(t)}$ of the vector $\mathbf{W}_{(t)}$ and replace its elements. Store elements $W_{(t)}$ in the sequence of p tourist sight clusters $\mathbf{C}_{(u)}$, shown in Figure 1. Clusters $\mathbf{C}_{(u)}$ contain feature attribute labels respectively. Set that the cluster which has the maximum quantity of the feature attribute label $W_{(t)}$ contains $\max p_{(u)}$ quantity of labels $W_{(t)}$.

Interest feature data vector $\mathbf{W}_{(t)}$ is the basis for tourists to confirm the travel tendency and for the smart recommendation system to choose the optimal tourist sights. Arbitrary element $\forall W_{(t)}$ of the vector $\mathbf{W}_{(t)}$ is taken as an interest label stored in the data list in the text format. The interest feature data vector $\mathbf{W}_{(t)}$ is used to set up the interest text mining algorithm based on tourist sight text knowledge. The feature attribute matching factor $k_{(1)}$ is confirmed, which represents the tourists' knowledge and cognition on the tourist sights' function. Meanwhile, tourists make requests on the traveling time $k_{(2)}$, the basic travel cost $k_{(3)}$ and the tourist sight attraction index $k_{(4)}$, and along with the feature attribute matching factor $k_{(1)}$, they form the tourist interest vector $\mathbf{K}_{(r)}$. This vector $\mathbf{K}_{(r)}$ is the basis for matching the optimal tourist sights.

Def 3.1: Tourist basic interest matrix $\mathbf{W}_{(p\times \max p_{(u)})}$. According to the interest feature data vector $\mathbf{W}_{(t)}$, the tourists choose certain quantity of feature attribute labels $W_{(t)}$ in each cluster $\mathbf{C}_{(u)}$ according to their sequence in the vector. And then, the labels $W_{(t)}$ are stored in the $p \times \max p_{(u)}$ dimension matrix by certain principle. This matrix is called the tourist basic interest matrix $\mathbf{W}_{(p\times \max p_{(u)})}$. The tourist basic interest matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ is the basis matrix for the interest text mining and outputting the knowledge and cognition on the tourist sights' function. The matrix matches tourist sight feature attributes and text knowledge and the feature attribute matching factor $k_{(1)}$ is confirmed. The arbitrary one element in the matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ represents tourists' one interest point. The same row represents the same cluster's tourist sights interest points. From the matrix $\mathbf{W}_{(p\times \max p_{(u)})}$, the cognition and tendency of the tourists on the tourist sights could

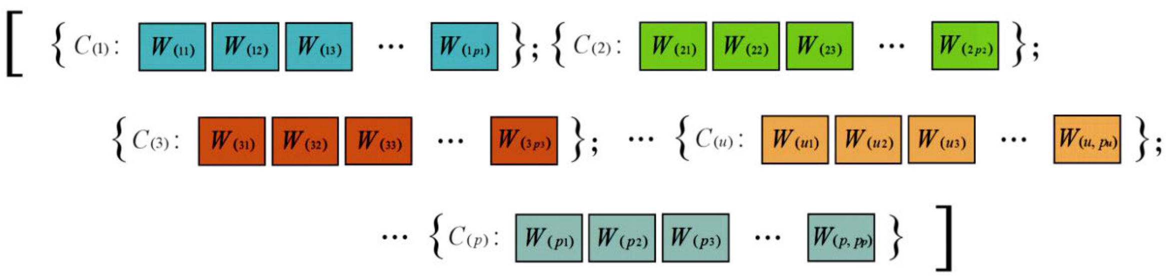


FIGURE 1. The vector $\mathbf{W}_{(t)}$ structure and its element distribution.

be concluded on the aspect of feature attributes. According to the definition, the principle for the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ to store the selected interest labels is as follows.

(1) The matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ is the nonzero matrix with p rows and $\max p_{(u)}$ columns;

(2) The No. u row is used to store the selected cluster $\mathbf{C}_{(u)}$ labels;

(3) Set the quantity of the selected cluster $\mathbf{C}_{(u)}$ labels as $e_{(u)}$. Thus, from the first element to the No. $e_{(u)}$ one in the No. u row, the cluster $\mathbf{C}_{(u)}$ labels are stored in sequence. The elements from the No. $e_{(u)}+1$ to the No. $\max p_{(u)}$ are set 0;

(4) The row rank and the column rank of the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ may not be full, that is, certain row or column's elements may all be 0. Thus, the row rank meets the condition $row \sim rank(\mathbf{W}_{(p\times\max p_{(u)})}) \leq p$, the column rank meets the condition $column \sim rank(\mathbf{W}_{(p\times\max p_{(u)})}) \leq \max p_{(u)}$.

Def 3.2: Tourist basic interest label vector $\mathbf{W}_{u^{(1\times\max p(u))}}$. All of the nonzero rows u of the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ are defined as the tourist basic interest label vectors $\mathbf{W}_{u^{(1\times\max p(u))}}$. The nonzero vector $\mathbf{W}_{u^{(1\times\max p(u))}}$ represents the tourists' knowledge and cognition on the tourist sights in the cluster $\mathbf{C}_{(u)}$, that is, the interest labels and interest tendency that tourists wish to get from the tourist sight cluster $\mathbf{C}_{(u)}$. The arbitrary one element in the vector $\mathbf{W}_{u^{(1\times\max p(u))}}$ represents the one interest point of the tourists on the cluster $\mathbf{C}_{(u)}$. It is used to match the tourist sight feature attribute and obtain the factor $k_{(1)}$ under the conditions of the cluster. Formula (3) is the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ general expression.

Formula (4) is the confirmed one sort of the basic interest matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ according to the labels chosen by the tourists. In the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$, arbitrary nonzero row is the basic interest label vector $\mathbf{W}_{u^{(1\times\max p(u))}}$. As to the Formula (3), the element $W(u,t)$ of the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ represents the one interest label selected by the tourists, that is, the one interest point of the element $W(u,t)$ in the row cluster $\mathbf{C}_{(u)}$. The element could be a text label, or an empty element 0, in the meantime, the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ is the full ranked matrix in both row and column. The elements of the same row belong to the same cluster while the elements in different rows doesn't belong to the same cluster. Formula (4) is a kind of random example of the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$. The element $W_{(q)}$ represents the last element of the random example matrix $\mathbf{W}_{(p\times\max p_{(u)})}$, the footnote q is a random value. (3) and (4), as shown at the bottom of the next page.

Def 3.3: Interest label homogeneous vector $\mathbf{W}^{\wedge}_{(1\times m)}$. As to the related tourist sight cluster $\mathbf{C}_{(u)}$ of the No. u row vector $\mathbf{W}(1\times\max p_{(u)})$ in the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$, the union set of the interest tendency labels and the feature attribute label vector $\mathbf{c}_{(i)}$ of the tourist sight $c_{(i)} \sim c_{(u,v)}$ in the cluster $\mathbf{C}_{(u)}$ is called the interest label homogeneous vector $\mathbf{W}^{\wedge}_{(1\times m)}$.The vector $\mathbf{W}^{\wedge}_{(1\times m)}$ represents the choice of tourists on one cluster $\mathbf{C}_{(u)}$ interest points and the matching degree with one certain tourist sight $c_{(i)}$. The higher the matching degree is, the closer relationship between the tourist sight $c_{(i)}$ with the selected interest points on the aspect of feature attributes, the more easily the tourist sight meets the tourists' interests. The process to get the interest label homogeneous vector $\mathbf{W}^{\wedge}_{(1\times m)}$ is as follows.

Step 1 Set up the empty vector $\mathbf{W}^{0}_{(1\times m)}$, and its dimension is $1\times m$.

Step 2 Search the first row vector $\mathbf{W}_{1^{(1\times\max p_{(u)})}}$ of the matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ relating to the cluster $C_{(1)}$ and match the homogeneous elements.

Sub-step 1 Search the first element $W_{(11)}$ and the vector $\mathbf{c}_{(i)}$ elements.

(1) Compare $W_{(11)}$ and $w_{(i,1)}$:

① If $W_{(11)} = w_{(i,1)}$, then $W_{(11)}$ and $w_{(i,1)}$ are homogeneous labels. Store $W_{(11)}$ into the first element $\mathbf{W}^{0}_{(11)}$ of the vector $\mathbf{W}^{0}_{(1\times m)}$. End the comparison, jump to Sub-step 2;

② If $W_{(11)} \neq w_{(i,1)}$, then $W_{(11)}$ and $w_{(i,1)}$ are heterogeneous labels. Jump to step (2).

(2) Compare $W_{(11)}$ and $w_{(i,2)}$:

① If $W_{(11)} = w_{(i,2)}$, then $W_{(11)}$ and $w_{(i,2)}$ are homogeneous labels. Store $W_{(11)}$ into the first element $\mathbf{W}^{0}_{(11)}$ of the vector $\mathbf{W}^{0}_{(1\times m)}$. End the comparison, jump to Sub-step 2;

② If $W_{(11)} \neq w_{(i,2)}$, then $W_{(11)}$ and $w_{(i,2)}$ are heterogeneous labels. Jump to step (3).

(3) Compare $W_{(11)}$ and $w_{(i,t(i))}$, $t_{(i)} = 1, 2, \ldots, s$. The method is the same as step (15) and step (2). When $t_{(i)} = s$, if $W_{(11)} \neq w(i,s)$, the label $W_{(11)}$ and the labels of the vector $\mathbf{c}_{(i)}$ are all heterogeneous labels, jump to Sub-step 2.

Sub-step 2 Search the second element $W_{(12)}$ and the vector $\mathbf{c}_{(i)}$ elements. Traverse $w_{(i,t(i))}$, and compare the label $W_{(12)}$ with $w_{(i,t(i))}$ with the same method in Sub-step 1.

(1) If label $W_{(12)}$ and arbitrary element $\forall w_{(i,t(i))}$ in the vector $\mathbf{c}_{(i)}$ are homogeneous:

① If the first element of the vector $\mathbf{W}^0_{(1\times m)}$ meets the condition $\mathbf{W}^0_{(11)} = 0$, store the label $W_{(12)}$ into the first element $\mathbf{W}^0_{(11)}$ of the vector $\mathbf{W}^0_{(1\times m)}$. End the comparison, jump to Sub-step 3.

② If the first element of the vector $\mathbf{W}^0_{(1\times m)}$ meets the condition $\mathbf{W}^0_{(11)} \neq 0$, store the label $W_{(12)}$ into the second element $\mathbf{W}^0_{(12)}$ of the vector $\mathbf{W}^0_{(1\times m)}$. End the comparison, jump to Sub-step 3.

(2) If the label $W_{(12)}$ and all of the elements $w_{(i,t(i))}$ in the vector $\mathbf{c}_{(i)}$ are heterogeneous, jump to Sub-step 3.

Sub-step 3 Search the No. $p_{(u)}$ element $W_{(1,p(u))}$ and the vector $\mathbf{c}_{(i)}$ elements, $p_{(u)} = 3, 4, \ldots, \max p_{(u)}$. Traverse $w_{(i,t(i))}$, compare $W_{(1,p(u))}$ and $w_{(i,t(i))}$ with the same method as Sub-step 1. Output the related interest label homogeneous vector $\mathbf{W}^1_{(1\times m)}$ of the cluster $C_{(1)}$.

Step 3 Search the second row vector $\mathbf{W}_{2(1\times \max p(u))}$ of the matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ relating to the cluster $C_{(2)}$ and match the homogeneous elements. The method is the same as Step 2. Output the related interest label homogeneous vector $\mathbf{W}^2_{(1\times m)}$ of the cluster $C_{(2)}$.

Step 4 Search the No. u row vector $\mathbf{W}_{u(1\times \max p(u))}$ of the matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ relating to the cluster $\mathbf{C}_{(u)}$ and match the homogeneous elements. The method is the same as Step 2. Output the related interest label homogeneous vector $\mathbf{W}^u_{(1\times m)}$ of the cluster $\mathbf{C}_{(u)}$, $u = 3, 4, \ldots, p$.

Def 3.4: The tourist interest label word frequency vector $\mathbf{F}_{(\mathbf{W}_{u(1\times \max p(u))})}$, tourist interest label word frequency matrix $\mathbf{F}_{(\mathbf{W}_{(p\times \max p_{(u)})})}$ and tourist sight feature attribute label word frequency vector $\mathbf{F}(c_{(i)(1\times s)})$. As to the interest label vector $\mathbf{W}_{u(1\times \max p(u))}$ in $\mathbf{W}_{(p\times \max p_{(u)})}$ of one cluster $\mathbf{C}_{(u)}$ confirmed by tourist, the appearance frequency $t_{1(u,p_{(u)})}$ of arbitrary label $\forall W_{(u,p_{(u)})}$ in the matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ is defined as the The tourist interest label $W_{(u,p_{(u)})}$ word frequency. Confirm all the word frequency $t_{1(u,p_{(u)})}$ of the labels $W_{(u,p_{(u)})}$ in the vector $\mathbf{W}_{u(1\times \max p(u))}$ and then store the word frequencies into the $1 \times \max p_{(u)}$ dimension vector $\mathbf{F}_{(\mathbf{W}_{u(1\times \max p(u))})}$, and this vector $\mathbf{F}_{(\mathbf{W}_{u(1\times \max p(u))})}$ is called the tourist interest label word frequency vector. The matrix that is formed by p word frequency vectors $\mathbf{F}_{(\mathbf{W}_{u(1\times \max p(u))})}$ relating to cluster $\mathbf{C}_{(u)}$ stored in the sequence of matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ rows is defined as the tourist interest label word frequency matrix $\mathbf{F}_{(\mathbf{W}_{(p\times \max p_{(u)})})}$. The appearance frequency of the arbitrary label $\forall w_{(i,t(i))}$ of the vector $\mathbf{c}_{(i)}$ in the tourist sight text encyclopedia knowledge is defined as the word frequency $t_2(i, t(i))$ for the label $w_{(i,t(i))}$. Confirm the word frequency for all of the labels $w_{(i,t(i))}$ in the vector $\mathbf{c}_{(i)}$ and store them into the $1 \times s$ dimension vector $\mathbf{F}(c_{(i)(1\times s)})$, and this vector $\mathbf{F}(c_{(i)(1\times s)})$ is called the tourist sight feature attribute label word frequency vector. The vector $\mathbf{F}_{(\mathbf{W}_{u(1\times \max p(u))})}$ and the matrix $\mathbf{F}_{(\mathbf{W}_{(p\times \max p_{(u)})})}$ represent the appearance frequency of the interest points relating to the labels selected by the tourists, that is, the tendency of the tourists to the feature attribute labels and related tourists sights. The larger the word frequency value is, the much closer tendency to the feature attribute will be. The vector $\mathbf{F}(c_{(i)(1\times s)})$ represents the appearance frequency of the feature attribute label in the tourist sight text knowledge and definition. Therefore, The matrix $\mathbf{F}_{(\mathbf{W}_{(p\times \max p_{(u)})})}$ and the vector $\mathbf{F}(c_{(i)(1\times s)})$ makes up the basic calculating condition for the factor $k_{(1)}$.

According to the word frequency definition and the extracting process, the element word frequencies of the interest label homogeneous vector $\mathbf{W}^{\wedge}_{(1\times m)}$ are obtained. The process of generating interest label matrix, label word frequency matrix and feature attribute label word frequency matrix is the process of mining tourists' interest data. The following pseudocode is for the interest data mining algorithm.

Algorithm 2 The Interest Data Mining Algorithm

Section 1: Set up the vector $\mathbf{W}_{(t)}$.

Step 1: Initialize the vector $\mathbf{W}^0_{(t)}$.

Step 2: Store $\mathbf{c}_{(i)}$ label into $\mathbf{W}^0_{(t)}$ and form $\mathbf{W}_{(t)}$.

For each element $w_{(i,t)}$ in $\mathbf{c}_{(i)}$:

For each element $W_{(t)}$ in $\mathbf{W}_{(t)}$:

Compare $w_{(i,t)}$ and $W_{(t)}$ until $t = s$ and $i \sim (0, n] \in Z^+$.

Get the vector $\mathbf{W}_{(t)}$.

Section 2: Set up the vector $\mathbf{W}^{\wedge}_{(1\times m)}$.

For each cluster $\mathbf{C}_{(u)}$, $u \sim (0, p] \in Z^+$:

For each element $W_{(1,p(u))}$:

Compare $w_{(i,t(i))}$ and $W_{(1,p(u))}$, judge if they are in the same cluster.

Get the vector $\mathbf{W}^u_{(1\times m)}$ for each $\mathbf{C}_{(u)}$.

$$\mathbf{W}_{(p\times \max p_{(u)})} = \begin{bmatrix} W_{(1,1)} & W_{(1,2)} & \cdots & W_{(1,\max p_{(u)})} \\ W_{(2,1)} & \cdots & W_{(2,\max p_{(u)}-1)} & W_{(2,\max p_{(u)})} \\ \cdots & \cdots & \cdots & \cdots \\ W_{(p,1)} & \cdots & W_{(p,\max p_{(u)}-1)} & W_{(p,\max p_{(u)})} \end{bmatrix}, \tag{3}$$

$$\mathbf{W}_{(p\times \max p_{(u)})} = \begin{bmatrix} W_{(2)} & W_{(4)} & \cdots & W_{(p1)} \\ W_{(p2)} & \cdots & W_{(10)} & 0 \\ \cdots & \cdots & 0 & 0 \\ W_{(p(u1))} & \cdots & W_{(p(u2))} & W_{(q)} \end{bmatrix}. \tag{4}$$

B. INTEREST TOURIST SIGHT MINING ALGORITHM

According to the definitions and the process of interest text mining, the interest tourist sight mining algorithm based on interest label matrix $\mathbf{W}_{(p\times\max p_{(u)})}$, the word frequency matrix $\mathbf{F}_{(\mathbf{W}_{(p\times\max p_{(u)})})}$ and the feature attribute word frequency vector $\mathbf{F}(c_{(i)(1\times s)})$ is set up. The basic principle of the algorithm is as follows. Firstly, the matching model of the tourist interest and tourist sight feature attribute is set up and the feature attribute matching factor $k_{(1)}$ is output. Then, combining with the travel time $k_{(2)}$, basic cost $k_{(3)}$ and tourist sight attraction index $k_{(4)}$, the interest tourist sight mining function is formed. This function is the key model to confirm the affinity between tourist interest and tourist sight feature attribute, and it outputs the quantified affinity value. Store the affinity values in the descending order in vector and extract the matched interest tourist sights according to tourists' interests. The algorithm flow is as follows.

Step 1 Set up the interest label matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ and word frequency matrix $\mathbf{F}_{(\mathbf{W}_{(p\times\max p_{(u)})})}$.

Step 2 Perform the text knowledge mining on all tourist sights $c_{(i)}$ in the domain C. In this mining process, tourist sight feature attribute label $w_{(i,t(i))}$ is obtained. Set up the tourist sight feature attribute label vector $\mathbf{c}_{(i)}$ and obtain the feature attribute word frequency vector $\mathbf{F}(c_{(i)(1\times s)})$.

Step 3 Set up the algorithm of feature attribute matching factor $k_{(1)}$ between the interest label matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ and the tourist sight $c_{(i)}$.

Sub-step 1 As to arbitrary cluster $\forall C_{(u)}$, confirm the tourist sight interest label vector $\mathbf{W}_{u(1\times\max p(u))}$ and the word frequency vector $\mathbf{F}_{(\mathbf{W}_{u(1\times\max p(u))})}$ relating to the cluster, and obtain the word frequency $t_{1(u,p_{(u)})}$ for each interest label $W_{(u,p_{(u)})}$.

Sub-step 2 As to arbitrary tourist sight $\forall c_{(i)}$, confirm its feature attribute label vector $\mathbf{c}_{(i)}$ and word frequency vector $\mathbf{F}(c_{(i)(1\times s)})$, and obtain the word frequency $t_2(i, t(i))$ for each feature attribute label $w_{(i,t(i))}$.

Sub-step 3 Set up the interest label homogeneous vector $\mathbf{W}^{\wedge}_{(1\times m)}$ between the vector $\mathbf{W}_{u(1\times\max p(u))}$ and vector $\mathbf{c}_{(i)}$, including m quantity of homogeneous labels. Meanwhile, obtain the homogeneous labels' word frequency $t_{1(u,p_{(u)})}$ in the vector $\mathbf{W}_{u(1\times\max p(u))}$ and the word frequency in the vector $\mathbf{c}_{(i)}$. According to the definition, it meets the conditions:

(1) $t_{1(u,p_{(u)})}, t_2(i, t(i)) \in Z^+$;

(2) $t_{1(u,p_{(u)})} \neq 0 \vee t_2(i, t(i)) \neq 0$;

(3) $t_{1(u,p_{(u)})} \ll t_2(i, t(i))$, that is, $|t_{1(u,p_{(u)})} - t_2(i, t(i))| = 0$ is the small probability event.

For the convenience of setting up the algorithm, of all the m quantity of homogeneous labels, the label word frequency belonging to the vector $\mathbf{W}_{u(1\times\max p(u))}$ is noted as $t_1(u, a)$, and the label word frequency belonging to the vector $\mathbf{c}_{(i)}$ is noted as $t_2(i, b)$. The $t_1(u, a)$ and $t_2(i, b)$ meet the conditions in the Sub-step 3, and $a, b \in (0, m] \subset Z^+$, the counting principle are $a = b, a = a + 1$ and $b = b + 1$.

Sub-step 4 Set up the algorithm of the feature attribute matching factor $k_{(1)}$. Then, if the vector meets the condition $\mathbf{W}^{\wedge}_{(1\times m)} \neq 0$,that is, there is at least one homogeneous label of the vector $\mathbf{W}_{u(1\times\max p(u))}$ and the vector $\mathbf{c}_{(i)}$. It meets the condition to form the factor $k_{(1)}$ algorithm, the algorithm is shown as Formula (5). Formula (5) is the improved text similarity algorithm based on the cosine similarity function, in which $t_1(u, a)$ and $t_2(i, b)$ are the word frequencies of the same labels in the vector $\mathbf{W}_{u(1\times\max p(u))}$ and the vector $\mathbf{c}_{(i)}$ respectively. The $t_{1(u,p_{(u)})}$ and $t_2(i, t(i))$ represent the label word frequencies of the vector $\mathbf{W}_{u(1\times\max p(u))}$ and the vector $\mathbf{c}_{(i)}$ respectively.

Sub-step 5 Set up the tourist interest vector $\mathbf{K}_{(r)}$. The feature attribute matching factor $k_{(1)}$, the expected travel time $k_{(2)}$, basic travel cost $k_{(3)}$ and tourist sight attraction index $k_{(4)}$ are combined to set up the tourist interest vector $\mathbf{K}_{(r)}$, it meets the condition $\mathbf{K}(r) = [k_{(1)}, k_{(2)}, k_{(3)}, k_{(4)}]$, $r \in (0, 4] \subset Z^+$. The tourist interest vector $\mathbf{K}_{(r)}$ is the key to set up interest tourist sight mining objective function.

$$k_{(1)}(\mathbf{W}_u, \mathbf{c}_{(i)}) = \sum_{a=1,b=1}^{m} t_1(u,a)\times t_2(i,b)\times\Big(\sum_{p_{(u)}=1}^{\max p_{(u)}} t_1^2(u,p_{(u)}) \times \sum_{t(i)=1}^{s} t_1^2(i,t(i))\Big)^{-\frac{1}{2}}, \quad (5)$$

Step 4 Set up the tourist sight mining objective function of $G_{(\mathbf{K}_{(r)},c_{(i)})}$. The tourist sight mining objective function is determined by the tourist interest vector and the function attributes the tourist sights can provide. It is the function to judge the extent on how tourist sights could satisfy tourists. The larger the function $G_{(\mathbf{K}_{(r)},c_{(i)})}$ value is, the much closer of the tourist sights' function attributes to tourists' interests will be. To make the factors $k_{(1)}, k_{(2)}, k_{(3)}$ and $k_{(4)}$ all have the same scale impact on the function results, the disturbing coefficient $\varepsilon_{(r)}$ is introduced for the factor $k_{(r)}$. The function of each disturbing coefficient $\varepsilon_{(r)}$ is to normalize the factors into the same scale.

Obtain one factor $k_{(1)}, k_{(2)}, k_{(3)}$ and $k_{(4)}$ respectively and compare their orders of magnitudes $\tau_{(r)}$. Their orders of magnitudes $\tau_{(r)}$ are $\ldots, 100, 10, 1, 0.1, 0.01, 0.001, \ldots$. And select the middle one value's order of magnitudes as the standard one. Compare the other factors' orders of magnitudes with the standard one and get the coefficient $\varepsilon_{(r)}$. Take the order of magnitudes $\tau_{(2)}$ for the factor $k_{(2)}$ as the standard one to confirm the coefficient $\varepsilon_{(r)}$.

(1) Compare $\tau_{(1)}$ with $\tau_{(2)}$.
Calculate $\tau_{(1)}/\tau_{(2)} = p_1$, then $\tau_{(1)} = p_1\tau_{(2)}$, $\varepsilon_{(1)} = 1/p_1$;
(2) Compare $\tau_{(3)}$ with $\tau_{(2)}$.
Calculate $\tau_{(3)}/\tau_{(2)} = p_2$, then $\tau_{(3)} = p_2\tau_{(2)}$, $\varepsilon_{(3)} = 1/p_2$;
(3) Compare $\tau_{(4)}$ with $\tau_{(2)}$.
Calculate $\tau_{(4)}/\tau_{(2)} = p_3$, then $\tau_{(4)} = p_3\tau_{(2)}$, $\varepsilon\,(4) = 1/p_3$.
In which $p_1, p_2, p_3 \in (\ldots, 100, 10, 1, 0.1, 0.01, 0.001 \ldots)$.

$$\begin{cases} G_{(\mathbf{K}_{(r)},c_{(i)})} = \varepsilon_1 k_{(1)} + \left[\sum_{r=2}^{n} \varepsilon_i |k(i) - k(i^*)|^d\right]^{-1/d}, \\ JG(\mathbf{K}_{(r)}, c_{(i)}) = \max\{G_{(\mathbf{K}_{(r)},c_{(i)})}\}, \\ s.t.\ i \in (1, n] \subset Z^+, \quad \varepsilon_i \in (0, 1] \subset R^+. \end{cases} \quad (6)$$

Set the quantitative values for the function attributes of the tourist sight $c_{(i)}$ as the best travel time $k_{(2^*)}$, basic travel cost $k_{(3^*)}$, attraction index $k_{(4^*)}$. Combining with the feature attribute matching factor $k_{(1)}$, the interest tourist sight mining objective function $G_{(\mathbf{K}_{(r)},c_{(i)})}$ is set up as Formula (6), and $n = 4$. The Formula (6) is the matching degree function between the tourists' interests and the tourist sight feature attributes based on the improved Minkowski distance function, whose aim is to search and obtain the minimum difference value between the tourists' interests and the tourist sight feature attributes. According to the definition, the maximum value of the matching degree is the calculating criterion for the factor $k_{(1)}$, while the minimum absolute difference values between the factors $k_{(2)}, k_{(3)}, k_{(4)}$ and their related quantitative values $k_{(2^*)}, k_{(3^*)}$ and $k_{(4^*)}$ for the function attributes of the tourist sight $c_{(i)}$ are the calculating criterion for the factors $k_{(2)}, k_{(3)}, k_{(4)}$.

Thus, the reciprocal on the sum of the difference values between the factors $k_{(2)}, k_{(3)}, k_{(4)}$ and $k_{(2^*)}, k_{(3^*)}, k_{(4^*)}$ is brought forward, the exponent sign is $-1/d$. The function of the exponent sign is to maximize the function value $G_{(\mathbf{K}_{(r)},c_{(i)})}$. The value of the parameter d determines the order of the Minkowski distance function. When $d = 1$, the Minkowski distance is the Manhattan distance, which could represent the right-angle side distance between two points in the two dimension space. Its amount of calculation is small. When $d = 2$, the Minkowski distance is the Euclidean distance, which could represent the straight-line distance between two points in the two dimension space. Its amount of calculation is relatively large. When the improved algorithm $G_{(\mathbf{K}_{(r)},c_{(i)})}$ represents the degree of approximation between the tourists' interests and the quantitative values for the function attributes of the tourist sight, the value of the parameter d should keep the reciprocal on the sum of the difference values between the factors $k_{(2)}, k_{(3)}, k_{(4)}$ and $k_{(2^*)}, k_{(3^*)}, k_{(4^*)}$ along with the factor $k_{(1)}$ in the same order of magnitudes, and reduce the calculating space complexity and time complexity. According to the definition, the value $d = 1$ is used to calculate the function.

Step 5 Set up the descending order vector $\mathbf{V}_{(uv)}$ of the cluster $\mathbf{C}_{(u)}$ tourist sights' objective function $G_{(\mathbf{K}_{(r)},c_{(i)})}$. Calculate the objective function $G_{(\mathbf{K}(r),c(i))}$ values for the $n_{(u)}$ tourist sights in the cluster $\mathbf{C}_{(u)}$ and form the vector $\mathbf{V}_{(uv)}$.

Sub-step 1 When u is confirmed, set up the $1 \times u$ dimension vector $\mathbf{V}_{(uv)}$.

Sub-step 2 Calculate the objective function of $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ value between the first tourist sight $c_{(u,1)}$ in cluster $\mathbf{C}_{(u)}$ and the vector $\mathbf{K}_{(r)}$, and the objective function $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ value between the second tourist sight $c_{(u,2)}$ and the vector $\mathbf{K}_{(r)}$.

(1) If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} \geq G_{(\mathbf{K}_{(r)},c_{(u,2)})}$, store the value $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ into the first element $\mathbf{V}_{(u1)}$ of the vector $\mathbf{V}_{(uv)}$, and store the value $G(\mathbf{K}_{(r)}, c_{(u,2)})$ into the second element $\mathbf{V}_{(u2)}$ of the vector $\mathbf{V}_{(uv)}$;

(2) If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} < G_{(\mathbf{K}_{(r)},c_{(u,2)})}$, store the value $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ into the first element $\mathbf{V}_{(u1)}$ of the vector $\mathbf{V}_{(uv)}$, and store the value $G(\mathbf{K}_{(r)}, c_{(u,2)})$ into the second element $\mathbf{V}_{(u2)}$ of the vector $\mathbf{V}_{(uv)}$.

Sub-step 3 Calculate the objective function of $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$ value between the third tourist sight $c_{(u,3)}$ in cluster $\mathbf{C}_{(u)}$ and the vector $\mathbf{K}_{(r)}$.

(1) If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} \geq G_{(\mathbf{K}_{(r)},c_{(u,2)})}$:

① If $G_{(\mathbf{K}_{(r)},c_{(u,3)})} > G_{(\mathbf{K}_{(r)},c_{(u,1)})} \geq G_{(\mathbf{K}_{(r)},c_{(u,2)})}$: store the value $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$ into the first element $\mathbf{V}_{(u1)}$ of the vector $\mathbf{V}_{(uv)}$, descend to store the value $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ into the second element $\mathbf{V}_{(u2)}$ of the vector $\mathbf{V}_{(uv)}$ and the value $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ into the third element $\mathbf{V}_{(u3)}$ of the vector $\mathbf{V}_{(uv)}$.

② If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} \geq G_{(\mathbf{K}_{(r)},c_{(u,3)})} > G_{(\mathbf{K}_{(r)},c_{(u,2)})}$: keep the value $G(\mathbf{K}(r), c_{(u,1)})$ element unchanged.

And then store the value $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$ into the second element $\mathbf{V}_{(u2)}$ of the vector $\mathbf{V}_{(uv)}$, and descend to store the value $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ into the third element $\mathbf{V}_{(u3)}$ of the vector $\mathbf{V}_{(uv)}$.

③ If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} \geq G_{(\mathbf{K}_{(r)},c_{(u,2)})} \geq G_{(\mathbf{K}_{(r)},c_{(u,3)})}$: keep the two values $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ and $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ elements unchanged. Store the value $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$ into the third element $\mathbf{V}_{(u3)}$ of the vector $\mathbf{V}_{(uv)}$.

(2) If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} < G_{(\mathbf{K}_{(r)},c_{(u,2)})}$:

① If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} < G_{(\mathbf{K}_{(r)},c_{(u,2)})} < G_{(\mathbf{K}_{(r)},c_{(u,3)})}$, store the value $G(\mathbf{K}(r), c_{(u,3)})$ into the first element $\mathbf{V}_{(u1)}$ of the vector $\mathbf{V}_{(uv)}$, and descend to store the value $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ into the second element $\mathbf{V}_{(u2)}$ of the vector $\mathbf{V}_{(uv)}$ and the value $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ into the third element $\mathbf{V}_{(u3)}$ of the vector $\mathbf{V}_{(uv)}$.

② If $G_{(\mathbf{K}_{(r)},c_{(u,1)})} < G_{(\mathbf{K}_{(r)},c_{(u,3)})} < G_{(\mathbf{K}_{(r)},c_{(u,2)})}$, keep the value $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ element unchanged. And then store the value $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$ into the second element $\mathbf{V}_{(u2)}$ of the vector $\mathbf{V}_{(uv)}$, and descend to store the value $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ into the third element $\mathbf{V}_{(u3)}$ of the vector $\mathbf{V}_{(uv)}$.

③ If $G_{(\mathbf{K}_{(r)},c_{(u,3)})} < G_{(\mathbf{K}_{(r)},c_{(u,1)})} < G_{(\mathbf{K}_{(r)},c_{(u,2)})}$: keep the two values $G(\mathbf{K}(r), c_{(u,1)})$ and $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$ elements unchanged, store the value $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$ into the third element $\mathbf{V}_{(u3)}$ of the vector $\mathbf{V}_{(uv)}$.

Sub-step 4 Calculate the objective function of $G(\mathbf{K}_{(r)}, c_{(u,v)})$ value between the No. v tourist sight $c_{(u,v)}$ in cluster $\mathbf{C}_{(u)}$ and the vector $\mathbf{K}_{(r)}$. According to the method of Sub-step 2 and Sub-step 3, compare the objective function values from the first to the No. $v-1$ tourist sight and descend to store the v quantity of objective function values of related tourist sights into the top v quantity of elements in the vector $\mathbf{V}_{(uv)}$. Traverse $v = 4, 5, \ldots, n_{(u)}$ and calculate $n_{(u)}$ quantity of objective function values of related tourist sights into the vector $\mathbf{V}_{(uv)}$. At this time, the vector $\mathbf{V}_{(uv)}$ is full rank, $rank(\mathbf{V}_{(uv)}) = n_{(u)}$.

Step 6 Traverse $u = 1, 2, \ldots, p$. Set up all the vectors $\mathbf{V}_{(uv)}$ with the objective function values $G_{(\mathbf{K}_{(r)},c_{(i)})}$ for all the clusters.

Step 7 Mine and obtain the tourist sights that best meet the tourists' interests. Take one-day trip as an example. Under the condition that the basic interest matrix $\mathbf{W}_{(p\times \max p_{(u)})}$ is confirmed, tourists consider the time schedule, travel cost, physical conditions, etc., they choose the $x_{(u)}$ quantity of the expected tourist sights in each cluster $\mathbf{C}_{(u)}$, $u \in (0, p]$. Set up the $1 \times r$ dimension vector $\mathbf{C}_{(i)}$ which is used to store the tourist sights to be visited, whose quantity of element is r.

Store related tourist sights into the vector $\mathbf{C}_{(i)}$ in the sequence of cluster $C_{(1)}, C_{(2)}, \ldots, C_{(u)}, \ldots, C(p)$. The element meets the condition of Formula (7). In the Formula (7), the total sum of the tourist sights to be visited r is determined by the tourists' interests and the expected tourist sights in each cluster. The interests determine the expected tourist sight cluster, the quantity of the expected tourist sights in one cluster determines the tourist sights provided by the recommendation system.

$$r = \sum_{u=1}^{p} x_{(u)}, \quad s.t.\ x_{(u)} \sim C_{(u)}, \quad u \in (0, p]. \tag{7}$$

Sub-step 1 Extract the top $x_{(1)}$ quantity of elements in the vector $\mathbf{V}_{(1v)}$ of the cluster $C_{(1)}$, they are related to the top $x_{(1)}$ quantity of tourist sights $c_{(1,v)}$ with objective function values in the descending order. And then store the $x_{(1)}$ quantity of tourist sights $c_{(1,v)}$ into the vector $\mathbf{C}_{(i)}$ from the first element to the No. $x_{(1)}$ element.

Sub-step 2 Extract the top $x_{(2)}$ quantity of elements in the vector $\mathbf{V}_{(2v)}$ of cluster $C_{(2)}$, they are related to the top $x_{(2)}$ quantity of tourist sights $c_{(2,v)}$ with objective function values in the descending order. Store the $x_{(2)}$ quantity of tourist sights $c_{(2,v)}$ into the vector $\mathbf{C}_{(i)}$ from the No. $x_{(1)} + 1$ element to the No. $x_{(1)} + x_{(2)}$ element.

Sub-step 3 Extract the top $x_{(u)}$ quantity of elements in the vector $\mathbf{V}_{(uv)}$ of cluster $\mathbf{C}_{(u)}$, they are related to the top $x_{(u)}$ quantity of tourist sights $c_{(u,v)}$ with objective function values in the descending order. Store the $x_{(u)}$ quantity of tourist sights $c_{(u,v)}$ into the vector $\mathbf{C}_{(i)}$ from the No. $\sum_1^{u-1} x_{(u)} + 1$ element to the No. $\sum_1^{u} x_{(u)}$ element, $u = 3, 4, \ldots, p$.

Sub-step 4 Finish searching the p quantity of clusters and obtain the vector $\mathbf{C}_{(i)}$. The r quantity of tourist sights stored in the vector $\mathbf{C}_{(i)}$ match tourists' interests, and they are the optimal tourist sights the recommendation system will provide for tourists, meanwhile, they are also the key nodes to schedule the tour route chain.

The following pseudo-code is for the interest tourist sight mining algorithm.

Algorithm 3 The Interest Tourist Sight Mining Algorithm

Step 1: Set up matrix $\mathbf{W}_{(p\times\max p_{(u)})}$ and $\mathbf{F}_{(\mathbf{W}_{(p\times\max p_{(u)})})}$.
Set up the matrix $\mathbf{F}(c_{(i)(1\times s)})$ for the vector $\mathbf{c}_{(i)}$.
Step 2: Set up $k_{(1)}$ algorithm for $\mathbf{W}_{(p\times\max p_{(u)})}$ and $\mathbf{c}_{(i)}$.
Sub-step 1: As to $\forall C_{(u)}$, confirm $\mathbf{W}_{u^{(1\times\max p(u))}}$ and $\mathbf{F}(\boldsymbol{W}_u(1\times\max p_{(u)}))$.
Get $t_{1^{(u,p_{(u)})}}$ for each $W_{(u,p_{(u)})}$.
Sub-step 2: As to $\forall c_{(i)}$, confirm $\mathbf{c}_{(i)}$ and $\mathbf{F}(c_{(i)(1\times s)})$.
Get $t_2(i, t(i))$ for each $w_{(i,t(i))}$.
Sub-step 3: Set up the vector $\mathbf{W}^{\wedge}_{(1\times m)}$.
Sub-step 4: Get $k_{(1)}$ and vector $\mathbf{K}_{(r)}$.
Step 3: Set up the function $G_{(\mathbf{K}_{(r)},c_{(i)})}$.
Step 4: Set up the vector $\mathbf{V}_{(uv)}$.
Sub-step 1: Calculate the value $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ and $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$.
Compare $G_{(\mathbf{K}_{(r)},c_{(u,1)})}$ and $G_{(\mathbf{K}_{(r)},c_{(u,2)})}$, sort in descending order.
Sub-step 2: Calculate the value $G_{(\mathbf{K}_{(r)},c_{(u,3)})}$.
Compare the Three values and sort them in descending order.
Sub-step 3: **For each** No. v and $u = 1, 2, \ldots, p$: calculate $G(\mathbf{K}_{(r)}, c_{(u,v)})$ and compare the v amount values, sort them in descending order. Get $\mathbf{V}_{(uv)}$.
Step 5: For each $u = 1, 2, \ldots, p$: extract $x_{(u)}$ amount of elements of $\mathbf{V}_{(uv)}$.
Store $x_{(u)}$ amount of tourist sights $c_{(u,v)}$ into the elements from No. $\sum_1^{u-1} x_{(u)} + 1$ to No. $\sum_1^{u} x_{(u)}$. Get the vector $\mathbf{C}_{(i)}$.

III. THE TOUR ROUTE CHAIN ALGORITHM BASED ON THE MP NERVE CELL MODEL OF THE MULTIVARIATE TRANSPORTATION MODES

After all the interest tourist sights are confirmed, they are used as key nodes to form the tour route chains. Tourists start from the point P, visit the r quantity of tourist sights, and return to the point P. This process relates to a whole trip. Since different tourists may choose the different transportation modes, it may cause different satisfaction degree in the same tour route. This is determined by city geographic information service, traffic information service and tourist sight information [12]. According to the recommended tourist sights, combining with the city basic data, the tour route chain algorithm based on tourists' transportation modes is set up. It outputs the tour routes which conform to tourists' interests and travel motive.

A. NERVE CELL MODEL BASED ON MULTIVARIATE TRANSPORTATION MODE MATCHING KEYS

The MP Nerve cell structure is a typical structure with multiple inputs and single output. Its input terminals include n quantity of signal paths x_i from the previous nerve cell, $i \in (0, n] \in Z^+$. Each signal path will be impacted by the connecting weight ω_i and then will further impact the current nerve cell. Meanwhile, the nerve cell itself has certain accommodation coefficients θ. Along with the impact of all the accommodation coefficients θ, the output signal path y of the nerve cell is generated, and this output signal path is one of the next nerve cell's input signal paths. Figure 2 is the basic structure of MP nerve cell. Formula (8) is the MP nerve cell signal input and output model. In the Formula (8), each input signal x_i is iterated by the connecting weight and added with each other, and then they are regulated by the coefficients θ and output Generally, the signal y as the input signal x_{i+1} for the next nerve cell. The connecting weight ω_i is the positive real number and the coefficient θ is the negative real number. In the Formula, the symbol f is the expression for the MP nerve cell iteration function. The improved MP nerve cell model is set up on the signal processing and transmitting mode of the MP nerve cell and its basic structure

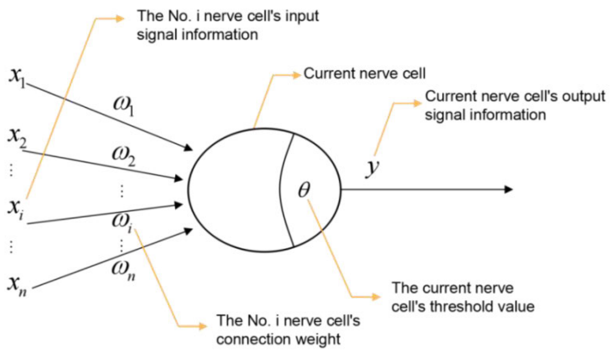


FIGURE 2. MP nerve cell structure.

in Figure 2 and the Formula (8).

$$y = f\left(\sum_{i=1}^{n} \omega_i x_i - \theta\right). \tag{8}$$

Inspired by the MP nerve cell structure and its signal transmission mode, each one of the tourist sight in the tour route could be seen as a nerve cell. The process that tourists travel from one tourist sight to the next one could be abstracted to the mode of signal transmission from one nerve cell to the next one. The signal transmitted in the tour route relates to tourists motive satisfaction degree. Set the initial motive satisfaction degree value as x_i. Set certain geographic information factors and traffic information factors as connecting weight ω_i and accommodation coefficients θ. From the transmission model, the tourist sight nerve cell output y could be obtained, and it is the next nerve cell's input x_i.

Def 4.1: Tourist sight nerve cell motive weight $\omega_{(k)}$ and accommodation coefficients θ. When tourists travel among tourist sights, the geographic information service and traffic information service will influence tourists' motive satisfaction. In the process of traveling, the factors that directly influence tourists' motive satisfaction are defined as the tourist sight nerve cell motive weights $\omega_{(k)}$, while the factors that indirectly influence tourists' motive satisfaction are defined as the accommodation coefficients θ. According to the real world environment geographic information data and traffic information data, each calculating method for each tourist sight nerve cell motive weight $\omega_{(k)}$ and accommodation coefficient θ are as follows [36], [37]. In order to confirm that each factor's impact on motive satisfaction is in the same scale, set the normalization parameter $\varepsilon_{(k)}$ and $\sigma_{(k)}$ for the weights $\omega_{(k)}$ and accommodation coefficients θ [38]. Obtain one factor $\omega_{(15)}, \omega_{(2)}, \omega_{(3)}, \omega_{(4)}, \omega_{(5)}$ and $\omega_{(6)}$ respectively, and then compare their orders of magnitudes $\tau_{(r)}$. Their orders of magnitudes $\tau_{(r)}$ are ..., 100, 10, 1, 0.1, 0.01, 0.001, ... And select the middle one value's order of magnitudes as the standard one. Compare the other factors' orders of magnitudes with the standard one and get the coefficient $\varepsilon_{(k)}$. Take the order of magnitudes $\tau_{(3)}$ for the factor $\omega_{(3)}$ as the standard one to confirm the coefficient $\varepsilon_{(k)}$ [53], [54].

(1) Compare $\tau_{(1)}$ with $\tau_{(3)}$.
Calculate $\tau_{(1)}/\tau_{(3)} = p_1$, then $\tau_{(1)} = p_1\tau_{(3)}$, $\varepsilon_{(1)} = 1/p_1$;
(2) Compare $\tau_{(2)}$ with $\tau_{(3)}$.
Calculate $\tau_{(2)}/\tau_{(3)} = p_2$, then $\tau_{(2)} = p_2\tau_{(3)}$, $\varepsilon\ (2) = 1/p_2$;
(3) Compare $\tau_{(4)}$ with $\tau_{(3)}$.
Calculate $\tau_{(4)}/\tau_{(3)} = p_3$, then $\tau_{(4)} = p_3\tau_{(3)}$, $\varepsilon\ (4) = 1/p_3$;
(4) Compare $\tau_{(5)}$ with $\tau_{(3)}$.
Calculate $\tau_{(5)}/\tau_{(3)} = p_4$, then $\tau_{(5)} = p_4\tau_{(3)}$, $\varepsilon\ (5) = 1/p_4$;
(5) Compare $\tau_{(6)}$ with $\tau_{(3)}$.
Calculate $\tau_{(6)}/\tau_{(3)} = p_5$, then $\tau_{(6)} = p_5\tau_{(3)}$, $\varepsilon\ (6) = 1/p_5$;
In which $p_1, p_2, p_3, p_4, p_5 \in$(...,100,10,1,0.1,0.01,0.001...).

The confirmation method for the coefficient $\sigma_{(k)}$ is the same as the $\varepsilon_{(k)}$. The specific algorithms are as follows.

(1) $\omega_{(15)}$: the route distance from one tourist sight to another one, $z_1(km)$, $\omega_{(15)} = z_1^{-1}$, $z_1 \in R^+$, $\varepsilon_1 = 1$;
(2) $\omega_{(2)}$: the spatial coordinate distance z_2, $z_2 \in R^+$, $\varepsilon_2 = 10$, $\omega_{(2)} = z_2 = (B^2 + l^2)^{1/2}$;
(3) $\omega_{(3)}$: the quantity of public bus z_3, $\omega_{(3)} = z_3$, $z_3 \in Z^+$, $\varepsilon_3 = 0.1$;
(4) $\omega_{(4)}$: the average running time of the public bus $z_4(h)$, $\omega_{(4)} = z_4^{-1}$, $z_4 \in R^+$, $\varepsilon_4 = 0.1$;
(5) $\omega_{(5)}$: the taxi fee z_5(¥*yuan*), $\omega_{(5)} = z_5^{-1}$, $z_5 \in R^+$, $\varepsilon_5 = 1$;
(6) $\omega_{(6)}$: the average road congestion index z_6, $\omega_{(6)} = 1 - z_6$, $z_6 \in R^+$, $\varepsilon_6 = 1$;
(7) θ (15): the quantity of traffic light z_7, $\theta(1) = z_7$, $z_7 \in Z^+$, $\sigma_1 = -0.01$;
(8) θ (2): the distance from tourist sight to bus station z_8 (km), $\theta\ (2) = z_8$, $z_8 \in R^+$, $\sigma_2 = -0.1$;
(9) $\theta(3)$: the average waiting time for the taxi $z_9(h)$, $z_9 \in R^+$, $\theta(3) = z_9$, $z_9 > 0$, $\sigma_3 = -0.1$;
(10) θ (4): the average quantity of the congestion road z_{10}, $\theta\ (4) = z_{10}$, $z_{10} \in Z^+$, $\sigma_4 = -0.01$.

Def 4.2: Tourist sight nerve cell motive value $X_{(i)}$. According to the definition of vector $\mathbf{C}_{(i)}$, one tour route contains r tourist sights to be visited. The motive value that is output by the No. $i-1$ tourist sight and input to the No. i tourist sight is defined as the input motive value $X_{(i)}$ of the No. i nerve cell.

According to the definition, the input motive value $X_{(i)}$ is the output value of the No. $i - 1$ tourist sight. As to the tour route mode from point P to the terminal P, the whole tour route will output $r + 1$ quantity of nerve cell motive values $X_{(i)}$. Different from the principle of signal processing and transmitting of the MP nerve cell in the Figure 2 and Formula (8), the signal information on the tour route chain transmits in a single input and single output mode.

In the improved MP model, each influence factor affects the input signal information $X_{(i)}$ and generate the output signal information, and then transmit to the next nerve cell. Thus, the signal information $X_{(i)}$ changes along with the transmitting process. It is also influenced by the tour sequence, influence factors and the traffic transportation modes.

According to the MP nerve cell structure, the transportation mode, tourist sight nerve cell motive weight ω_i,

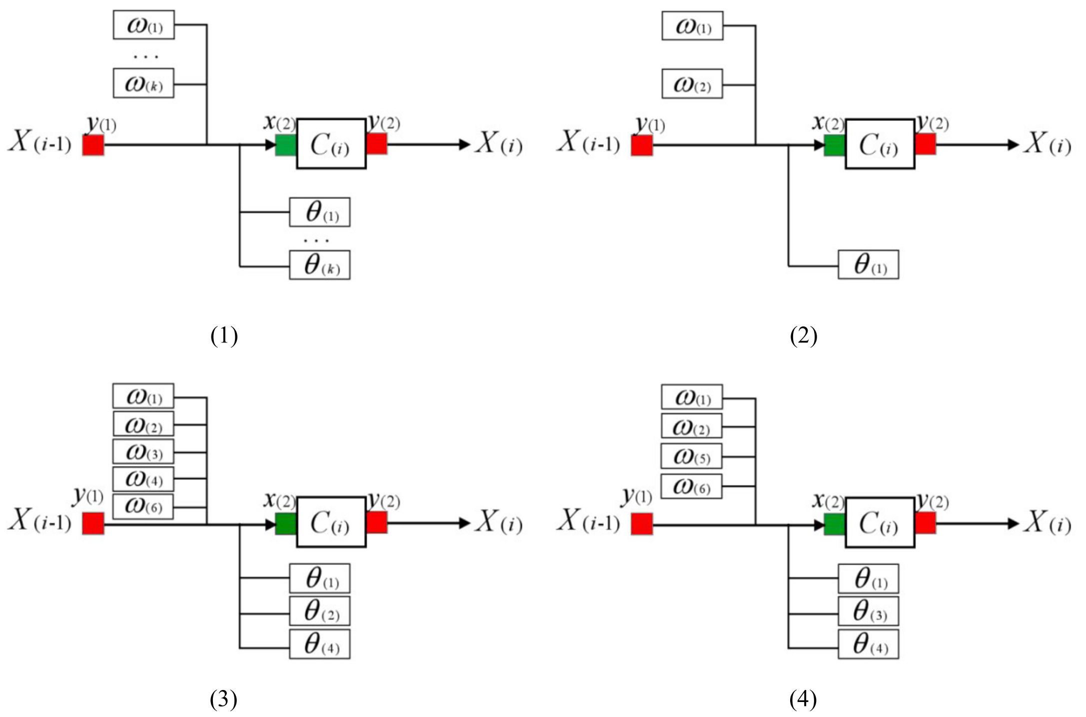


FIGURE 3. The nerve cell model based on multivariate transportation mode matching keys.

TABLE 1. The conditions of different factors' impacts on the multivariate transportation modes.

	$\omega_{(1)}$	$\omega_{(2)}$	$\omega_{(3)}$	$\omega_{(4)}$	$\omega_{(5)}$	$\omega_{(6)}$	$\theta_{(1)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(4)}$
$\xi=1$	1	1	0	0	0	0	1	0	0	0
$\xi=2$	1	1	1	1	0	1	1	1	0	1
$\xi=3$	1	1	0	0	1	1	1	0	1	1

accommodation coefficient θ_i and the tourist sight nerve cell motive values $X_{(i)}$, the motive nerve cell model based on multivariate transportation mode matching keys is set up, as Formula (9) shows.

In the Formula (9), the initial value for the output signal information $X^{\xi}_{(i+1)}$ is $X^{\xi}_{(i)}$. The initial signal information is influenced by the weight value $\omega_{(k)}$ and coefficient $\varepsilon_{(k)}$. The iteration value is generated. And then, it is affected by the coefficient $\theta_{(k)}$ and the parameter $\sigma_{(k)}$ to iterate the output signal $X^{\xi}_{(i+1)}$.

$$
\begin{aligned}
X^{\xi}_{(i+1)} = X^{\xi}_{(i)} &+ \sum_{k(\xi)=1}^{\max k(\xi)} X^{\xi}_{(i)} \cdot \varepsilon(k(\xi))\omega(k(\xi)) \\
&+ \sum_{k(\xi)=1}^{\max k(\xi)} \sigma(k(\xi))\theta(k(\xi)), \\
s.t.\ \ i \in (0, r] \subset \mathrm{Z}^{+}. \qquad (9)
\end{aligned}
$$

In this model, parameter ξ represents the transportation mode tourists choose. When $\xi = 1$, the tourists choose walking or cycling. When $\xi = 2$, the tourists choose public transportation. When $\xi = 3$, the tourists choose taxi or self-driving [32], [33]. When the tourists choose the different modes of transportation, the trip will be influenced by different nerve cells' motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$.

Thus, when tourists choose different transportation modes, the recommendation system will provide them different optimal tour routes. Based on the chosen transportation mode and the MP nerve cell model, the motive nerve cell model based on multivariate transportation mode matching keys is set up, shown as Figure 3. Figure 3(15) is an improved MP nerve cell model with the inputs of multivariate transportation mode matching keys, meanwhile, it has the single output of tourist sight nerve cell motive value $X_{(i)}$. In a whole trip, if the transportation modes are different, the identical tour route will generate different motive satisfaction degrees, and this is caused by the different impacts of the influence factors [46], [50].

Table 1 shows the conditions of different factors' impacts on the multivariate transportation modes. In the table, the value 1 means that the transportation mode is influenced by this related factor, while the value 0 means that the transportation mode is not influenced by this related factor. According to the Table 1 data, the nerve cell model with different matching input keys is set up, shown in Figure 3(2) and 3(4).

According to Table 1 data and Figure 3 the improved MP nerve cell model with the inputs of multivariate transportation mode matching keys, the improved motive nerve cell algorithm based on tourist sight nerve cell motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$ is set up as Formulas (10)~(12), representing the tourist sight nerve cell motive values $X_{(i)}$ algorithm of the modes $\xi = 1, \xi = 2$ and $\xi = 3$ respectively, $i \in (0, r] \subset Z^{+}$. As to one tour route, by the iterating process of Formulas (10)~(12), each tourist sight nerve cell motive output value $X(i+1)$ is obtained under the different transportation modes.

When $i = r$, the motive value $X^{\xi}_{(r+1)}$ of one tour route under the condition of certain transportation mode is output. Analyze the Formula (10)~(12). Different transportation modes are influenced by the different weights $\omega_{(k)}$ and coefficients $\theta_{(k)}$. The specific algorithm is generated on the Formula (9). When the tourists choose one certain kind of transportation mode, the recommendation system will choose one Formula in (10)~(12) to calculate the output iteration value of the tour route chain, and then recommend tour routes for the tourists.

$$X^{1}_{(i+1)} = X^{1}_{(i)} + \sum_{k_{(\xi)}=1}^{2} X^{1}_{(i)} \cdot \varepsilon_{(k(\xi))}\omega_{(k(\xi))} + \sigma_{(1)}\theta_{(1)}, \tag{10}$$

$$\begin{aligned} X^{2}_{(i+1)} &= X^{2}_{(i)} + \sum_{k_{(\xi)}=1}^{4} X^{2}_{(i)} \cdot \varepsilon_{(k(\xi))}\omega_{(k(\xi))} + X^{2}_{(i)} \cdot \varepsilon_{(6)}\omega_{(6)} \\ &\quad + \sum_{k_{(\xi)}=1}^{2} \sigma_{(k(\xi))}\theta_{(k(\xi))} + \sigma_{(4)}\theta_{(4)}, \end{aligned} \tag{11}$$

$$\begin{aligned} X^{3}_{(i+1)} &= X^{3}_{(i)} + \sum_{k_{(\xi)}=1}^{2} X^{3}_{(i)} \cdot \varepsilon_{(k(\xi))}\omega_{(k(\xi))} \\ &\quad + \sum_{k_{(\xi)}=5}^{6} X^{3}_{(i)} \cdot \varepsilon_{(k(\xi))}\omega_{(k(\xi))} + \sigma_{(1)}\theta_{(1)} \\ &\quad + \sum_{k_{(\xi)}=3}^{4} \sigma_{(k(\xi))}\theta_{(k(\xi))}. \end{aligned} \tag{12}$$

B. TOUR ROUTE CHAIN ALGORITHM MODEL BASED ON THE TOURIST SIGHT MOTIVE NERVE CELL

One tourism city's commercial service will bring certain influence on tourists' traveling process. For example, based on the confirmation of the tourist sights to be visited, tourists will choose and confirm the temporary accommodation in the tourism city, that is, reserving the restaurant or hotel to live in. As to a one-day trip, tourists usually start the trip from the restaurant or hotel and then visit the confirmed tourist sights. Usually, after visiting all the tourist sights, they will return to the restaurant or hotel for rest. On the analysis of the optimization on the spot geographic distribution, the confirmation of the restaurant or hotel is very important. In the design of the proposed algorithm, the location of the restaurant or hotel determines the selection of the tourist sights to be visited as well as the optimal tour route formed by the tourist sights, because according to the algorithm, when the location of the restaurant or hotel changes, the optimal tourist sights will also change, then the output optimal tour route along with its iteration value will simultaneously change, too. Therefore, when designing the tour route algorithm, the location of the restaurant or hotel is set as the starting point P and absorbed into the tour route critical nodes. Besides the commercial service of restaurant and hotel, the shopping malls are also important in tourism recommendation. In the proposed algorithm, the shopping malls are specially clustered into a classification, which is used as the cluster to match tourists' interest labels. If tourists' interests tend to approach the cluster in utmost extent, certain amount of shopping malls will be confirmed as the tourist sights to be visited according to tourists' required quantity. In the sample experiment, Wan Da mall, Wangfujing mall, Zhang gong qiao and Changjiang market are set as the tourist sight cluster $C_{(2)}$ for the research. For this cluster, the interest element conditions and labels are also considered and designed in Algorithm 2 of the interest data mining algorithm and Algorithm 3 of the interest tourist sight mining algorithm.

Tourists starts to travel from the point P, visit r quantity of tourist sights, and finally return to the point P, this process forms an integrated nerve chain structure. Each tourist sight motive nerve cell is the node of the nerve chain structure. According to the definition, there will be $r + 2$ nodes in the nerve chain, and the starting point and the terminal point are both the point P. In the chain, there are r quantity of nerve cell nodes. Under the analysis of the geographic spatial distribution, the layout of tourist sights and roads, and the traffic association degree, arbitrary two urban tourist sights are connected by city roads [24], [25]. Thus, between the two tourist sights, the related factors of geographic information data, traffic information data and tourist sight data are formed, that is the defined factors of tourist sight nerve cell motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$. Therefore, arbitrary tourist sight $C_{(1)}$ or point P are all connected with other r quantity of nerve cell nodes, and there exist different factors between two nodes.

In one chain, the motive satisfaction degree could be considered as the signal information stream. The initial signal information of the point P is $X_{(0)}$. In the process of traveling from the starting point P to the first tourist sight $C_{(1)}$, according to the Formula (9), the initial signal information is iterated by the tourist sight nerve cell motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$. After the iteration of all nerve cells, the processed signal information $X_{(1)}$ is output. After visiting the first tourist sight, the tourist will visit the second tourist sight. Input the initial signal information $X_{(1)}$, the mode of signal information input, processing and output is the same as the first section. The iteration process ends with

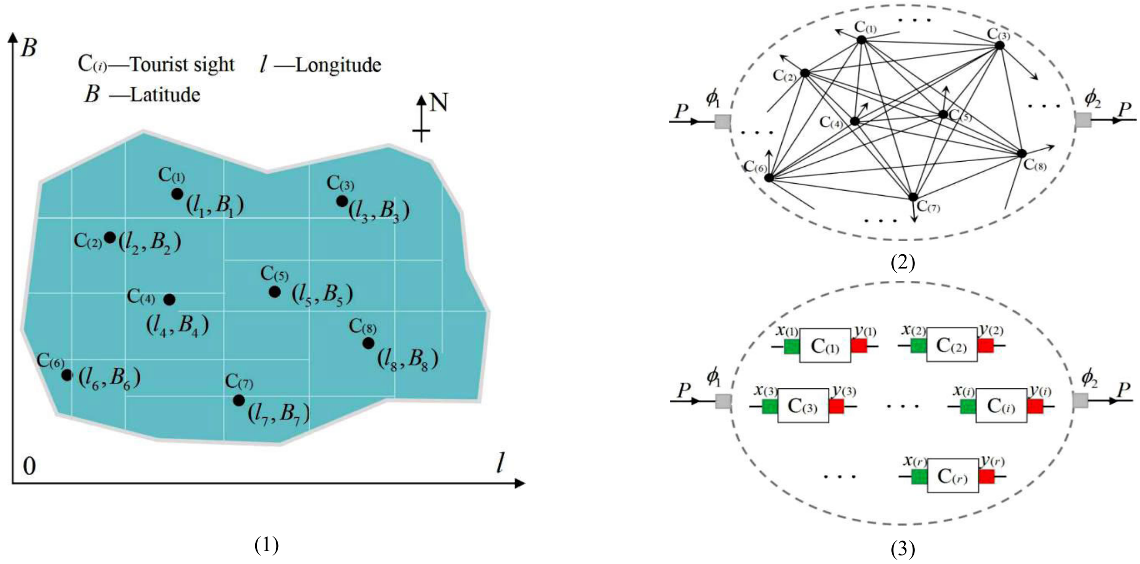


FIGURE 4. City tourist sight geographic spatial distribution, nerve cell network model and the dispersed nerve cell connection interfaces set.

the last tourist sight $C_{(r)}$ and the point P and outputs the whole trip's signal information $X_{(r+1)}$. The signal information transmission mode on the nerve chain is the same as the motive satisfaction degree transmission mode in the tour route, that is, the previous trip process will influence the next one. The final output signal information value $X_{(r+1)}$ could be considered as the ability of the tour route to meet the tourists' interests. When the initial value $X_{(0)}$ is identical while the tour route sequence is different, the same transportation mode will output different motive values on different chains [26], [27]. When the initial value $X_{(0)}$ and the tour route sequence are identical while the transportation mode is different, the same chain will output the different motive values. This is the key issue on tour route recommendation according to the optimal interest tourist sights [28], [34]. As to the definition process, The tour route chain composed by $r + 2$ quantity of nodes could form multiple sequences. The core thought of the tour route chain algorithm model is to traverse to search all the tour route chains and get the globally optimal solution on the motive satisfaction degree.

1) THE FOUNDATION OF THE TOURIST SIGHT NERVE CELL NETWORK MODEL AND THE TOUR ROUTE CHAIN MODEL

According to the tourist sights' geographic spatial distribution and their connection roads distribution, the tourist sight nerve cell network based on the point P and r quantity of nerve cells is set up, shown in Figure 4. Figure 4(1) is the coordinate distribution of the urban tourist sights, in which the horizontal axis is the longitude while the vertical axis is the latitude. The blue area is the city's urban built-up area. The internal gray lines are the main roads of the city, and they divide the built-up area into different blocks. On the aspect of terrain, a city may include plain, hill land or mountain region. Thus, in the same city, different tourist sights or leisure areas may be distributed at different altitudes. Do orthographic projection of the tourist sights on the different altitudes to the two dimension coordinate system,but not considering the terrain and elevation. Here are the three reasons. First, the ferry of the tourists between two tourist sights is influenced by the geographic factors. The motive weights $\omega_{(k)}$ and coefficients $\theta_{(k)}$ which are used to set up the algorithm model all come from the geographic information database and the electronic map. The obtained basic data have considered and contained the terrain and elevation factors. For example, the ferry distance between two tourist sights obtained from the geographic information database and the electronic map, it is the real world distance under the condition of the terrain and elevation. Second, the tourism activities in the downtown area of the city are mainly the horizontal movement process, that is, the tourist sight traveling presents the two dimension relationship between two tourist sights, but not the three dimension relationship. Usually, the three dimension tourism activities mainly include mountain climbing, paragliding, parachuting, hot-air balloon, etc. But the research work in the paper mainly focuses on the downtown two dimension tourist sights, including parks and greenland, venues, shopping malls and amusement parks. The tourism activities in these places are all in the two dimension plane.Third, the algorithm modeling process and the application area are both for the two dimension plane. The basic data, influence factors, the thought for the modeling,and the experiment process, etc, are all in the two dimension plane. In all, the terrain and elevation could not be considered. Each tourist sight $C_{(i)}$ is on the fixed coordinates. Figure 4(2) is the abstract visualized form of the tourist sight nerve cell, and it is the commonly used layout of points for the shortest path searching between the point P and r quantity of tourist sights. In the layout, the left and right points are both P, in the dotted circle, all tourist sights $C_{(i)}$ are distributed in accordance with the longitude and latitude (l, B), the connecting lines among the tourist sights are the nerve chains. The signal information on the chains are influenced by the motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$. The transmission has the two-way conductibility feature, that is, the signal information direction could be forward or backward. Each tourist sight is the signal information receiving point and the output point. On the left side and the right side, there is an interface ϕ_1 and ϕ_2 respectively. The nerve interface ϕ_1 is the starting point from which the iteration signal information is generated and input to the nerve network when tourists start the trip from the point P. The nerve interface ϕ_2 is the terminal point at which the iterated signal information is output after the tourists visit all the tourist sights. In the nerve network system, each tourist sight nerve cell radiates one nerve interconnection line, and this nerve interconnection line has three conditions:

(1) It is connected to the interface ϕ_1;

(2) It is connected to the interface ϕ_2;

(3) It is not connected to interface but to the nerve cell.

If the nerve interconnection line of one tourist sight $C_{(i)}$ is under the condition (15) or (2), the tourist sight $C_{(i)}$ must be the first or the No. $r-1$ tourist sight to be visited in one tour route chain. If the nerve interconnection line of one tourist sight $C_{(i)}$ is under the condition (3), the tourist sight $C_{(i)}$ must be the internal tourist sight to be visited in one tour route chain. In the model, two tourist sights or one tourist and the point P could only be connected by one nerve interconnection line. The r quantity of tourist sights are connected in the single direction way one after one, in which, two tourist sights $C(p)$ and $C(q)$ are connected to the point P, $p, q \in (0, r] \in Z^+$, and the other tourist sights are connected with each other. This connection mode conforms to the process in which the tourists start from the point P, visit the r quantity of tourist sights and return to the point P.

Disperse the tourist sight nerve cell model into the set of single nerve connection interface models, shown in Figure 4(3). Th set includes r quantity of nerve cells, and each nerve cell's code is the tourist sight code $C_{(i)}$. The left green block stores the input nerve cell's code $x_{(i)}$ and the right red block stores the output nerve cell's code $y_{(i)}$. The left and right sides of the point P represent the last input tourist sight code and the first input tourist sight code. The signal information stream enters the tourist sight nerve cell network model from the point P and finally come out, and different tour route chains with different transportation modes will output different signal information values, that is the motive values. Each tourist sight connection interface's input nerve cell and output nerve cell are all different in each tour route chain, which generates the different distributions and forms of the nerve cell network connection interface codes. The input and output connection interfaces meet the following conditions:

(1) The input nerve cell $C_{(x_{(i)})}$ of the tourist sight nerve cell $C_{(i)}$ is the only input nerve cell of $C_{(i)}$, but not other nerve cell's input nerve cell, that is $x_{(i)} \neq x_{(\neg i)}$;

(2) The output nerve cell $C_{(y_{(i)})}$ of the tourist sight nerve cell $C_{(i)}$ is the only output nerve cell of $C_{(i)}$, but not other nerve cell's output nerve cell, that is $y_{(i)} \neq y_{(\neg i)}$;

(3) The code of the tourist sight nerve cell $C_{(i)}$, the code of the input nerve cell $C_{(x_{(i)})}$ and the code of the output nerve cell $C_{(y_{(i)})}$ meet the conditions:

① $0 < \left|x_{(i)} - i\right| < r$;

② $0 < \left|y_{(i)} - i\right| < r$;

③ $0 < \left|x_{(i)} - y\right| < r$;

④ $x_{(i)} \neq i$;

⑤ $y_{(i)} \neq i$;

⑥ $x_{(i)} \neq y_{(i)}$.

The tourist sight nerve cell network model and the dispersed nerve cell connection interfaces set are the basis for setting up the tour route chain model. According to the principle of connection interface coding and tourists' traveling, the tour route chain model is set up. Considering the quantity of the tourist sights to be visited, the traveling mode and process, the signal information transmission direction, etc., there are $P(r, r)$ sorts of tour route chains between the point P. Define the No. i visited tourist sight in the chain as $K_{(i)}$, $i \in (0, r] \in Z^+$. The tourist sight $K_{(i)}$ is the arbitrary tourist sight $C_{(i)}$ in the interest tourist sight vector $\mathbf{C}_{(i)}$, which is determined by the tour sequence. Design the tour route chain model with the $K_{(i)}$ nerve cell element. According to the definition, one chains contains $r+2$ quantity of nerve cells, in which the first and the last one are both the point P, with the internal nerve cells $K_{(i)}$. The initial signal information enters from the point P and it is influenced by each nerve cell's motive weight $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$, and finally the iterated maximum signal information is output.

Figure 5 shows the universal model of the tour route chain, in which the nerve cell $K_{(i)}$ is the empty set Ø. According to the tour route chain algorithm, the specific tourist sight of the nerve cell in each chain as well as the output signal information is output. In the chain, each nerve cell's input signal information is influenced by certain motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$, and is determined by the transportation mode, as Table 1 shows. In the figure, the black bold arrows and blocks form the main chain structure. The upper layer structure of the main chain is the motive weight $\omega_{(k)}$, and the under layer structure of the main chain is the accommodation coefficient $\theta_{(k)}$.

The motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$ between the nerve cell $K_{(i)}$ and $K_{(i+1)}$ are determined by the tourist sights $C(x)$ which are absorbed into the chain and stored in the nerve cells $K_{(i)}$ and $K_{(i+1)}$, that is, the nerve cell $K_{(i)}$ and $K_{(i+1)}$ relate to the relational data of the tourist sight $C(x1)$ and $C(x2)$ in the real world environment. When the tour sequence changes, the stored tourist sight in the nerve cell $K_{(i)}$ will also change, and then the motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$ will change

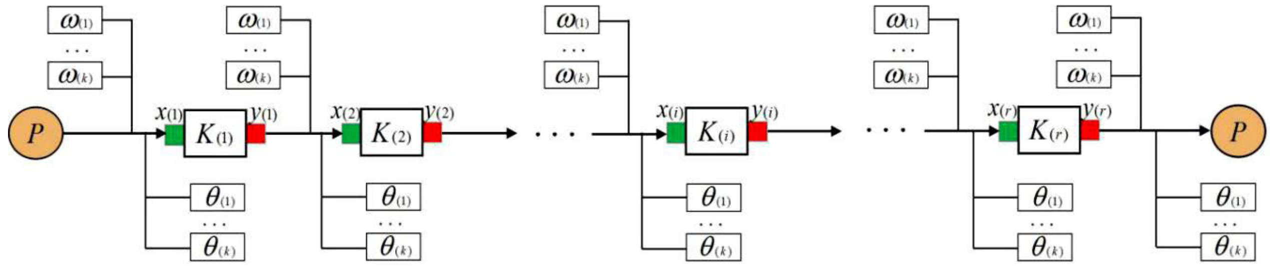


FIGURE 5. The universal model of the tour route chain based on tourist sight nerve cell.

simultaneously. The green block $x_{(i)}$ on the left side of the nerve cell $K_{(i)}$ is the connected previous nerve cell $K(i-1)$, and the red block $y_{(i)}$ on the right side of the nerve cell $K_{(i)}$ is the connected next nerve cell $K_{(i+1)}$.

2) TOUR ROUTE CHAIN ALGORITHM BASED ON THE TOURIST SIGHT MOTIVE NERVE CELL

According to the universal model of the tour route chain and the relationship between the transportation mode and the motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$, the tour route chain algorithm based on the tourist sight motive nerve cell is set up. The basic thought of the algorithm principle is as follows. Under the condition of one transportation mode, the upper layer connection interfaces $\omega_{(k)}$ and under layer connection interfaces $\theta_{(k)}$ are determined by the Table 1. First, the $K_{(i)}$ storage capacity of the tour route is empty set $\emptyset$. Randomly store the tourist sight $C_{(i)}$ into the nerve cell $K_{(i)}$ and make the arbitrary $\forall K_{(i)}$ non-empty, and then an integrated tour route chain is generated. When the input initial signal information is $X_{(0)}$, the signal information is iterated by r quantity of nerve cells and their parameters, and finally the maximum value max $X_{(i)}$ is obtained at the terminal point P. In the process of signal information iteration, each nerve cell relates to one input signal information $X_{(i)}$ and one output signal information $X_{(i+1)}$. Arbitrary nerve cell $\forall K_{(i)}$ and $\forall K_{(\neg i)}$ tourist sights change with each other, it will generate a new tour route chain. Under the condition of the same signal information, the final output signal information will change, too. Traverse all $\mathrm{P}(r, r)$ quantity of tour route chains, output each chain's motive value and finally output the globally optimal solution. The specific algorithm flow is as follows.

Step 1 Set up the tour route chain, and set the storage capacity $K_{(i)}$ is empty set $\emptyset$. Shown in Figure 5, the input nerve cell of arbitrary tourist sight nerve cell $\forall K_{(i)}$ is $x_{(i)}$, and its output nerve cell is $y_{(i)}$.

Step 2 Confirm the tourist's selected transportation mode, and confirm the related upper layer connection interfaces $\omega_{(k)}$ and under layer connection interfaces $\theta_{(k)}$ in the tour route chain according to the transportation mode ξ.

Step 3 Randomly store the tourist sight $C_{(i)}$ into the nerve cell $K_{(i)}$, and $\forall K_{(i)} \neq \emptyset$. The tour route chain $L_{(e)}$ is formed. In the chain, nerve cell $K_{(i)}$ relates to one tourist sight $C_{(i)}$,but their footnotes may not be the same. When the No. e tour route chain is iterated, the iteration time on its nerve cells is noted as $\kappa_{(e,1)}$. When the iteration process of arbitrary one chain $\forall L_{(e)}$ is finished, the quantity of the iterated chains is noted as $\kappa_{(2)}$.

According to the definition and the real world travel process, the time of $\kappa_{(e,1)}$ and $\kappa_{(2)}$ meet the condition $0 < \kappa_{(e,1)} \leq r+1, 0 < \kappa_{(2)} \leq \mathrm{P}(r, r)$. The starting point P is noted as $K_{(0)}$, the terminal point P is noted as $K_{(r+1)}$. The tour route chain $L_{(e)}$ meets the condition $1 < i \leq j < r, i, j, r \in \mathrm{Z}^+$, $L_{(e)} = P, K_{(1)}, K_{(2)}, \ldots, K_{(i)}, \ldots, K_{(j)}, \ldots, K_{(r)}, P$.

Sub-step 1 Iterate the first tour route chain $L(1)$ and output the maximum signal information's related motive value $X^{\xi}_{(r+1)}$. The initial iteration value is $\kappa_{(2)} = 0$, the counting method is $\kappa_{(2)} = \kappa_{(2)} + 1$.

(1) Set the initial signal information is $X^{\xi}_{(0)}$, the initial iteration time is $\kappa_{(1,1)} = 0$. The counting method is $\kappa_{(e,1)} = \kappa_{(e,1)} + 1, e = 0$;

(2) Confirm all the related tourist sights $C_{(i)}$ of the nerve cells $K_{(i)}$ in the first chain $L_{(1)}$ and their input nerve cells and output nerve cells. Note the related tourist sight of the nerve cell $K_{(i)}$ as $C_{(s(i))}$. For example, if the tourist sight $C(6)$ is stored in $K_{(3)}$, then $i = 3, s_{(3)} = 6$. According to the definition, the input nerve cell of $K_{(i)}$ is $K(i-1)$, and the input tourist sight is $x_{(i)} = C_{(s(i-1))}$. The output nerve cell of $K_{(i)}$ is $K_{(i+1)}$, and the output tourist sight is $y_{(i)} = C_{(s(i+1))}$.

(3) Iterate the tour route chain $L_{(1)}$:

① Iterate the interval $L_{(e\sim 1)} = P, K_{(1)}$. Input the initial signal information $X^{\xi}_{(0)}$, embed the upper layer connection interfaces $\omega_{(k)}$ and under layer connection interfaces $\theta_{(k)}$ of the point P and the tourist sight $C_{(s(1))}$. Output the signal information $X^{\xi}_{(1)}$ at the tourist sight nerve cell $K_{(1)}$, noted as $\kappa_{(1,1)} = 1$;

i) If $r = 1$, the point $K_{(2)}$ is P, jump to the step ④;

ii) If $r > 1$, jump to the step ②, continue iterating.

② Iterate the interval $L(e \sim 2) = K_{(1)}, K_{(2)}$. Input the signal information $X^{\xi}_{(1)}$, embed the upper layer connection interfaces $\omega_{(k)}$ and under layer connection interfaces $\theta_{(k)}$ of the tourist sight $C_{(s(1))}$ and $C_{(s(2))}$. Output the signal information $X^{\xi}_{(2)}$ at the tourist sight nerve cell $K_{(2)}$, noted as $\kappa_{(1,1)} = 2$;

i) If $r = 2$, the point $K_{(3)}$ is P, jump to the step ④;

ii) If $r > 2$, jump to the step ③, continue iterating.

③ Iterate the interval $L(e \sim i+1) = K_{(i)}, K_{(i+1)}, i \in (2, r-1]$. Input the signal information $X^{\xi}_{(i)}$, embed the upper layer connection interfaces $\omega_{(k)}$ and under layer connection interfaces $\theta_{(k)}$ of the tourist sight $C_{(s(i))}$ and $C_{(s(i+1))}$. Output the signal information $X^{\xi}_{(i+1)}$ at the tourist sight nerve cell $K_{(i+1)}$, noted as $\kappa_{(1,1)} = i+1$;

i) If $r = i+1$, the point $K_{(i+2)}$ is P, jump to the step ④;

ii) If $r > i+1$, continue iterating until the step ④.

④ Iterate the interval $L_{(e \sim r+1)} = K_{(r)}, P$. Input the signal information $X^{\xi}_{(r)}$, and embed the upper layer connection interfaces $\omega_{(k)}$ and under layer connection interfaces $\theta_{(k)}$ of the tourist sight $C(s(r))$ and the point P. Output the signal information $X^{\xi}_{(r+1)}$ at the terminal point P, noted as $\kappa_{(1,1)} = r+1$. Finish the iteration and output the maximum motive value $X^{\xi}_{(r+1)}$ of the tour route chain $L_{(1)}$, noted as $X^{\xi}_{L_{(1)}}$.

(4) Set up the motive value storage vector $\mathbf{X}^{\xi}_{L_{(e)}}$ for the tour route chain with the dimension $1 \times \mathrm{P}(r,r)$. Initialize the vector $\mathbf{X}^{\xi}_{L_{(e)}}$ as the zero set. Store $X^{\xi}_{L_{(1)}}$ into the first element $\mathbf{X}^{\xi}_{1}$ of the vector $\mathbf{X}^{\xi}_{L_{(e)}}$.

Sub-step 2 Iterate the second tour route chain $L_{(2)}$ and output the maximum signal information's related motive value $X^{\xi}_{(r+1)}$. The initial iteration value is $\kappa_{(2)} = 1$, the counting method is $\kappa_{(2)} = \kappa_{(2)} + 1$.

(1) Randomly choose arbitrary tourist sight nerve cell $\forall K_{(i)}$ in the chain $L_{(1)}, 0 < i < r, i \in \mathrm{Z}^{+}$. The tourist sight, the input and the output connection interfaces are $C_{(s(i))}, x_{(i)} = C_{(s(i-1))}, y_{(i)} = C_{(s(i+1))}$;

(2) Randomly choose the arbitrary tourist sight nerve cell $\forall K_{(j)}$ in the chain $L_{(1)}, 0 < j < r, j \in \mathrm{Z}^{+}$ and $i \neq j$. The tourist sight, the input and output connection interfaces are $C_{(s(j))}, x_{(j)} = C_{(s(j-1))}, y_{(j)} = C_{(s(j+1))}$, and then:

① If $|i-j| = 1$:

the related tourist sights $C_{(s(i))}$ and $C_{(s(j))}$ of the nerve cells $K_{(i)}$ and $K_{(j)}$ are connected in the chain, and they are the input or output nerve cells respectively.

② If $|i-j| > 1$:

the related tourist sights $C_{(s(i))}$ and $C_{(s(j))}$ of the nerve cells $K_{(i)}$ and $K_{(j)}$ are not connected in the chain, and they are not the input or output nerve cells respectively.

(3) Exchange the storage location of the tourist sights of the nerve cells $K_{(i)}$ and $K_{(j)}$. Make the nerve cell $K_{(i)}$ store the tourist sight $C_{(s(j))}$, and the nerve cell $K_{(j)}$ store the tourist sight $C_{(s(i))}$. And the input and output connection interfaces which have changed are $y_{(i-1)} = C_{(s(j))}, x_{(j+1)} = C_{(s(i))}$.

Then:

① If $i-j = 1$, then $x_{(i)} = C_{(s(i))}, y_{(j)} = C_{(s(j))}$;

② If $j-i = 1$, then $x_{(j)} = C_{(s(j))}, y_{(i)} = C_{(s(i))}$.

(4) Exchange the tourist sights of the nerve cells $K_{(i)}$ and $K_{(j)}$ and form the new chain $L_{(2)}$. Use the same method in the Sub-step 1 and output the maximum motive value $X^{\xi}_{(r+1)}$ of the chain $L_{(2)}$, noted as $X^{\xi}_{L_{(2)}}$.

(5) Compare the value $X^{\xi}_{L_{(1)}}$ with $X^{\xi}_{L_{(2)}}$:

① If $X^{\xi}_{L_{(1)}} \geq X^{\xi}_{L_{(2)}}$, keep the storage of $X^{\xi}_{L_{(1)}}$ in the first element $\mathbf{X}^{\xi}_{1}$ of the vector $\mathbf{X}^{\xi}_{L_{(e)}}$ unchanged, and store $X^{\xi}_{L_{(2)}}$ into the second element $\mathbf{X}^{\xi}_{2}$ of the vector $\mathbf{X}^{\xi}_{L_{(e)}}$;

② If $X^{\xi}_{L_{(1)}} < X^{\xi}_{L_{(2)}}$, descend to store $X^{\xi}_{L_{(1)}}$ into the second element $\mathbf{X}^{\xi}_{2}$ of the vector $\mathbf{X}^{\xi}_{L_{(e)}}$, and store $X^{\xi}_{L_{(2)}}$ into the first element $\mathbf{X}^{\xi}_{1}$ of the vector $\mathbf{X}^{\xi}_{L_{(e)}}$.

Sub-step 3 According to the method of the random exchanging nerve cell tourist sights, delete the repeated chains and store the new one. Iterate the No. e tour route chain $L_{(e)}, e \in (3, \mathrm{P}(r,r)] \subset \mathrm{Z}^{+}$. Output the related maximum motive value $X^{\xi}_{(r+1)}$ of the signal information. The initial iteration value is $\kappa_{(2)} = e-1$, and the counting principle is $\kappa_{(2)} = \kappa_{(2)} + 1$.

(1) Output e quantity of motive values $X^{\xi}_{L(1)}, X^{\xi}_{L(2)}, \ldots, X^{\xi}_{L(e)}$. According to the method of Sub-step 2, step (5), descend to store the motive values $X^{\xi}_{L_{(1)}}, X^{\xi}_{L_{(2)}}, \ldots, X^{\xi}_{L_{(e)}}$ into the top e quantity of elements in the vector $\mathbf{X}^{\xi}_{L_{(e)}}$.

① If $e < \mathrm{P}(r,r)$, and then continue iterating and store the value $X^{\xi}_{L_{(e)}}, \kappa_{(2)} = \kappa_{(2)} + 1$;

② If $e = \mathrm{P}(r,r)$, and then jump to the step (2).

(2) Output the value $\kappa_{(2)} = \mathrm{P}(r,r)$, and obtain the vector $\mathbf{X}^{\xi}_{L_{(e)}}$ with descending order elements. Its rank is $rank(\mathbf{X}^{\xi}_{L_{(e)}}) = \mathrm{P}(r,r)$.

Step 4 Output the priority sequence of the tour route chains. According to the full rank vector $\mathbf{X}^{\xi}_{L_{(e)}}$, output the priority sequence of the tour route chains by the motive values $X^{\xi}_{L_{(e)}}$. And then, the motive value $X^{\xi}_{L_{(e)}}$ of the first element $\mathbf{X}^{\xi}_{1}$ in the vector $\mathbf{X}^{\xi}_{L_{(e)}}$ relates to the maximum motive value of all the tour route chains, and it has the strongest ability to match tourists' motive. It is the globally optimal solution while other chains are the sub-optimal ones.

The following pseudo-code is for the optimal tour route chain algorithm.

IV. EXPERIMENT AND THE RESULT ANALYSIS

In order to testify the algorithm's feasibility and practicalness in the real world environment, an experiment is carried out and the result is output based on tourists' interests and the real world tourism geographic spatial environment [44], [45]. Meanwhile, the commonly used shortest path searching algorithms are set as the control group to compare with the developed algorithm. In this research, the Leshan city is taken as the experimental city. As an international tourism city, Leshan have multiple and abundant tourism resources. Its urban tourist sights are discretely distributed. All the tourist sights are connected by city roads and avenues. Since Leshan city's built-up area is much smaller than other first-tier cities, the tourists could choose any one of the three transportation modes to travel in the city, including walking (or cycling), taking public bus, taking taxi (or self-driving) [29], [31]. In this research, Leshan city's urban tourist sights are taken as the experimental subjects. Based on the tourists' interests,

Algorithm 4 The Optimal Tour Route Chain Algorithm

Step 1: Set up the empty $\emptyset$ tour route chain model.
  As to $\forall K_{(i)}$, the input nerve is $x_{(i)}$, the output nerve is $y_{(i)}$.
Step 2: Confirm mode ξ, $\omega_{(k)}$ and $\theta_{(k)}$.
Step 3: Iterate the tour route chain.
  For each $e \in (0, \mathrm{P}(r, r)] \subset \mathrm{Z}^{+}, \kappa_{(2)} = \kappa_{(2)} + 1$:
   For each interval $L_{(e\sim i+1)} = K_{(i-1)}, K_{(i)}, i \in (0, r+2]$:
    Set the initial value $X^{\xi}_{(0)}$.
    Confirm $x_{(i)}$ and $y_{(i)}$ for each node nerve cell.
    Output each nerve cell's value $X^{\xi}_{(i)}$.
    $\kappa_{(e,1)} = \kappa_{(e,1)} + 1$.
   If $\kappa_{(2)} = \mathrm{P}(r, r)$, stop the iteration.
    Sort $X^{\xi}_{L_{(1)}}, X^{\xi}_{L_{(2)}}, \ldots, X^{\xi}_{L_{(e)}}$ in descending order.
    Get the vector $\mathbf{X}^{\xi}_{L_{(e)}}$.
Step 4: Get the optimal tour route chain by $\mathbf{X}^{\xi}_{L_{(e)}}$.

the optimal tour route under each transportation mode is obtained in the experiment. The experiment's basic principle is as follows. First, the urban tourist sights are clustered and analyzed according to their feature attributes, and obtain the tourist sight clusters and the subordinate degree matrix. Then, the tourist interest matrix is set up, and each interest vector of the cluster is obtained, which is used to mine the matched tourist sights and output the tourist sight vector. Take the mined tourist sights as nodes in the tour route chain, and the tour route chains' signal information under the three transportation modes are studied respectively to obtain the globally optimal solution. Compare with the control group algorithms and analyze the results. Set the evaluation factors for the evaluation of the proposed method as follows.

(1) Feature attribute matching factor $k_{(1)}$;
(2) Tourist sight mining objective function $G_{(\mathbf{K}(r),c(i))}$;
(3) Motive weight $\omega_{(k)}$;
(4) Coefficient factor $\theta_{(k)}$;
(5) The signal information motive value of each nerve cell $X^{\xi}_{(i)}$;
(6) The signal information motive value of the tour route chain $X^{\xi}_{(r+1)}$;
(7) The increased signal information motive value between two nerve cells $\Delta X^{\xi}_{(i)}$;
(8) The signal information motive difference value between the experimental group algorithm and the control group algorithm;
(9) The space complexity of the algorithm;
(10) The time complexity of the algorithm.

A. DATA COLLECTION

According to the basic thought of the experiment design, the collected data for the experiment include Leshan city's map data, urban tourist sights, longitude and latitude data, traffic information data and tourist sight knowledge text data, etc. Take the interest data provided by tourists as the primary data to obtain the tourist sight clusters, tourist sight objective function values, tourist sight vector and the globally optimal solution under the three transportation modes [42], [43].

1) THE TOURIST SIGHT DATA

Take the 15 quantity of typical tourist sights $c_{(i)}$ in Leshan city's urban area as the experimental subjects, and they form the tourist sight domain $C, n = 15, 0 < i \leq 15, i \in \mathrm{Z}^{+}$. C ={$c_{(1)}$: Leshan Giant Buddha; $c_{(2)}$: Wan Da mall; $c_{(3)}$: the cultural center; $c_{(4)}$: Wangfujing mall; $c_{(5)}$: the children amusement park; $c_{(6)}$: Tian gong kai wu; $c_{(7)}$: LvXin park; $c_{(8)}$: Leshan museum; $c_{(9)}$: Zhang gong qiao; $c_{(10)}$: Jia zhou chang juan; $c_{(11)}$: Leshan amusement park; $c_{(12)}$: Hai tang park; $c_{(13)}$: Leshan art museum; $c_{(14)}$: Changjiang market; $c(15)$: Yuancheng amusement park}. Take Leshan city's urban main roads and avenues as the basic structure to confirm the geographic location of the tourist sights $c_{(i)}$, shown in Figure 6. Figure 6(1) shows the tourist sight $c_{(i)}$ distribution on the map. Figure 6(2) shows the abstract nerve cell distribution of the tourist sight nerve cell network model.

2) TOURIST SIGHT BASIC INFORMATION DATA

According to the developed tourist sight clustering algorithm, tourist sight mining algorithm and the tour route chain algorithm, the tourist sight basic information data include feature attribute keyword label, travel time schedule, travel cost, attraction index, longitude and latitude, etc. The knowledge text data for each tourist sight $c_{(i)}$ is crawled from the Internet, and the feature attribute keyword labels are extracted from the text data to calculate feature attribute matching factors $k_{(1)}$. From the tourism website, the best travel time schedule $k_{(2)}$ (*hour*), travel cost $k_{(3)}$ (*yuan*) and attraction index $k_{(4)}$ are obtained. The longitude and latitude (l, B) data are obtained from the website GPSspg. The Table 2 shows the basic tourist sight $c_{(i)}$ information of the domain r.

3) THE GENERATION OF THE TOURIST SIGHT CLUSTERS

According to the clustering algorithm, the feature attribute label vector $\mathbf{c}_{(i)}$ for each tourist sight $c_{(i)}$ in the domain C is confirmed. Set the feature attribute threshold value as $T_{(c_{(i)},c_{(j)})} = 0.200$. The label vector maximum matching value δ and the feature attribute clustering correlation degree $D_{(c_{(i)},c_{(j)})^{-1}}$ between the tourist sight $c_{(i)}$ and $c_{(j)}$ are calculated, shown in the Table 3. Compare the Table 3 data with the feature attribute threshold value $T_{(c_{(i)},c_{(j)})}$, and the closeness between two tourist sights could be obtained. A $p \times \max n(i)$ dimension matrix $\mathbf{C}_{(p\times \max n(i))}$ is set up to store tourist sights $c_{(i)}$ of the domain C, shown in Formula(13). According to the Table 3 data and the algorithm flow, the selected seed points are $c_{(1)} \sim \Delta c_{(1)}$: Leshan giant Buddha; $c_{(2)} \sim \Delta c_{(2)}$: Wan Da mark; $c_{(3)} \sim \Delta c_{(3)}$: the cultural center; $c_{(5)} \sim \Delta c_{(5)}$: the children amusement park. Starting from each seed point $\Delta c_{(i)}$, search the subordinate degree of other tourist sights to the point $\Delta c_{(i)}$ and set up the subordinate degree

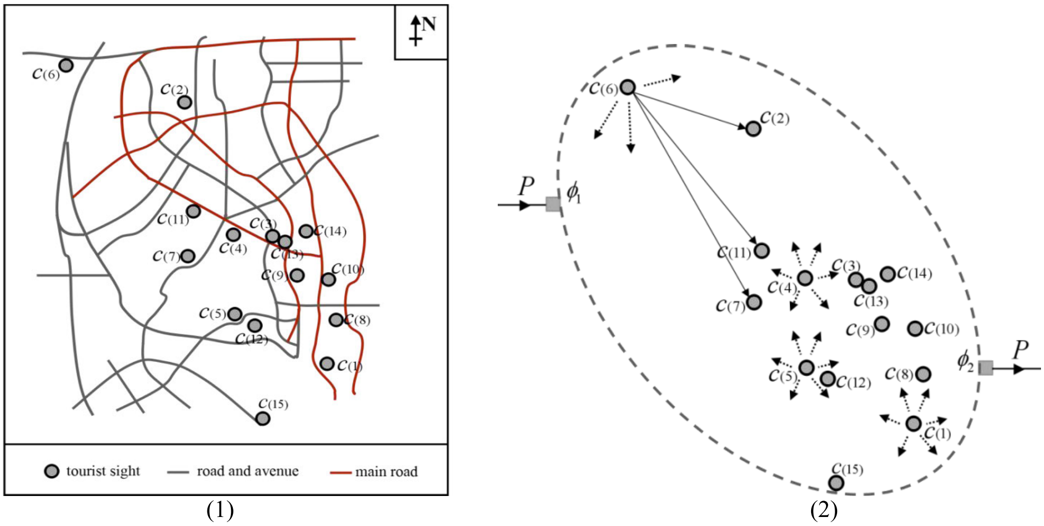


FIGURE 6. The distribution of the tourist sights and the abstract nerve cell nodes distribution extracted from the network model.

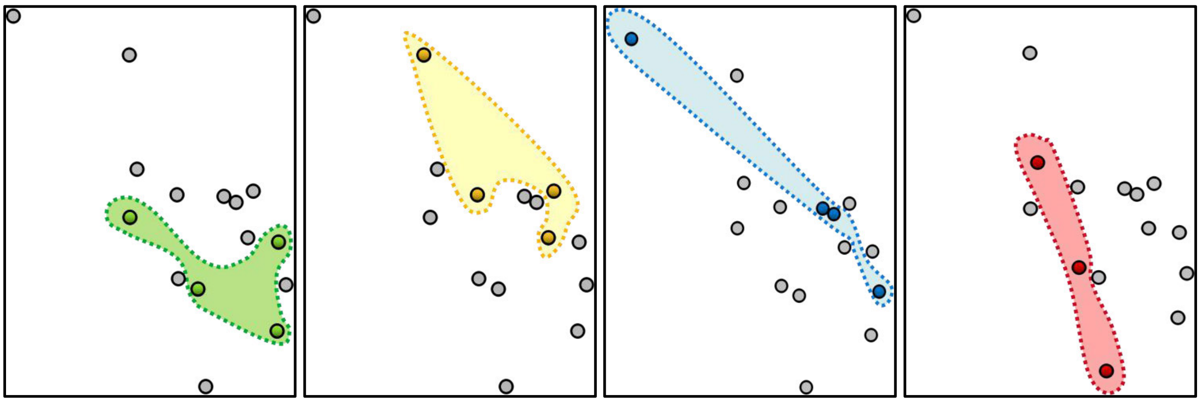
TABLE 2. Tourist sight $c_{(i)}$ basic information data of the domain C.

	l,B	$k_{(2)}$	$k_{(3)}$	$k_{(4)}$		l,B	$k_{(2)}$	$k_{(3)}$	$k_{(4)}$
$c_{(1)}$	103.772，29.548	2.0~3.0	80	0.912	$c_{(9)}$	103.769，29.569	1.0~3.0	0	0.901
$c_{(2)}$	103.737，29.619	1.5~2.0	0	0.829	$c_{(10)}$	103.775，29.571	1.5~2.0	0	0.826
$c_{(3)}$	103.762，29.581	1.5~2.0	0	0.729	$c_{(11)}$	103.742，29.588	2.0~3.0	10	0.765
$c_{(4)}$	103.751，29.583	1.0~3.0	0	0.598	$c_{(12)}$	103.764，29.556	1.5~2.0	0	0.778
$c_{(5)}$	103.763，29.556	2.0~3.0	20	0.612	$c_{(13)}$	103.763，29.582	1.5~2.0	0	0.602
$c_{(6)}$	103.704，29.616	1.0~2.0	30	0.512	$c_{(14)}$	103.766，29.589	1.5~2.0	0	0.524
$c_{(7)}$	103.732，29.578	1.0~2.5	0	0.711	$c_{(15)}$	103.754，29.521	2.0~3.0	50	0.571
$c_{(8)}$	103.776，29.563	1.5~2.0	0	0.596					

TABLE 3. Feature attribute correlation degree $D_{(c_{(i)},c_{(j)})^{-1}}$ of the tourist sight $c_{(i)}$ and $c_{(j)}$.

	$c_{(1)}$ / $c_{(9)}$	$c_{(2)}$ / $c_{(10)}$	$c_{(3)}$ / $c_{(11)}$	$c_{(4)}$ / $c_{(12)}$	$c_{(5)}$ / $c_{(13)}$	$c_{(6)}$ / $c_{(14)}$	$c_{(7)}$ / $c_{(15)}$	$c_{(8)}$ /- -
$c_{(1)}$	- - / 1.000	0.500 / 0.167	0.333 / 0.500	0.500 / 0.200	0.500 / 0.500	0.333 / 1.000	0.143 / 0.500	0.333 / - -
$c_{(2)}$	0.500 / 0.167	- - / 0.500	1.000 / 0.500	0.125 / 0.500	1.000 / 0.500	0.500 / 0.111	0.500 / 1.000	1.000 / - -
$c_{(3)}$	0.333 / 1.000	1.000 / 0.500	- - / 1.000	1.000 / 0.500	1.000 / 0.143	0.167 / 0.500	0.500 / 0.500	0.167 / - -
$c_{(4)}$	0.500 / 0.125	0.125 / 0.500	1.000 / 1.000	- - / 0.500	1.000 / 0.500	0.500 / 0.111	0.500 / 1.000	1.000 / - -
$c_{(5)}$	0.500 / 1.000	1.000 / 1.000	1.000 / 0.167	1.000 / 0.500	- - / 0.500	1.000 / 1.000	0.500 / 0.167	0.500 / - -
$c_{(6)}$	0.333 / 0.500	0.500 / 0.333	0.167 / 1.000	0.500 / 0.500	1.000 / 0.111	- - / 1.000	0.500 / 0.500	0.111 / - -
$c_{(7)}$	0.143 / 0.500	0.500 / 0.143	0.500 / 0.500	0.500 / 0.111	0.500 / 0.500	0.500 / 1.000	- - / 0.500	0.500 / - -
$c_{(8)}$	0.333 / 1.000	1.000 / 0.500	0.167 / 1.000	1.000 / 0.500	0.500 / 0.143	0.111 / 1.000	0.500 / 1.000	- - / - -
$c_{(9)}$	1.000 / - -	0.167 / 0.500	1.000 / 0.500	0.125 / 1.000	1.000 / 0.500	0.500 / 0.167	0.500 / 1.000	1.000 / - -
$c_{(10)}$	0.167 / 0.500	0.500 / - -	0.500 / 1.000	0.500 / 0.111	1.000 / 0.500	0.333 / 1.000	0.143 / 0.500	0.500 / - -
$c_{(11)}$	0.500 / 0.500	0.500 / 1.000	1.000 / - -	1.000 / 0.500	0.167 / 0.500	1.000 / 1.000	0.500 / 0.111	1.000 / - -
$c_{(12)}$	0.200 / 1.000	0.500 / 0.111	0.500 / 0.500	0.500 / - -	0.500 / 1.000	0.500 / 1.000	0.111 / 0.500	0.500 / - -
$c_{(13)}$	0.500 / 0.500	0.500 / 0.500	0.143 / 0.500	0.500 / 1.000	0.500 / - -	0.111/ 1.000	0.500 / 1.000	0.143 / - -
$c_{(14)}$	1.000 / 0.167	0.111 / 1.000	0.500 / 1.000	0.111 / 1.000	1.000 / 1.000	1.000 / - -	1.000 / 0.500	1.000 / - -
$c_{(15)}$	0.500 / 1.000	1.000 / 0.500	0.500 / 0.111	1.000 / 0.500	0.167 / 1.000	0.500 / 0.500	0.500 / - -	1.000 / - -

matrix $\xi_{(p\times\max n(i))}$, shown in Formula (14). In the matrix, the black bold value 1 represents the seed point location. The abstract four clusters are shown in Figure 7, and the four clusters are as follows.



(1) cluster $C_{(1)}$ (2) cluster $C_{(2)}$ (3) cluster $C_{(3)}$ (4) cluster $C_{(4)}$

FIGURE 7. The output tourist sight clusters from the clustering algorithm.

$C_{(1)}$ ={$c_{(1)}$: Leshan Giant Buddha; $c_{(7)}$: LvXin park; $c_{(10)}$: Jia zhou chang juan; $c_{(12)}$: Hai tang park};

$C_{(2)}$ ={$c_{(2)}$: Wan Da mall; $c_{(4)}$: Wangfujing mall; $c_{(9)}$: Zhang gong qiao; $c_{(14)}$: Changjiang market};

$C_{(3)}$ ={$c_{(3)}$:the cultural center; $c_{(6)}$: Tian gong kai wu; $c_{(8)}$: Leshan museum; $c_{(13)}$: Leshan art museum};

$C_{(4)}$ ={$c_{(5)}$: the children amusement park; $c_{(11)}$: Leshan amusement park; $c_{(15)}$: Yuancheng amusement park}. (13) and (14), as shown at the bottom of the page.

B. THE RESULT OF THE MINED TOURIST SIGHTS MATCHING THE TOURISTS' INTERESTS

Based on the tourist sight basic information data and the tourist sight clusters, the tourist sights that match tourists' interests are mined and obtained. In the experiment, the interest feature attribute data vector $\mathbf{W}_{(t)}$ is obtained by the feature attribute label vector $\mathbf{c}_{(i)}$. The confirmed vector $\mathbf{W}_{(t)}$ is $\mathbf{W}_{(t)}$ = {$\mathrm{W}_{(11)}$: park; $\mathrm{W}_{(12)}$: greenland; $\mathrm{W}_{(13)}$: sightseeing; $\mathrm{W}_{(14)}$: flower; W(1): scenery; $\mathrm{W}_{(21)}$: leisure; $\mathrm{W}_{(22)}$: shopping; $\mathrm{W}_{(23)}$: synthesis; $\mathrm{W}_{(24)}$: entertainment; $\mathrm{W}_{(25)}$: catering; $\mathrm{W}_{(31)}$:venue; $\mathrm{W}_{(32)}$: memorial; $\mathrm{W}_{(33)}$: history; $\mathrm{W}_{(34)}$: science; $\mathrm{W}_{(35)}$ knowledge; $\mathrm{W}_{(36)}$: nature; $\mathrm{W}_{(41)}$: amusement; $\mathrm{W}_{(42)}$: theme park; $\mathrm{W}_{(43)}$: swimming; $\mathrm{W}_{(44)}$: athletics; $\mathrm{W}_{(45)}$: cartoon;}. From the vector $\mathbf{W}_{(t)}$, tourists choose certain quantity of labels and they are composed of the interest matrix $\mathbf{W}_{(p\times\max p(u))}$, in which each row represents the interest label vector $\mathbf{W}_{u(1\times\max p(u))}$, shown in Formula (15). The matrix $\mathbf{W}_{(p\times\max p(u))}$ is the basis for the tourists to get the tourist sights to be visited, and it is used for text mining to get the feature attribute matching factor $k_{(1)}$. Meanwhile, considering the self conditions and time schedule, the tourist provides the requirements in which the expected travel time $k_{(2*)} \leq 1.5(hour)$, the travel cost $k_{(3*)} = 0(yuan)$, and the tourist sight attraction index $k_{(4*)} \geq 0.700$. According to the basic interest matrix $\mathbf{W}_{(p\times\max p(u))}$ data, the experiment sets that this tourist expects to visit two tourist sights of the $C_{(1)}$ cluster, one $C_{(2)}$ cluster tourist sight and one $C_{(3)}$ cluster tourist sight. The feature attribute matching factor $k_{(1)}$ is obtained by the tourist sight text mining, combing with the Table 2 tourist sight feature attributes $k_{(2)}, k_{(3)}$ and $k_{(4)}$, the tourist sight mining objective function $G_{(\mathbf{K}(r),c(i))}$ is obtained. According to the definition, set the disturbance parameters as $\varepsilon_{(1)} = 1, \varepsilon_{(2)} = 1, \varepsilon_{(3)} = 0.1, \varepsilon_{(4)} = 1$. Through the tourist sight mining algorithm, the feature attribute matching factors $k_{(1)}$ values and objective function $G_{(\mathbf{K}(r),c(i))}$ values of the interest label vector $\mathbf{W}_{u(1\times 4)}$ are calculated and obtained,

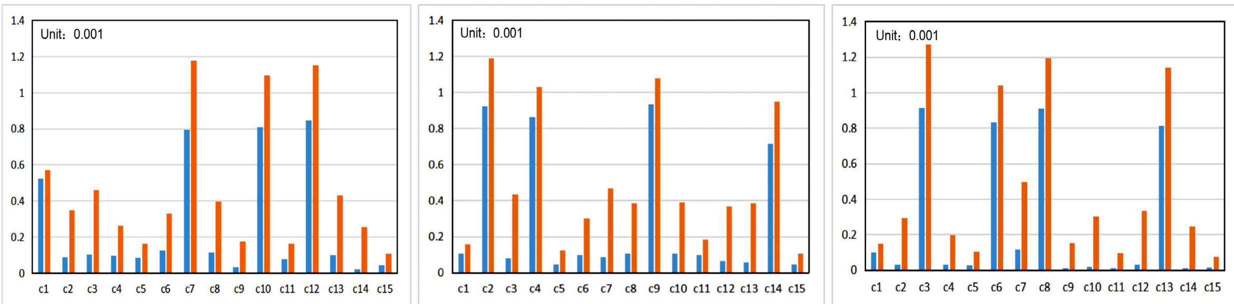
$$\mathbf{C}_{(4\times4)} = \begin{bmatrix} c_{(1)} & c_{(2)} & c_{(3)} & c_{(4)} \\ c_{(5)} & c_{(6)} & c_{(7)} & c_{(8)} \\ c_{(9)} & c_{(10)} & c_{(11)} & c_{(12)} \\ c_{(13)} & c_{(14)} & c_{(15)} & 0 \end{bmatrix} \tag{13}$$

$$\xi_1(4\times4) = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\xi}_2(4\times4) = \begin{bmatrix} 0 & \mathbf{1} & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \boldsymbol{\xi}_3(4\times4) = \begin{bmatrix} 0 & 0 & \mathbf{1} & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{\xi}_4(4\times4) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{14}$$

TABLE 4. The feature attribute matching factors $k_{(1)}$ values and the objective function $G_{(\mathbf{K}(r),c(1))}$ values.

	$k_{(1)}$			$G(\mathbf{K}(r),c(i))$				$k_{(1)}$			$G(\mathbf{K}(r),c(i))$		
	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$		$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$
$c(1)$	0.523	0.106	0.098	0.573	0.156	0.148	$c(9)$	0.032	0.935	0.011	0.175	1.078	0.154
$c(2)$	0.087	0.925	0.032	0.351	1.189	0.296	$c(10)$	0.810	0.105	0.018	1.095	0.390	0.303
$c(3)$	0.102	0.078	0.913	0.460	0.436	1.271	$c(11)$	0.077	0.098	0.010	0.163	0.184	0.096
$c(4)$	0.098	0.865	0.033	0.264	1.031	0.199	$c(12)$	0.846	0.065	0.032	1.151	0.370	0.337
$c(5)$	0.086	0.045	0.028	0.164	0.123	0.106	$c(13)$	0.099	0.056	0.813	0.429	0.386	1.143
$c(6)$	0.125	0.098	0.834	0.330	0.303	1.039	$c(14)$	0.023	0.717	0.011	0.257	0.951	0.245
$c(7)$	0.796	0.088	0.116	1.176	0.468	0.496	$c(15)$	0.046	0.045	0.016	0.107	0.106	0.077
$c(8)$	0.114	0.104	0.912	0.396	0.386	1.194							



(1) Vector $\mathbf{W}_1$ (2) Vector $\mathbf{W}_2$ (3) Vector $\mathbf{W}_3$

FIGURE 8. The distribution figure of the $k_{(1)}$ values and objective function $G_{(\mathbf{K}(r),c(1))}$ values of the label vector.

shown in the Table 4. From the data, the distribution figure of the $k_{(1)}$ values and objective function $G_{(\mathbf{K}(r),c(i))}$ values is obtained, shown in Figure 8. The Figure 8(1)~(3) represent the $k_{(1)}$ values and $G_{(\mathbf{K}(r),c(i))}$ values of the vector $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$, in which the blue data bars represent the $k_{(1)}$ values, the orange data bars represent the $G_{(\mathbf{K}(r),c(i))}$ values. In the Figure 8(1), the blue lines relate to the $k_{(1)}$ values on the $\mathbf{W}_1$ of the tourist sights $c_{(1)} \sim c(15)$, and the orange lines relate to the $G_{(\mathbf{K}(r),c(i))}$ values on the $\mathbf{W}_1$ of the tourist sights $c_{(1)} \sim c_{(1)}$. In the Figure 8(2), the blue lines relate to the $k_{(1)}$ values on the $\mathbf{W}_2$ of the tourist sights $c_{(1)} \sim c_{(1)}$, and the orange lines relate to the $G_{(\mathbf{K}(r),c(i))}$ values on the $\mathbf{W}_2$ of the tourist sights $c_{(1)} \sim c_{(1)}$. In the Figure 8(3), the blue lines relate to the $k_{(1)}$ values on the $\mathbf{W}_3$ of the tourist sights $c_{(1)} \sim c_{(1)}$, and the orange lines relate to the $G_{(\mathbf{K}(r),c(i))}$ values on the $\mathbf{W}_3$ of the tourist sights $c_{(1)} \sim c_{(1)}$. According to the tourist sight descending order vector $\mathbf{V}_{(uv)}$, combining with tourists' interests, the Table 4 data and Figure 8 data distribution, the tourist sight vector $\mathbf{C}_{(i)}$ is obtain, $\mathbf{C}_{(i)} = \{c_{(7)}, c_{(12)}, c_{(2)}, c_{(3)}\}$.

$$\mathbf{W}_{(4\times4)} = \begin{bmatrix} W_{(11)} & W_{(14)} & W_{(1)} & 0 \\ W_{(21)} & W_{(22)} & W_{(24)} & W_{(25)} \\ W_{(32)} & W_{(35)} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (15)$$

C. THE OUTPUT RESULT OF THE TOUR ROUTE CHAIN ALGORITHM BASED ON MULTIVARIATE TRANSPORTATION MODES

Based on the tourist sight mining, the tour route chains and signal information motive values under the conditions of multivariate transportation modes and the nerve nodes of vector $\mathbf{C}_{(i)}$ tourist sights are output. According to the longitude and latitude of the point P, the starting and terminal point of the chain are confirmed. The internal nerve cell nodes are the longitude and latitude of the vector $\mathbf{C}_{(i)}$ tourist sights. When tourists choose different transportation modes, the tour route chain will be influenced by the travel sequence and the different factors, which results in different motive values. The tour route chain's motive value will change if any one of the factors on tourists' interests, the point P coordinate, the vector $\mathbf{C}_{(i)}$ tourist sights and the traffic transportation mode changes. It is also influenced by the geographic information data and traffic information data. In the experiment, the motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$ are obtained under the condition that the tourists' interests, the point P coordinate and the vector $\mathbf{C}_{(i)}$ tourist sights are confirmed. Considering the Table 1 data, the signal information motive value descending order vector $\mathbf{X}^{\xi}_{L(e)}$ under each transportation mode is output. To testify the algorithm's feasibility and advantages, the experiment chooses three shortest path searching algorithms as the control group, and each

TABLE 5. The motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$.

	$\omega_{(1)}$	$\omega_{(2)}$	$\omega_{(3)}$	$\omega_{(4)}$	$\omega_{(5)}$	$\omega_{(6)}$	$\theta_{(1)}$	$\theta_{(2)}$	$\theta_{(3)}$	$\theta_{(4)}$
$Pc_{(2)}$	0.208	0.030	5.000	1.364	0.133	0.632	8.000	0.200	0.167	1.000
$Pc_{(3)}$	0.167	0.054	5.000	1.091	0.103	0.544	8.000	0.200	0.250	2.000
$Pc_{(7)}$	0.192	0.031	5.000	1.006	0.128	0.365	9.000	0.200	0.233	2.000
$Pc_{(12)}$	0.112	0.069	5.000	1.034	0.071	0.287	8.000	0.200	0.269	3.000
$c_{(2)}c_{(3)}$	0.175	0.045	5.000	1.071	0.117	0.233	8.000	0.150	0.298	3.000
$c_{(2)}c_{(7)}$	0.182	0.041	5.000	1.002	0.128	0.431	9.000	0.150	0.196	1.000
$c_{(2)}c_{(12)}$	0.115	0.069	5.000	1.003	0.077	0.129	10.000	0.150	0.188	3.000
$c_{(3)}c_{(7)}$	0.286	0.030	5.000	1.006	0.191	0.762	5.000	0.100	0.098	1.000
$c_{(3)}c_{(12)}$	0.227	0.025	5.000	1.714	0.152	0.512	4.000	0.100	0.101	2.000
$c_{(7)}c_{(12)}$	0.189	0.039	5.000	1.200	0.126	0.416	5.000	0.150	0.167	2.000

TABLE 6. The output of nerve cell's signal information motive values $X^{\xi}_{(i)}$, the tour route chains' signal information motive values $X^{\xi}_{(r+1)}$ and the increased signal information motive values $\Delta X^{\xi}_{(i)}$.

	Tour route chain	$X^{\xi}_{(i)}$					$\Delta X^{\xi}_{(i)}$				
		$K_{(1)}$	$K_{(2)}$	$K_{(3)}$	$K_{(4)}$	$P{:}K_{(5)}$	$PK_{(1)}$	$K_{(1)}K_{(2)}$	$K_{(2)}K_{(3)}$	$K_{(3)}K_{(4)}$	$K_{(4)}P$
						$\xi=1$					
1	$Pc_{(12)}c_{(2)}c_{(7)}c_{(3)}P$	1.722	3.008	4.699	7.403	12.556	0.722	1.286	1.691	2.704	5.153
2	$Pc_{(3)}c_{(7)}c_{(2)}c_{(12)}P$	1.627	2.530	3.938	7.009	12.550	0.627	0.903	1.408	3.071	5.541
3	$Pc_{(12)}c_{(7)}c_{(2)}c_{(3)}P$	1.722	2.669	4.159	6.679	11.320	0.722	0.947	1.49	2.52	4.641
4	$Pc_{(12)}c_{(2)}c_{(3)}c_{(7)}P$	1.722	3.008	4.808	7.576	11.289	0.722	1.286	1.8	2.768	3.713
5	$Pc_{(3)}c_{(2)}c_{(7)}c_{(12)}P$	1.627	2.564	3.992	6.253	11.188	0.627	0.937	1.428	2.261	4.935
						$\xi=2$					
1	$Pc_{(3)}c_{(7)}c_{(12)}c_{(2)}P$	2.740	8.009	20.860	52.720	146.261	1.74	5.269	12.851	31.86	93.541
2	$Pc_{(2)}c_{(7)}c_{(3)}c_{(12)}P$	2.666	6.880	20.215	53.710	144.478	1.666	4.214	13.335	33.495	90.768
3	$Pc_{(12)}c_{(3)}c_{(7)}c_{(2)}P$	2.562	6.747	19.824	51.888	143.952	1.562	4.185	13.077	32.064	92.064
4	$Pc_{(3)}c_{(7)}c_{(2)}c_{(12)}P$	2.740	8.009	20.895	52.810	142.056	1.74	5.269	12.886	31.915	89.246
5	$Pc_{(12)}c_{(2)}c_{(7)}c_{(3)}P$	2.562	6.349	16.539	48.698	139.161	1.562	3.787	10.19	32.159	90.463
						$\xi=3$					
1	$Pc_{(3)}c_{(7)}c_{(12)}c_{(2)}P$	2.229	5.590	11.769	23.518	53.351	1.229	3.361	6.179	11.749	29.833
2	$Pc_{(2)}c_{(7)}c_{(3)}c_{(12)}P$	2.166	4.540	11.457	24.460	52.698	1.166	2.374	6.917	13.003	28.238
3	$Pc_{(2)}c_{(12)}c_{(7)}c_{(3)}P$	2.166	4.208	8.838	22.369	52.532	1.166	2.042	4.63	13.531	30.163
4	$Pc_{(12)}c_{(3)}c_{(7)}c_{(2)}P$	2.023	4.261	10.750	23.003	52.180	1.023	2.238	6.489	12.253	29.177
5	$Pc_{(3)}c_{(7)}c_{(2)}c_{(12)}P$	2.229	5.590	11.904	23.790	51.248	1.229	3.361	6.314	11.886	27.458

algorithm outputs the tour route chain with the shortest path under each transportation mode. Since the maximum motive value of the tour route is determined by the quantity of the vector $\mathbf{C}_{(i)}$ tourist sights, the experiment chooses the top five tour route chains in the vector $\mathbf{X}^{\xi}_{L_{(e)}}$ output by the developed algorithm, and the top three tour route chains output by the control group algorithms. Analyze the comparison results and conclude the algorithm's advantages.

1) THE GENERATION OF THE TOURIST SIGHT NERVE CELL'S MOTIVE WEIGHTS AND ACCOMMODATION COEFFICIENTS

According to the definition of the motive weight $\omega_{(k)}$ and accommodation coefficient $\theta_{(k)}$, the Leshan city's geographic information data and traffic information data are obtained from the Baidu map and Leshan city's basic geographic information database, including the route distance from one tourist sight to another one $z_1(km)$,the spatial coordinate distance z_2, the quantity of public bus z_3, the average running time of public bus $z_4(h)$, the taxi fee z_5(¥*yuan*), the average road congestion index z_6, the quantity of traffic light z_7, the distance from tourist sight to bus station $z_8(km)$, the average waiting time for the taxi $z_9(h)$, the average quantity of congestion road z_{10}. The motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$ are shown in the Table 5. The parameter ξ will influence the $\omega_{(k)}$ and $\theta_{(k)}$ values. In the experiment, the starting point P is set as the Leshan railway station, whose coordinate is $l = 103.712$, $B = 29.602$.

2) THE RESULT OF TOUR ROUTE CHAINS BASED ON MULTIVARIATE TRANSPORTATION MODES

According to the quantity of the vector $\mathbf{C}_{(i)}$ tourist sights $C_{(i)}$, $r = 4$, there will be $\mathrm{P}(r, r) = 24$ sorts of tour route chains that meet tourists' interests. Set the initial signal information motive value as $X^{\xi}_{(0)} = 1.000$. Substitute the motive weights $\omega_{(k)}$ and the accommodation coefficients $\theta_{(k)}$ under the condition of the tourists' different transportation modes ξ into the nerve cell iteration model Formulas (10)~(12). The iteration results are shown in the Table 6.

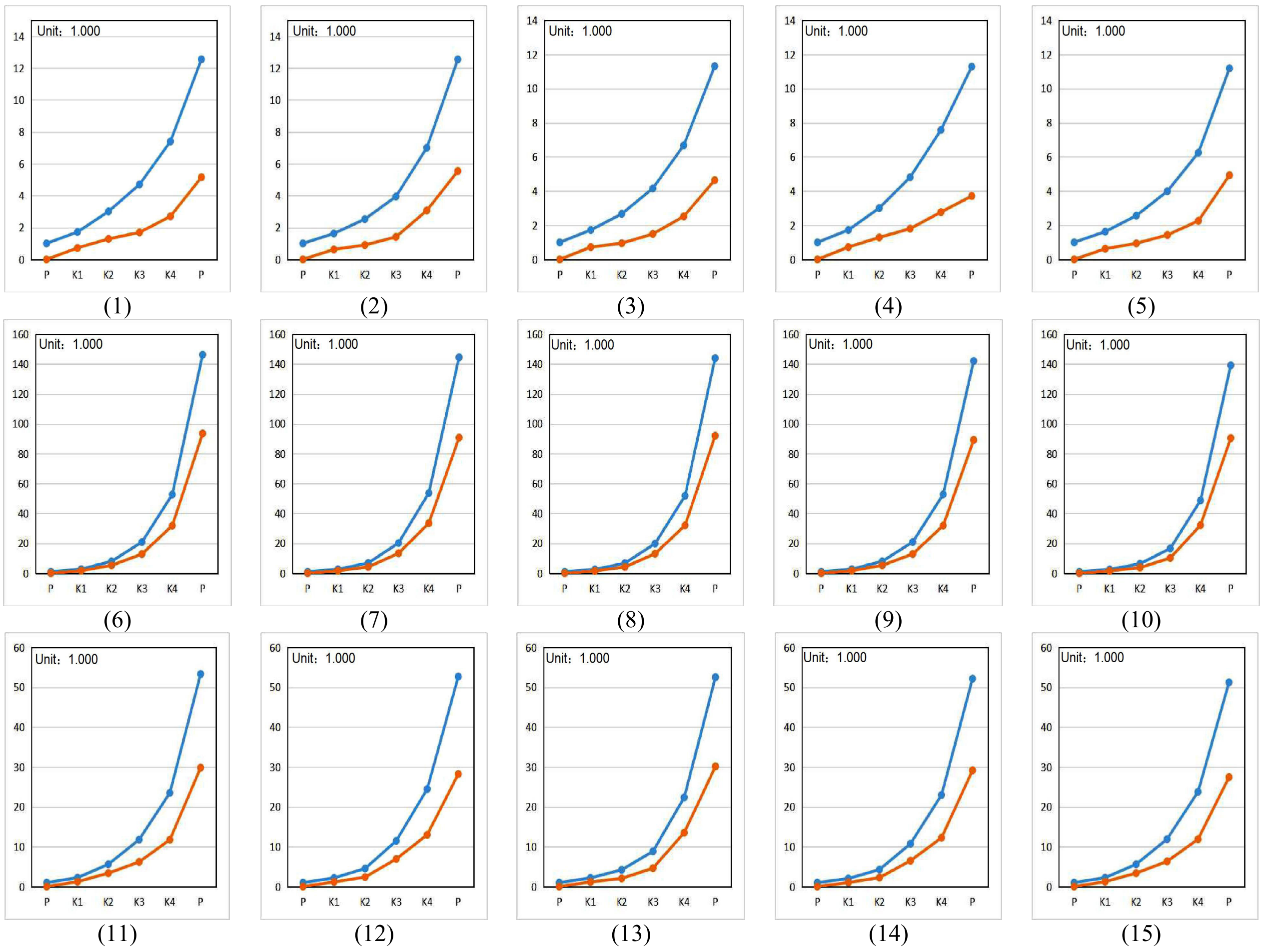


FIGURE 9. The increasing tendency of the nerve cell signal information motive values and the nerve cell signal information increased values.

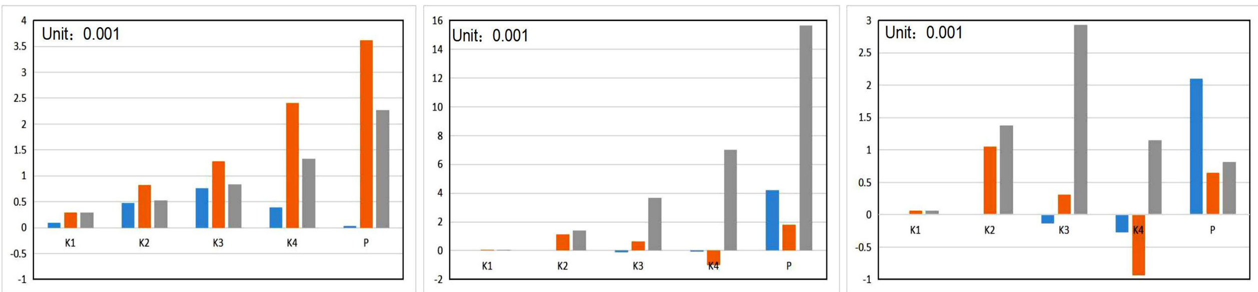
The top five tour route chains of the vector $\mathbf{X}^{\xi}_{L(e)}$ for each transportation mode are displayed in the table. The data of the column $K_{(i)}$ represent the signal information motive values $X^{\xi}_{(i)}$ of the nerve cell nodes $K_{(i)}$. The columns of $K_{(i)}K_{(i+1)}$(or $PK_{(1)}, K_{(4)}P$) represent the increased signal information motive values $\Delta X^{\xi}_{(i)}$ of the next nerve cell node to the previous one. Figure 9 shows the tendency of nerve cell signal information motive values $X^{\xi}_{(i)}$ and the increased signal information motive values $\Delta X^{\xi}_{(i)}$ of the top five tour route chains under the condition of transportation mode ξ. The Figure 9(1)$\sim$(5) represents the transportation mode $\xi = 1$, The Figure 9(6)$\sim$(10) represents the transportation mode $\xi = 2$, The Figure 9(11)$\sim$(15) represents the transportation mode $\xi = 3$. The blue curves represent the increasing tendency of the nerve cell signal information motive values $X^{\xi}_{(i)}$ for the tour route chains. The nodes relate to the starting point P, the points $K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$ and the terminal point P. The orange curves represent the tendency of the nerve cell signal information increased values $\Delta X^{\xi}_{(i)}$ for the chains' node intervals, relating to the interval between the two points.

3) THE CONTROL GROUP ALGORITHMS AND THE OUTPUT RESULTS

In tour route planning, the shortest path searching algorithms are usually used as the basic method, whose principle is searching for the smallest distance between two points with certain quantity of internal nodes. Its purpose is to optimize the travel schedule, travel time and the travel cost. The shortest path searching algorithm only considers the distance in the whole process of tourists' traveling among the tourist sights, not considering the tourist sights' geographic spatial distribution, specific coordinates, geographic information data and traffic information data, etc., and the algorithms all have different searching modes and features. When the quantity of the tourist sight is relatively small, the usually used shortest path searching algorithms include Dijkstra algorithm, Floyd-Warshall algorithm, Bellman-Ford algorithm, etc. The experiment chooses the three algorithms as the control group. According to the three algorithms' iteration method in searching shortest path, the shortest paths under the three algorithms are output, which are the optimal tour route chains for the three algorithms. Since the three algorithms

TABLE 7. Each algorithm's optimal tour route chain and the motive different values on the tourist sight nerve cell modes of the control group algorithms to the developed algorithm.

Dijkstra	Floyd-Warshall	Bellman-Ford	Proposed algorithm	ξ	$K_{(1)}$	$K_{(2)}$	$K_{(3)}$	$K_{(4)}$	$P{:}K_{(5)}$
					0.095	0.478	0.761	0.394	0.006
			$Pc_{(12)}c_{(2)}c_{(7)}c_{(3)}P$	1	0.294	0.825	1.286	2.402	3.625
					0.294	0.53	0.837	1.328	2.266
					0	0	-0.035	-0.09	4.205
$Pc_{(3)}c_{(7)}c_{(2)}c_{(12)}P$	$Pc_{(2)}c_{(7)}c_{(3)}c_{(12)}P$	$Pc_{(2)}c_{(12)}c_{(7)}c_{(3)}P$	$Pc_{(3)}c_{(7)}c_{(12)}c_{(2)}P$	2	0.074	1.129	0.645	-0.99	1.783
					0.074	1.397	3.653	7.014	15.656
					0	0	-0.135	-0.272	2.103
			$Pc_{(3)}c_{(7)}c_{(12)}c_{(2)}P$	3	0.063	1.05	0.311	-0.942	0.651
					0.063	1.382	2.931	1.148	0.819



(1) $\xi = 1$ (2) $\xi = 2$ (3) $\xi = 3$

FIGURE 10. The motive different values on the tourist sight nerve cell nodes of the control group algorithms to the developed algorithm.

do not consider the factor of the transportation mode, as to the different transportation modes, the shortest paths are the same one, and they relate to the certain tour route chains with the signal information motive values generated by the developed algorithm, and have the difference values as to the optimal tour routes of the developed algorithm. In order to ensure the comparison between the proposed method and the comparative methods is fair, the research sets the identical experimental environment for the algorithms. First, the proposed algorithm and the control group algorithms all have the same starting point and the terminal point, also, all the experimental tourist sights are the same, which are used in all of the algorithms. Second, the real world experimental environment of the algorithms are identical, that is, each algorithm combines with the same geographic information data, traffic information data and the tourist sight data. Third, the modeling, calculating and running environment of the algorithms are all identical, which ensures the fairness on the aspect of space complexity and time complexity. In the left part of the Table 7, the optimal tour route chains of the control group algorithms and the developed algorithms are displayed. In the right part of the Table 7, the motive different values on the tourist sight nerve cell nodes of the control group algorithms to the developed algorithm are displayed, the condition is the transportation mode ξ. Figure 10 shows the distribution of the motive different values on the tourist sight nerve cell nodes of the control group algorithms to the developed algorithm, in which the Figure 10(1)~(3) represent the transportation modes of $\xi = 1, \xi = 2$ and $\xi = 3$.

The blue data lines represent the signal information difference values $X_{(i)}^{\xi}$ of the developed algorithm to the Dijkstra algorithm on the nodes $P, K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$. The orange data lines represent the signal information difference values $X_{(i)}^{\xi}$ of the developed algorithm to the Floyd-Warshall algorithm on the nodes $P, K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$. The gray data lines represent the signal information difference values $X_{(i)}^{\xi}$ of the developed algorithm to the Bellman-Ford algorithm on the nodes $P, K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$. By comparing the difference values of each algorithm on the nodes, the algorithms' capacities on generating motive values could be compared and analyzed. Since the modeling method and mechanism are different among the proposed algorithm and the control group algorithms, the space complexity and time complexity in the running process will be different, too. When the starting point and the nodes on the tour route are identical, all the algorithms' running efficiency on outputting the signal information motive values are different. According to the experimental conditions, the starting point is the point P, the nodes are $K_{(1)}, K_{(2)}, K_{(3)}$ and $K_{(4)}$, and the terminal point is also the point P. Thus, the quantity of the node on one tour route chain is $n = 6$. The space complexity and time complexity for all the algorithms are shown in the Table 8.

TABLE 8. The comparison on the time complexity and space complexity among the proposed algorithm and the control group algorithms.

	Time complexity	Time complexity $n=6$	Space complexity	Space complexity $n=6$
The proposed algorithm	$O(2n\log_2 n)$	$O(31.02)$	$O(1)$	$O(1)$
Dijkstra algorithm	$O(n\log_2 n+n^2)$	$O(51.51)$	$O(n)$	$O(6)$
Floyd-Warshall algorithm	$O(n^3)$	$O(216)$	$O(n^2)$	$O(36)$
Bellman-Ford algorithm	$O(15n)$	$O(90)$	$O(n)$	$O(6)$

4) THE ANALYSIS OF THE EXPERIMENTAL RESULTS

The statistical method is used to make statistics and analyze the experimental results and compare each tour route's performance under different transportation modes, and finally judge about the significance of the method. Firstly, the statistical characteristics of the experimental results are discussed and analyzed.

(1) The experimental data results are not random. In the research, the nerve cell signal information motive values $X_{(i)}^{\xi}$ and the increased values $\Delta X_{(i)}^{\xi}$ are calculated by the proposed algorithm, which simulates the signal processing and transmitting mode of the nerve cells and combines with the motive weights $\omega_{(k)}$ and coefficients $\theta_{(k)}$. The data results are not the randomly generated observation data, thus, they do not have the random characteristics.

(2) The experimental data results are not independent with each other. According to the algorithm design, on the same tour route chain, the previous one nerve cell's motive value is the precondition for the next one nerve cell. Thus, the adjacent two tourist sights have the direct close dependence relationship. And all the tourist sight nodes have the dependence relationship. According to the analysis, all the experimental data results are not the independent data.

(3) The experimental data results do not obey normal distribution. The signal information motive values on the tour route chains have the monotone increasing characteristics. Thus, the data generated by algorithms do not obey normal distribution.

(4) The evaluation analysis on the experimental data results. The evaluation on the experimental data does not preset certain criterion, that is, there is no certain tour route that is selected as the reference to evaluate other tour routes. The signal information motive difference values generated by each tour route are aroused by the differences on tourists' interests, tourist sights, motive weights $\omega_{(k)}$ and coefficients $\theta_{(k)}$, tour sequence, etc.

According to the analysis of the experimental data results on the statistical characteristics, the commonly used statistical test methods such as T test, F test, chi-square test, and Z test are not suitable for the experimental data results. According to the characteristics of the experimental data results, the single point analysis could be done to test the performance of the algorithms and the results. In order to test the stability and discreteness of each tour route under the different transportation modes on outputting the signal information motive values, in the experiment, the variance analysis and standard deviation analysis are used to compare and analyze the variance values and the standard deviation values on each nodes of the tour routes. By calculating and analyzing the values, the quantitative method could be used to test the stability of the algorithm when generating the signal information motive values. The statistical data are the calculated and output data in the Table 6. The equation of variance $S^2 = \sum(X - \overline{X})/(n-1)$ is used to calculate the variance values, and the standard deviation equation $S = \sqrt{S^2}$ is used to calculate the standard deviation values, and get the Table 9. In the Table 9, the left side represents the statistical variance values, the average variance values, the statistical standard deviation values and the average standard deviation values of each node's signal information motive values $X_{(i)}^{\xi}$ in the tour routes. For example, the first row data of the $X_{(i)}^{\xi}$ are the statistical variance values of the signal information motive values on the nodes of $K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$ and $P : K_{(5)}$ for all the tour routes under the transportation mode of $\xi = 1$, and the second row data are the related average variance values, the third row data are the statistical standard deviation values, and the fourth row data are the average standard deviation values. Other modes of $\xi = 2$ and $\xi = 3$ are the same. The right side represents the statistical variance values, the average variance values, the statistical standard deviation values and the average standard deviation values of each node's signal information increased motive values $\Delta X_{(i)}^{\xi}$ in the tour routes. For example, the first row data of the $\Delta X_{(i)}^{\xi}$ are the statistical variance values of the signal information increased motive values on the nodes of $K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$ and $P : K_{(5)}$ for all the tour routes under the transportation mode of $\xi = 1$, and the second row data are the related average variance values, the third row data are the statistical standard deviation values, and the fourth row data are the average standard deviation values. Other modes of $\xi = 2$ and $\xi = 3$ are the same. The Figure 11 shows the distribution of the statistical variance values, the average variance values, the statistical standard deviation values, the average standard deviation values of the signal information motive values and increased motive values on each node of the tour routes under the different transportation modes. In Figure 11, the Figure 11(1)∼(3) are the distributions of the statistical variance values and the average variance values of the signal

TABLE 9. The statistical variance values, average variance values, statistical standard deviation values and average standard deviation values of each node in the tour routes under different transportation modes.

	$X^{\xi}_{(i)}$					$\Delta X^{\xi}_{(i)}$				
	$K_{(1)}$	$K_{(2)}$	$K_{(3)}$	$K_{(4)}$	$P{:}K_{(5)}$	$PK_{(1)}$	$K_{(1)}K_{(2)}$	$K_{(2)}K_{(3)}$	$K_{(3)}K_{(4)}$	$K_{(4)}P$
$\xi=1$										
$S^2_{(1)}$	0.003	0.056	0.165	0.289	0.499	0.003	0.039	0.030	0.090	0.475
$\overline{S^2}_{(1)}$	0.202	0.202	0.202	0.202	0.202	0.127	0.127	0.127	0.127	0.127
$S_{(1)}$	0.052	0.236	0.407	0.537	0.707	0.052	0.196	0.173	0.301	0.689
$S_{(1)}$	0.388	0.388	0.388	0.388	0.388	0.282	0.282	0.282	0.282	0.282
$\xi=2$										
$S^2_{(2)}$	0.008	0.585	3.260	3.752	7.297	0.008	0.466	1.658	0.461	2.693
$S^2_{(2)}$	2.981	2.981	2.981	2.981	2.981	1.057	1.057	1.057	1.057	1.057
$S_{(2)}$	0.089	0.765	1.806	1.937	2.701	0.089	0.682	1.288	0.679	1.641
$\overline{S}_{(2)}$	1.460	1.460	1.460	1.460	1.460	0.876	0.876	0.876	0.876	0.876
$\xi=3$										
$S^2_{(3)}$	0.007	0.487	1.585	0.627	0.597	0.007	0.406	0.758	0.579	1.258
$S^2_{(3)}$	0.660	0.660	0.660	0.660	0.660	0.602	0.602	0.602	0.602	0.602
$S_{(3)}$	0.084	0.698	1.259	0.792	0.772	0.084	0.637	0.871	0.761	1.122
$\overline{S}_{(3)}$	0.721	0.721	0.721	0.721	0.721	0.695	0.695	0.695	0.695	0.695

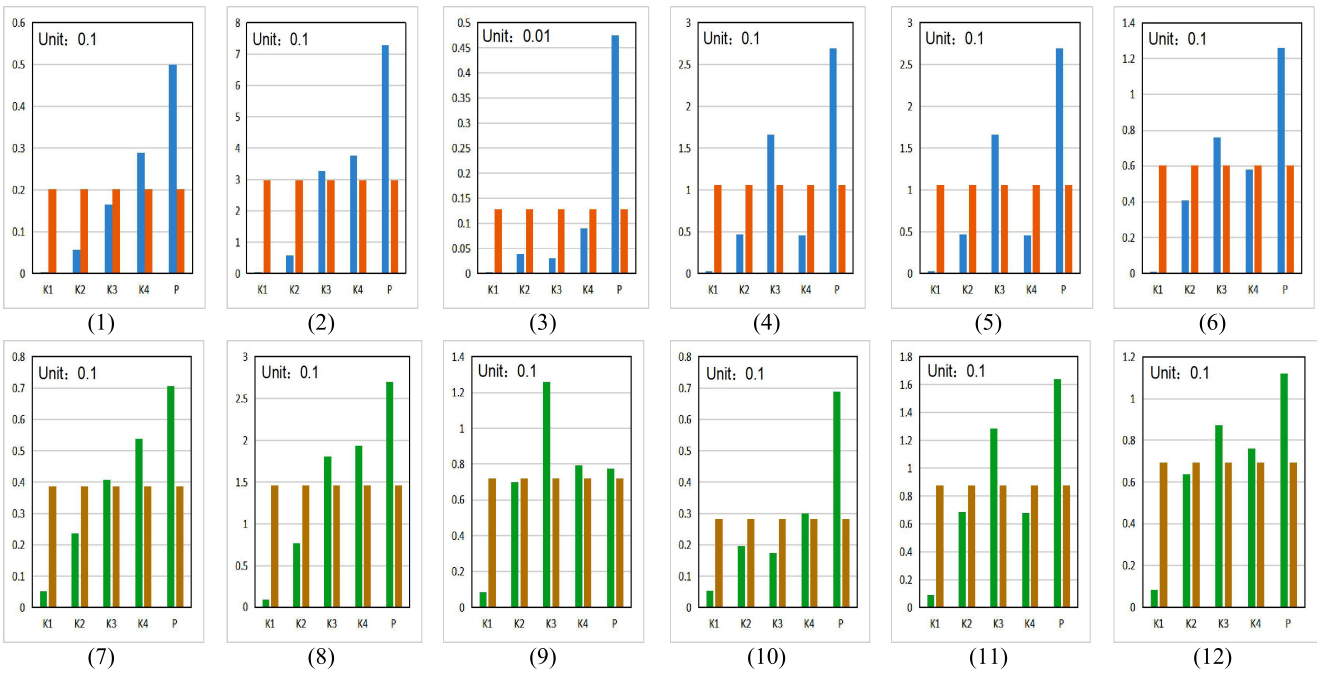


FIGURE 11. The distribution of the statistical variance values, the average variance values, the statistical standard deviation values, the average standard deviation values of the signal information motive values and increased motive values on each node of the tour routes under the different transportation modes.

information motive values under the transportation modes of $\xi = 1, \xi = 2$ and $\xi = 3$. The Figure 11(4)~(6) are the distributions of the statistical variance values and the average variance values of the signal information increased motive values under the transportation modes of $\xi = 1, \xi = 2$ and $\xi = 3$. The Figure 11(7)~(9) are the distributions of the statistical standard deviation values and the average standard deviation values of the signal information motive values under the transportation modes of $\xi = 1, \xi = 2$ and $\xi = 3$. The Figure 11(10)~(12) are the distributions of the statistical standard deviation values and the average standard deviation values of the signal information increased motive values under the transportation modes of $\xi = 1, \xi = 2$ and $\xi = 3$.

V. THE EXPERIMENTAL RESULTS ANALYSIS

According to the experiment and the output results, analyze and discuss the results on the experiment basic data, the generated clusters, the mined tourist sights, the output tour route chains and the comparison with the control group algorithms.

A. THE EXPERIMENT BASIC DATA ANALYSIS

Analyze the collected tourist sights' basic data, they have the following features. First, the classification and quantity of the tourist sights are abundant. All the tourist sights are the typical ones in the Leshan city, which have relatively high popularity and the best tourist source evaluation. Second, the geographic distributions of the tourist sights re relatively discrete. Third, arbitrary two tourist sights are connected by city roads and avenues, they are accessible in different transportation modes. Fourth, all the tourist sights have self feature attributes. Analyze the Table 2 data, arbitrary tourist sight has the quantitative feature attribute on the longitude and latitude, feature attribute matching factor $k_{(1)}$, the optimal travel time $k_{(2)}$, the basic travel cost $k_{(3)}$ and attraction index $k_{(4)}$, in which the factor $k_{(1)}$ is obtained by tourist sight knowledge text mining. It is the key factor to match the tourists' interests and it reflects the correlation degree on the aspect of the text definition and function. As to the tourist sight clustering algorithm and text mining algorithm, the feature attributes are influenced by the disturbance factors and are normalized on the impact to tourists' interests, and this is the basis for the algorithm.

B. THE TOURIST SIGHT CLUSTERING RESULT ANALYSIS

Analyze the Table 3 data, the correlation of arbitrary two tourist sights is mainly influenced by the keyword label. The higher the correlation degree is, the larger the label vector matching value δ will be, the smaller the feature attribute cluster correlation degree $D_{(c_{(i)},c_{(j)})^{-1}}$ will be. When the value of two tourist sights is smaller than the threshold value $T_{(c_{(i)},c_{(j)})} = 0.500$, the two tourist sights are the homogeneous cluster ones. In the process of generating clusters, the tourist sights $c_{(1)} \sim \Delta c_{(1)}$: Leshan giant Buddha, $c_{(2)} \sim \Delta c_{(2)}$: Wan Da mall, $c_{(3)} \sim \Delta c_{(3)}$: the cultural center, $c_{(5)} \sim \Delta c_{(5)}$:the children amusement park are chosen as the seed points, from which the subordinate degree matrix $\xi_{(p\times \max n(i))}$ is generated. Analyze the four subordinate degree matrix of Formula (14), the black bold value 1 represents each cluster's seed point, and the other elements are not the seed points. Of all the other non seed points, the correlation degree $D_{(c_{(i)},c_{(j)})^{-1}}$ of the tourist sights with the value 1 is smaller than or equal to the threshold value $T_{(c_{(i)},c_{(j)})}$, they are the homogeneous cluster tourist sights. The correlation degree $D_{(c_{(i)},c_{(j)})^{-1}}$ of the tourist sights with the value 0 is larger than the threshold value $T_{(c_{(i)},c_{(j)})}$, they are the heterogeneous cluster tourist sights. From the element values of the matrix $\xi(p \times \max n(i))$, the homogeneous cluster and the heterogeneous cluster tourist sights are randomly distributed, which is determined by the tourist sights' feature attributes, the label vector, the text mining, etc. And from the subordinate degree matrix, the clusters $C_{(1)}$, $C_{(2)}$, $C_{(3)}$ and $C_{(4)}$. From the distribution of the clusters, different clusters are discrete, and they are also interlaced. The tourists ferry between two tourist sights, it may also be the ferry between two clusters.

C. THE TOURIST SIGHT MINING RESULT ANALYSIS

The confirmed interest matrix $\mathbf{W}_{(4\times4)}$ is shown as Formula (15). Analyze the basic interest matrix, the arbitrary one row $\forall \mathbf{W}_{u(1\times4)}$ of the matrix represents the expected label on the related cluster's tourist sights. The feature attribute matching factors $k_{(1)}$ are confirmed by text mining, and it is the result of interest matrix row's label elements matching the tourist sights' text data. Combining with tourists' specific needs on the factors $k_{(2)}$, $k_{(3)}$ and $k_{(4)}$, the objective function $G_{(\mathbf{K}(r),c(i))}$ values are obtained. The values $k_{(1)}$ and $G_{(\mathbf{K}(r),c(i))}$ are shown in the Table 4, data distribution is shown in Figure 8.

Analyze the Table 4 data and the Figure 8, the matching values $k_{(1)}$ and $G_{(\mathbf{K}(r),c(i))}$ of the matrix $\mathbf{W}_{(4\times4)}$ row's interest label vector $\mathbf{W}_{u(1\times4)}$ to the related cluster $\mathbf{C}_{(u)}$ tourist sights are all different. In the descending order of the $G_{(\mathbf{K}(r),c(i))}$ values, store the cluster tourist sights into the vector $\mathbf{V}_{(uv)}$, and the interest tourist sights are extracted from the vector.

From the Figure 8, the $k_{(1)}$ and $G_{(\mathbf{K}(r),c(i))}$ values have different distributions at each cluster's tourist sights. The higher the data bar is, the higher the matching degree value of the tourist sight to the interest label vector $\mathbf{W}_{u(1\times4)}$ will be. The Figure 8 shows which tourist sights mostly match the tourists' interests.

D. THE OUTPUT RESULT OF THE TOUR ROUTE CHAINS

Analyze the Table 5 data, the tourist sight nerve cell motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$ among different tourist sights calculated by the Leshan city's geographic information data and traffic information data are discrepant in some extent, and the output signal information motive values at the nerve cells iterated by the factors $\omega_{(k)}$ and $\theta_{(k)}$ are also discrepant. This is determined by the discrepancy of the tourist sights' distribution, geographic information data and traffic information data. The factors $\omega_{(k)}$ and $\theta_{(k)}$ are adjusted by the disturbance coefficients to ensure that they are in the same order on the influence of the signal information motive value. In the same tour route chain, the signal information will not be influenced by the two tourists' motive weights $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$, they have the same impact in the forward direction and the reverse direction. Analyze the Table 6 data, the output top five tour route chains are different under the condition of multivariate transportation modes, and also, in each tour route chain, the output signal information motive values at the nerve cells are different, and this is determined by the cluster of the tourist sight, the travel sequence, weight values $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$. For each transportation mode, the first ranked tour route has the highest signal information motive value, which could provide tourists with the best satisfaction. The other tour route chains are the suboptimal ones. Analyze the variation tendency of the tour route chains' nerve cells signal information motive values and the variation tendency of the increased signal information motive values in the Figure 9, the signal information motive value of each tour route chain increases with the travel time and tourists' traveling distance along the tourist sights' sequence. At the terminal point P, the signal information motive value gets to the maximum one, and it the whole chain's maximum motive value. Meanwhile, the increased signal information motive value of each interval between two nerve cells also increase with the travel time and tourists' traveling distance along the tourist sights' sequence, and it also gets to the maximum value at the terminal point P. This phenomenon illustrates that the signal information motive value of each interval between two nerve cells is monotone increasing. Of all the tour route chains under the same transportation mode, the higher the signal information motive value of the tour route chain is, the increasing curve slope will be larger, the increasing speed will be higher, too.

Compare the results of the control group algorithms and the developed algorithm on the optimal tour route chains in the Table 7 and Figure 9. The searched shortest paths of the control group algorithms are the globally optimal solutions, which are different from the results of the developed algorithm. Table 7 shows the signal information motive value differences at the nerve cells and the whole chains' signal information motive difference values of the output optimal tour route chains of the control group algorithms and the developed algorithm under the condition of different transportation modes. Analyze the data, the signal information motive values of the control group algorithms and the developed algorithm have positive or negative difference, that is, the difference values at the different nerve cells are fluctuant, this is determined by the input initial signal information value and the impacted weight values $\omega_{(k)}$ and accommodation coefficients $\theta_{(k)}$. Its volatility reflects the algorithms' discrepancies and characteristics on the iterated signal information motive values when they output the optimal tour routes. Figure 10 shows the signal information motive value differences distribution of the output tour route chain nerve cells of the control group algorithms and the developed algorithm under the condition of different transportation modes. Analyze the distribution figures, the output shortest paths of the control group algorithms and the developed algorithm have difference on the signal information motive value at each nerve cell and each interval of two nerve cells.

(1) When $\xi = 1$, the difference values on the optimal tour route chains of the developed algorithm to the Dijkstra algorithm firstly increase and then decrease in the iteration process, and the difference values on the optimal tour route chains of the developed algorithm to the Floyd-Warshall algorithm and the Bellman-Ford algorithm gradually increase in the iteration process.

(2) When $\xi = 2$, the difference values on the optimal tour route chains of the developed algorithm to the Bellman-Ford algorithm gradually increase in the iteration process, the difference values on the optimal tour route chains of the developed algorithm to the Floyd-Warshall algorithm firstly decrease and then increase in the iteration process, and the difference value of the developed algorithm to the Dijkstra algorithm gets to the maximum value only at the terminal point while little difference at other nerve cells.

(3) When $\xi = 3$, the difference values on the optimal tour route chains of the developed algorithm to the Bellman-Ford algorithm firstly increase and then decrease in the iteration process, and the difference values on the optimal tour route chains of the developed algorithm to the Floyd-Warshall algorithm and the Dijkstra algorithm firstly decrease and the increase in the iteration process.

By comparing the differences between the algorithms, under the three transportation modes, the output signal information motive values of the proposed algorithm on most of the nodes are larger than the same nodes of the control group algorithms, while few amount of nodes are smaller. But in all, the maximum signal information motive values of the tour route chains output by the control group algorithms are smaller than the developed algorithm, which illustrates that the developed algorithm has advantages on its algorithm design principle and performance to satisfy tourists' interests. Compare the algorithms' time complexity and space complexity. The proposed algorithm's time complexity and space complexity are both lower than the control group algorithms, thus, the proposed algorithm runs fastest and takes up the lowest memory storage space in the computer, followed by the Dijkstra algorithm, the Bellman-Ford algorithm and the Floyd-Warshall algorithm. Thus, it has advantages on the aspect of algorithm operating.

Take the nerve cell nodes as the control points and analyze the Table 9 and the Figure 11. Under the different transportation modes, the statistical variance values and standard deviation values of the signal information motive values on each node in the tour routes have the maximum value and the minimum value, and the values all fluctuate up and down above or under the average line. The fluctuate range is small. The statistical variance value and the standard deviation value of the same node are all small. Of all the statistical variance values and standard deviation values, there is no exceptional value, which illustrates that the performance of the proposed algorithm is stable on outputting the signal information motive values. Meanwhile, the statistical variance values and standard deviation values of the signal information increased motive values on each node in the tour routes have the maximum value and the minimum value, and the values all fluctuate up and down above or under the average line. The fluctuate range is small. The statistical variance value and the standard deviation value of the same node are all small. Of all the statistical variance values and standard deviation values, there is no exceptional value, which illustrates that the performance of the proposed algorithm is stable on iterating signal information motive values between two tourist sights and in the whole tour route. By analyzing the statistical variance values and standard deviation values, it testifies that the proposed algorithm is stable. When the tour sequence, motive weights $\omega_{(k)}$ and coefficients $\theta_{(k)}$, transportation modes change, the performance of the proposed algorithm is good.

E. THE PRACTICALNESS OF THE ALGORITHM AND THE EXPERIMENT

The proposed algorithm is set up on the real world tourism environment and the practical data, and it is mainly used to develop the intelligent recommendation system. The tourist interest data mining, the precise tourist sight mining and the optimal tour route chain algorithm in the proposed algorithm are all set up on the practical experimental data, and the output results are all practical and feasible in the real world environment. Thus, the proposed algorithm could be used as the embedded algorithm for the intelligent recommendation system, and the output results could be directly provided for tourists for reference. The experiment is not a simulation one, but a real world environmental experiment. The basic tourist sight data, geographic information data and traffic information data are all collected from the practical experimental data of the tourism city Leshan, also, the evaluation factors and data are based on the true geographic data and tourism data. Thus, the recommendation results could be used directly, but not the computer simulation results. The control group algorithms that generate the tour routes are also based on the same real world environment, geographic data and tourism data. The proposed algorithm and the control group algorithms all have the same starting point and nodes on the tour routes, and they have the same running and operating environment, that is, the geographic information data, traffic information data and the tourist sight data are identical for the algorithms.

Compare the signal information motive values on the nodes $K_{(1)}, K_{(2)}, K_{(3)}, K_{(4)}$ and P, and the data results come from the real world environment, which reflects the satisfaction difference on the tour routes that the tourists will obtain under the circumstance of the Leshan basic tourism data. On this aspect, the proposed algorithm has advantages than the control group algorithms, it is practical and feasible.

F. THE ANALYSIS OF THE ALGORITHM MEETING TOURISTS' INTERESTS

The aim of the proposed algorithm is to output the optimal tourist sights and tour routes to match tourists' interests.

First, analyze the results of tourist sights matching tourists' interests. The algorithm uses interest labels to quantify the tourists' interests and the feature attributes that the tourist sights could provide. Through setting up the objective function $G_{(\mathbf{K}(r),c(i))}$, the approach values between the tourists' interest labels and the tourist sight feature attributes are calculated, of which the tourist sights with the best approached values are the matched tourist sights. In the experimental results, the function values of the Table 4 reflect the approach extent between tourists' interests and the tourist sights' feature attributes. The best matching tourist sights could meet all the requirements of the tourists. On this aspect, the confirmed four tourist sights in the experiment could meet all the tourists' requirements, that is, each tourist sight's visiting time $k_{(2^*)} \leq 1.5$ (Unit: *hour*), the cost $k_{(3^*)} = 0$ (Unit:¥*yuan*), tourist sight attraction index $k_{(4^*)} \geq 0.700$. The tourist expects to visit two cluster $C_{(1)}$ tourist sights, one cluster $C_{(2)}$ tourist sight and one cluster $C_{(3)}$ tourist sight, and should be the optimal one to match the interests.

Second, on the basis of the matched tourist sights, the tour routes are output under the three transportation modes via the proposed algorithm. The algorithm combines with the real world geographic information data and traffic information data, it is close to the actual tourism activities and reflects tourists' traveling process. Since each of the tourist sight in the tour routes matches tourists' interests, the tour routes all could meet the interests, too. Of all the tour routes, the algorithm outputs the optimal ones and provides for the tourists. The experiment testifies that the proposed algorithm can meet the needs of tourists, the recommended tourist sights and tour routes are all the best ones for the traveling activities.

Third, the proposed algorithm's modeling thought, modeling method, parameter selecting and algorithm processing are not based on a certain city. The parameters used in the algorithm modeling are the geographic information data and the traffic information data that could be provided by any tourism city. When the research range changes, the only operation for the algorithm is to alter the basic data of the studied tourism city. Thus, the proposed algorithm is a universal algorithm, which has the universality and portability. This is the significant purpose of the proposed algorithm, that is, to make the algorithm as the embedded algorithm for the intelligent recommendation system to support the software or Internet platform. The system demand data are all stored in the database, which can be used to recommend tourist sight and tour routes under the condition of different cities' tourism environment.

VI. CONCLUSION AND THE FUTURE WORK

Conclude the research thought and the key research content, the developed tour route chain recommendation algorithm based on interest mining and multivariate transportation modes mainly researches on the tourist sights that match tourists' interests on the aspect of interest mining and the condition of multivariate transportation modes, which outputs the tour route chains in the city's transportation modes. Of these two aspects, one is the precondition to generate the travel motive, and the other one is the indispensable condition for traveling. The tourists' interests determine the optimal tourist sights provided by the recommendation system before they visit the tourism city, that is, the matching degree with the tourist sights' feature attributes, and then influence the specific tourist sights in the nerve cells, finally influence the tour route chains' capacity on satisfying tourists' interests. Another factor, the traffic transportation modes directly influence the output results of the signal information motive values. When the tourists choose different transportation modes, and other factors are identical, the same tour route chain will output different signal information motive values. The tour route chain is the optimal one under one certain transportation mode, but may not be the optimal one for another transportation mode. This is the emphasis and key point of the research.

A. ANALYSIS OF PROBLEM SOLVING PROCESS AND RESULTS

The research process has the characteristics of waterfall model, the research content and result of the previous step is the precondition for the next research content. The process from the algorithm design to the results generation mainly includes three steps.

The first step: the city tourist sight domain is set up. The tourist sight cluster algorithm is set up to obtain the urban tourist sight clusters, and this is the precondition for mining the tourist sights to match tourists' interests. In the cluster algorithm, the text mining method is used to calculate the similarity of the tourist sights' feature attributes, and the tourist sights with the similar feature attributes are in the same cluster. On the basis of clusters, by matching the tourists' interests, the optimal and the best matched tourist sights are obtained.

The second step: simulate the nerve cell mode in receiving and transmitting the signal information, and design the tourist sight nerve cell model and the tour route chain model. The tourist sight nerve cell motive weight $\omega_{(k)}$ and coefficient $\theta_{(k)}$ are set as the factors that influence the signal information motive value, and the signal information transmitting mode simulates the nerve cell chain. Set up the tour route chain model based on the tourist sight nodes, through the model, the signal information motive values of the tour route chains under the different transportation modes could be calculated, and the optimal tour routes could be obtained.

The third step: carry out the experiment and analyze the results. The basic data are collected to perform the experiment. The proposed algorithm is used to output the precise tourist sights and the optimal tour routes. And then the comparison experiment is carried out, under the same experimental conditions, the control group algorithms also output the tour routes, and the results are compared with the proposed algorithm's results. The comparison experiment testifies that the proposed algorithm has advantages in generating the optimal tour routes than the control group algorithms.

In conclusion, the mined tourist sights by the objective function of the proposed algorithm could best match the tourists' interests and combine with the real world tourism data such as the geographic information data and the traffic information data, and the output the optimal tour routes. Compared with the control group algorithms, the proposed algorithm has higher motive benefits satisfaction, lower time complexity and space complexity.

B. ACADEMIC IMPLICATIONS

The research has some academic implications. Firstly, this research brings forward new thought and method for the development of the intelligent tourism recommendation system. The study of the intelligent recommendation system should not be limited on the collaborative filtering algorithm, content recommendation algorithm, knowledge recommendation algorithm, association rules algorithm, user characteristics recommendation algorithm. Thus, the proposed method in the paper provides new thought and method for the recommendation system research. The recommendation method in the research is based on the tourists' interests, combining with data mining algorithm, so it is more accurate and precise. The recommendation process takes the matched tourist sights as the nodes, which ensures that each node tourist sight could match tourists' interests. The geographic information data and the traffic information data are brought in the algorithm, also, the mode of nerve cell transmitting signal information is simulated to output the optimal tour routes. Therefore, the research method in the work is a comprehensive method.

Secondly, the proposed algorithm has relatively strong universality, feasibility and portability. The proposed algorithm in the paper has strong universality. The tourism city Leshan is taken as the research range and object. The basic data come from the Leshan geographic information and traffic information. When the basic data is replaced by the data of other tourism cities, the algorithm and the method are also usable and applicable. Thus, this proposed algorithm could be used as the embedded algorithm for the intelligent tourism recommendation system, and has strong portability.

Thirdly, the research has the characters of broadening the research thoughts and expanding the system functions. In the process of recommending tourist sights and tour routes, this algorithm combines with the tourists' interests mining, the tourist sights and tour routes searching, the transportation mode selecting, the geographic information data and traffic information data collecting, etc. It makes the research work have characters of broadening the research thoughts on tourism recommendation. Aim at the research points of interest data mining, algorithm optimizing and multiple influence factors, more research work could be done in future. Moreover, the algorithm refers to large amount of intermediate data, thus, these data could be used to do deep mining and obtain relative knowledge, which could be used by the tourism administration department to manage and mine the interest data, research on the hot popular tourist sights and tour routes, predict the tourist traveling volume, and it is also the reference for the government to optimize the public transportation system, launch the public vehicle and optimize the tourism transportation guarantee.

C. LIMITATION OF THE PAPER AND FUTURE WORK

The research in the paper focuses on tourists' interests mining, optimal tourist sights and tour routes searching under the multivariate transportation modes. The research range is the downtown area. The research range and object is the downtown area and its tourist sights. The tourist sights in the outskirts and the subordinate counties are not in the research range. Thus, the proposed method could not be used to plan the tour routes which include the outskirts tourist sights. Meanwhile, the transportation mode between the downtown area and the outskirts tourist sights are not considered as the transportation modes of the research only refer to the inner downtown area. The accessibility of the outskirts tourist sights is not studied. Also, the matching degree between the outskirts tourist sights and the tourists' interests are not studied, therefore, the proposed method could neither meet the needs of the recommendation system to match tourists' interests with the outskirts tourist sights, nor plan the tour routes containing the outskirts tourist sights.

Aim at the limitation of the research work, our research team will do more further work on this issue in future, including two aspects. Firstly, expand the research range and absorb the outskirts and counties tourist sights into the research range, meanwhile, the transportation modes between the downtown area and the outskirts should also be studied. The accessibility of the outskirts tourist sights would also be studied. Secondly, the relationship model between the outskirts tourist sights and the tourists' interests would be studied. The transportation modes in the outskirt counties are different from the transportation modes in the downtown area, thus, they should be further studied. Thirdly, the further studies could be done on the aspects of interest data mining, algorithm optimizing and multiple influence factors.

ABBREVIATIONS AND NOTATIONS

MP	McCulloch-Pitts
GIS	Geographic Information System
GPS	Global Positioning System
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory
SQL	Structured Query Language
C	Tourist sight domain
$\mathbf{C}_{(u)}$	Tourist sight cluster
$c_{(u,v)}$	Tourist sight meta data
$\mathbf{c}_{(i)}$	Tourist sight feature attribute label vector
$D(c_{(i)}, c_{(j)})$	The clustering criterion
$\Delta c_{(i)}$	Cluster generation point
$\xi_{(p\times \max n(i))}$	Subordinate degree matrix
$\mathbf{C}_{(p\times \max n(i))}$	Tourist sight storage matrix
$\mathbf{W}_{(t)}$	Interest feature data vector
$k_{(r)}$	Feature attribute matching factor
$\mathbf{W}_{(p\times \max p(u))}$	Tourist basic interest matrix
$\mathbf{W}_{u(1\times \max p(u))}$	Tourist basic interest label vector
$\mathbf{W}^{\wedge}_{(1\times m)}$	Interest label homogeneous vector
$\mathbf{F}_{(\mathbf{W}(p\times \max p(u)))}$	Word frequency matrix
$\mathbf{F}_{(c_{(i)}(1\times s))}$	Feature attribute word frequency vector
$G(\mathbf{K}_{(r),c_{(i)}})$	Interest tourist sight mining objective function
$\omega_{(k)}$	Tourist sight nerve cell motive weight
θ	accommodation coefficients
$\varepsilon_{(k)}$	normalization parameter for $\omega_{(k)}$
$\sigma_{(k)}$	normalization parameter for θ
$X_{(i)}$	Tourist sight nerve cell motive value

REFERENCES

[1] D. Gavalas, V. Kasapakis, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "A survey on mobile tourism recommender systems," in *Proc. 3rd Int. Conf. Commun. Inf. Technol. (ICCIT)*, Jun. 2013, pp. 131–135.

[2] J. A. Ivars-Baidal and J. F. V. Rebollo, "Tourism planning in Spain. From traditional paradigms to new approaches: Smart tourism planning," *Boletín de la Asociación de Geógrafos Españoles*, vol. 82, pp. 1–31, Aug. 2019.

[3] L. Yuan, Z. Yu, W. Luo, J. Zhang, and Y. Hu, "Clifford algebra method for network expression, computation, and algorithm construction," *Math. Methods Appl. Sci.*, vol. 37, no. 10, pp. 1428–1435, Jul. 2014.

[4] N. Daina, A. Sivakumar, and J. W. Polak, "Electric vehicle charging choices: Modelling and implications for smart charging services," *Transp. Res. C, Emerg. Technol.*, vol. 81, pp. 36–56, Aug. 2017.

[5] A. Chehreghan and R. A. Abbaspour, "A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm," *Cartography Geographic Inf. Sci.*, vol. 45, no. 3, pp. 255–269, May 2018.

[6] S. Kang, G. Lee, J. Kim, and D. Park, "Identifying the spatial structure of the tourist attraction system in South Korea using GIS and network analysis: An application of anchor-point theory," *J. Destination Marketing Manage.*, vol. 9, pp. 358–370, Sep. 2018.

[7] G.-P. Gwon, W.-S. Hur, S.-W. Kim, and S.-W. Seo, "Generation of a precise and efficient lane-level road map for intelligent vehicle systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4517–4533, Jun. 2017.

[8] J. Kim, B. Thapa, and S. Jang, "GPS-based mobile exercise application: An alternative tool to assess spatio-temporal patterns of visitors' activities in a national park," *J. Park Recreation Admin.*, vol. 37, pp. 124–134, Jan. 2019.

[9] A. Dolinayova, J. Masek, M. Kendra, J. Camaj, D. Grandsart, E. Marlier, P. Colzani, M. Arena, J. Paragreen, P. Navaratnam, and M. Brennan, "Research of the passenger's preferences and requirements for the travel companion application," *J. Adv. Transp.*, vol. 4, pp. 1–12, Jan. 2018.

[10] K. C. Ciesielski, A. X. Falcão, and P. A. V. Miranda, "Path-value functions for which Dijkstra's algorithm returns optimal mapping," *J. Math. Imag. Vis.*, vol. 60, no. 7, pp. 1025–1036, Sep. 2018.

[11] J. Qiu and R. Wang, "Automatic extraction of road networks from GPS traces," *Photogramm. Eng. Remote Sens.*, vol. 82, no. 8, pp. 593–604, Aug. 2016.

[12] L. Yang, X. Wang, G. Sun, and Y. Li, "Modeling the perception of walking environmental quality in a traffic-free tourist destination," *J. Travel Tourism Marketing*, vol. 37, no. 5, pp. 608–623, Jun. 2020.

[13] L. Liu, H. Mu, X. Yang, R. He, and Y. Li, "An oriented spanning tree based genetic algorithm for multi-criteria shortest path problems," *Appl. Soft Comput.*, vol. 12, no. 1, pp. 506–515, Jan. 2012.

[14] E. Jenelius, "Public transport experienced service reliability: Integrating travel time and travel conditions," *Transp. Res. A, Policy Pract.*, vol. 117, pp. 275–291, Nov. 2018.

[15] Y. Yang, C. Wang, W. Liu, and P. Zhou, "Understanding the determinants of travel mode choice of residents and its carbon mitigation potential," *Energy Policy*, vol. 115, pp. 486–493, Apr. 2018.

[16] D. He, H. Liu, K. He, F. Meng, Y. Jiang, M. Wang, J. Zhou, P. Calthorpe, J. Guo, Z. Yao, and Q. Wang, "Energy use of, and CO_2 emissions from China's urban passenger transportation sector–carbon mitigation scenarios upon the transportation mode choices," *Transp. Res. A, Policy Pract.*, vol. 53, pp. 53–67, Jul. 2013.

[17] A. Neves and C. Brand, "Assessing the potential for carbon emissions savings from replacing short car trips with walking and cycling using a mixed GPS-travel diary approach," *Transp. Res. A, Policy Pract.*, vol. 123, pp. 130–146, May 2019.

[18] M. D. Yap, G. Correia, and B. van Arem, "Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips," *Transp. Res. A, Policy Pract.*, vol. 94, pp. 1–16, Dec. 2016.

[19] C. Koo, K.-H. Yoo, J.-N. Lee, and M. Zanker, "Special section on generative smart tourism systems and management: Man–machine interaction," *Int. J. Inf. Manage.*, vol. 36, no. 6, pp. 1301–1305, Dec. 2016.

[20] K. Kabassi, "Personalizing recommendations for tourists," *Telematics Informat.*, vol. 27, no. 1, pp. 51–66, Feb. 2010.

[21] X. Hu, C. Zhang, G. Zhang, M. Zhang, H. Li, M. Leeson, and J. Liao, "Finding the k shortest paths by ripple-spreading algorithms," *Eng. Appl. Artif. Intel.*, vol. 87, pp. 1–12, Jan. 2020.

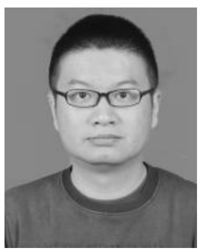
[22] H.-C. Huang and C. I. Hou, "Tourism demand forecasting model using neural network," *Int. J. Comput. Sci. Inf. Technol.*, vol. 9, no. 2, pp. 19–29, 2017.

[23] Y.-C. Hu, P. Jiang, and P.-C. Lee, "Forecasting tourism demand by incorporating neural networks into Grey–Markov models," *J. Oper. Res. Soc.*, vol. 70, no. 1, pp. 12–20, Jan. 2019.

[24] S. Lin, C. Dong, M. Chen, F. Zhang, and J. Chen, "Summary of new group intelligent optimization algorithms," *Comput. Eng. Appl.*, vol. 54, no. 12, pp. 6–14, 2018.

[25] L. Wang and X.-L. Zheng, "A knowledge-guided multi-objective fruit fly optimization algorithm for the multi-skill resource constrained project scheduling problem," *Swarm Evol. Comput.*, vol. 38, pp. 54–63, Feb. 2018.

[26] X. Hu and Z. Deng, "Research on perception bias of implementation benefits of urban intelligent transportation system based on big data," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–7, Dec. 2019.
[27] J. Xie and Y. K. Choi, "Hybrid traffic prediction scheme for intelligent transportation systems based on historical and real-time data," *Int. J. Distrib. Sens. Netw.*, vol. 13, no. 11, 2017, Art. no. 1550147717745009.
[28] W. Zheng and Z. Liao, "Using a heuristic approach to design personalized tour routes for heterogeneous tourist groups," *Tourism Manage.*, vol. 72, pp. 313–325, Jun. 2019.
[29] Q. Li, F. Wang, J. Wang, and W. Li, "LSTM-based SQL injection detection method for intelligent transportation system," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4182–4191, May 2019.
[30] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowl.-Based Syst.*, vol. 172, pp. 1–14, May 2019.
[31] Z. Diao, D. Zhang, X. Wang, K. Xie, S. He, X. Lu, and Y. Li, "A hybrid model for short-term traffic volume prediction in massive transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 935–946, Mar. 2019.
[32] W. T. Zhao, P. Li, Q. Liu, D. Liu, and X. W. Liu, "Selecting the optimal hidden layer of extreme learning machine using multiple kernel learning," *Ksii Trans. Internet Inf. Syst.*, vol. 12, pp. 5765–5781, Dec. 2018.
[33] X. Kong, F. Xia, J. Wang, A. Rahim, and S. K. Das, "Time-location-relationship combined service recommendation based on taxi trajectory data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1202–1212, Jun. 2017.
[34] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3135–3146, Nov. 2017.
[35] Y.-C. Chiou, R.-C. Jou, and C.-H. Yang, "Factors affecting public transportation usage rate: Geographically weighted regression," *Transp. Res. A, Policy Pract.*, vol. 78, pp. 161–177, Aug. 2015.
[36] T. Ma, G. Motta, and K. Liu, "Delivering real-time information services on public transit: A framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2642–2656, Oct. 2017.
[37] Y. Liu, K.-F. Lam, J. T. Wu, and T. T.-Y. Lam, "Geographically weighted temporally correlated logistic regression model," *Sci. Rep.*, vol. 8, no. 1, p. 1417, Dec. 2018.
[38] H. Liu, C. Jin, and A. Zhou, "Popular route planning with travel cost estimation from trajectories," *Frontiers Comput. Sci.*, vol. 14, no. 1, pp. 191–207, Feb. 2020.
[39] M. Meng, A. Rau, and H. Mahardhika, "Public transport travel time perception: Effects of socioeconomic characteristics, trip characteristics and facility usage," *Transp. Res. A, Policy Pract.*, vol. 114, pp. 24–37, Aug. 2018.
[40] K. H. Lim, J. Chan, S. Karunasekera, and C. Leckie, "Tour recommendation and trip planning using location-based social media: A survey," *Knowl. Inf. Syst.*, vol. 60, no. 3, pp. 1247–1275, 2019.
[41] J.-Q. Li, K. Zhou, L. Zhang, and W.-B. Zhang, "A multimodal trip planning system with real-time traffic and transit information," *J. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 60–69, Apr. 2012.
[42] D. Esztergár-Kiss and C. Csiszár, "Evaluation of multimodal journey planners and definition of service levels," *Int. J. Intell. Transp. Syst. Res.*, vol. 13, no. 3, pp. 154–165, Sep. 2015.
[43] S. R. M. Sakip, N. M. Akhir, and S. S. Omar, "Determinant factors of successful public parks in Malaysia," *Procedia Social Behav. Sci.*, vol. 170, pp. 422–432, Jan. 2015.
[44] A. Nuzzolo and A. Comi, "A normative optimal strategy in intelligent transit networks," *Transp. Res. Procedia*, vol. 27, pp. 380–387, Jan. 2017.
[45] Y. Sui, F. Shao, X. Yu, R. Sun, and S. Li, "Public transport network model based on layer operations," *Phys. A, Stat. Mech. Appl.*, vol. 523, pp. 984–995, Jun. 2019.
[46] E. Algaba, V. Fragnelli, N. Llorca, and J. Sánchez-Soriano, "Horizontal cooperation in a multimodal public transport system: The profit allocation problem," *Eur. J. Oper. Res.*, vol. 275, no. 2, pp. 659–665, Jun. 2019.
[47] Y. Gao, S. Rasouli, H. Timmermans, and Y. Wang, "Trip stage satisfaction of public transport users: A reference-based model incorporating trip attributes, perceived service quality, psychological disposition and difference tolerance," *Transp. Res. A, Policy Pract.*, vol. 118, pp. 759–775, Dec. 2018.
[48] B. Yang, X. Luan, and Y. Zhang, "A pattern-based approach for matching nodes in heterogeneous urban road networks," *Trans. GIS*, vol. 18, no. 5, pp. 718–739, Oct. 2014.
[49] J. Kim, B. Thapa, and S. Jang, "GPS-based mobile exercise application: An alternative tool to assess spatio-temporal patterns of visitors' activities in a national park," *J. Park Recreation Admin.*, vol. 47, pp. 124–134, Jan. 2019.
[50] A. Gunawan, H. C. Lau, and P. Vansteenwegen, "Orienteering problem: A survey of recent variants, solution approaches and applications," *Eur. J. Oper. Res.*, vol. 255, no. 2, pp. 315–332, Dec. 2016.
[51] X. Wu, Z. Huang, X. Peng, Y. Chen, and Y. Liu, "Building a spatially-embedded network of tourism hotspots from geotagged social media data," *IEEE Access*, vol. 6, pp. 21945–21954, 2018.
[52] D. Rachmawati, "Implementation of modified cheapest insertion heuristic on generating Medan city tourism route," *J. Phys., Conf. Ser.*, vol. 1566, Jun. 2020, Art. no. 012076.
[53] H. Liu, Y. Zhang, Q. Zhu, and J. Chu, "Environmental efficiency of land transportation in China: A parallel slack-based measure for regional and temporal analysis," *J. Cleaner Prod.*, vol. 142, pp. 867–876, Jan. 2017.
[54] S. Liu, S. Wang, C. Liu, and R. Krishnan, "Understanding taxi drivers' routing choices from spatial and social traces," *Frontiers Comput. Sci.*, vol. 9, no. 2, pp. 200–209, Apr. 2015.
[55] G. Wang, S. Shin, and W. Lee, "A text sentiment classification method based on LSTM-CNN," *J. Korea Soc. Comput. Inform.*, vol. 24, no. 12, pp. 1–7, 2019.
[56] S. W. G. AbuSalim, R. Ibrahim, M. Z. Saringat, S. Jamel, and J. A. Wahab, "Comparative analysis between dijkstra and bellman-ford algorithms in shortest path optimization," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 917, Sep. 2020, Art. no. 012077.
[57] Y. Chen, "Application of improved Dijkstra algorithm in coastal tourism route planning," *J. Coastal Res.*, vol. 106, no. 1, pp. 251–254, Jul. 2020.
[58] A. Khamami and R. Saputra, "The shortest path search application based on the city transport route in Semarang using the Floyd-warshall algorithm," *J. Phys., Conf. Ser.*, vol. 1217, May 2019, Art. no. 012116.
[59] B. R. Devi, K. K. Rao, and M. A. Rani, "Application of modified Bellman-ford algorithm for cooperative communication," *Wireless Pers. Commun.*, vol. 109, no. 3, pp. 2025–2049, Dec. 2019.
[60] R. Ramadiani, D. Buchori, A. Azainil, and N. Dengen, "Floyd-warshall algorithm to determine the shortest path based on android," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 144, no. 1, Apr. 2018, Art. no. 012019.



XIAO ZHOU received the B.S., M.S., and Ph.D. degrees from Information Engineering University, Zhengzhou, China, in 2009, 2012, and 2015, respectively. He is currently a Postdoctoral Researcher with the Chengdu Institute of Computer Application, Chinese Academy of Sciences. He is the program facilitator or main researcher of eight provincial-level research projects, got three patents for invention, and published three works. He has published over 50 research papers in journals and international conferences. His research interests include geospatial information analysis and modeling, smart tourism mathematics, and spatial data mining.

MINGZHAN SU received the B.S. and M.S. degrees from Information Engineering University, Zhengzhou, China, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with PLA Strategic Support Force Information Engineering University. His main research interests include geographic information data processing and modeling, spatial data analysis, and spatial data mining.

GUANGHUI FENG is currently pursuing the Ph.D. degree with Guangzhou University. He is currently an Associate Professor with the Zhengzhou University of Industrial Technology. He has presided over and participated in more than five research projects. He has published ten research papers in journals and international conferences. His main research interests include computer application technology, machine learning, and data mining.

XINGHAN ZHOU received the B.S. degree from the Jingdezhen Ceramic Institute, in 2016, and the M.S. degree from the Sichuan University of Science and Engineering, in 2019. His main research interests include intelligent tourism and geospatial data processing.