

A tourism preference construction model based on complex network analysis

Lina Zhong

Beijing Foreign Studies University

Baolin Deng

Beijing International Studies University

Jackie Rong

Monash University School of Information Technology

Linfeng Dong (✉ dlfphd@163.com)

Haoyu Wang

Hainan University

Haoyu Kong

Australian National University

Research

Keywords: complex network, tourists' preferences, model, Louvain

Posted Date: March 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-15544/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A tourism preference construction model based on complex network analysis

Lina Zhong¹², Baolin Deng¹², Jia Rong³, Linfeng Dong^{4*}, Haoyu Wang⁴, Haoyu Kong⁵

¹School of Tourism Sciences, Beijing International Studies University
Chaoyang District, Beijing, China
Email: zhonglina@bisu.edu.cn

²Institute for Big Data Research in Tourism, Beijing international studies university
Chaoyang District, Beijing, China
Email: dengbaolin0225@163.com

³ Department of Data Science & AI
Monash University, Melbourne, Australia
Email: jackie.rong@monash.edu

⁴Tourism College of Hainan University
Meilan District, Haikou, China
Email: dlfphd@163.com

⁵Tourism College of Hainan University
Meilan District, Haikou, China
Email: whycritical@163.com

⁵College of Business and Economic, Australian National University
Canberra, Australia
Email: haoyu.kong@anu.edu.au

* Corresponding Author
Linfeng Dong

Abstract: The travel notes contain a wealth of tourists' behavior information, which provides a new way to study tourists' preferences. How to mine the text of online travel notes accurately and efficiently has become the key to research tourists' preferences. In this paper, the theory and technology of text mining were introduced into the research of tourists' preference through a large number of online travel notes accumulated on the Internet. The main research work of this paper was as follows: (1) The tourists' preference model was constructed by complex network method; (2) The travel notes data of Sanya tourists as an example was crawled and analyzed. In this paper, the theory of network travel data and text mining is introduced into the study of tourists' preferences, which not only improves the data quality of traditional preference

research field, but also provides a new method for mastering tourists' preferences more accurately.

Keywords: complex network; tourists' preferences; model; Louvain

1 Introduction

More and more tourists prefer to share travel notes on the Internet, and online travel notes as an important part of UGC data source has become an important reference for other tourists in making travel decisions. Among them, the long travel notes involve more and more specific tourist information, which becomes an important communication channel for tourists to perceive the tourist destination, and has a high reference value for revealing tourists' preference for destination.

In this paper, aiming at the long text mining, we designed a method based on the complex network analysis to obtain the elements of travel notes and use the complex network analysis method to analyze the community, and carried out the automatic emotion recognition for the community and elements, so as to establish a set of methods to construct the tourism preference model from the long text travel notes. It is hoped that the tourist preference model based on the complex network method will be well known and even used by more tourism scholars. On the other hand, it will provide useful ideas for the study of tourist preference using long text travel notes.

The specific contributions of this paper include:(1)A literature survey about various existing research about tourists' preferences and application for complex network into

tourism and analyzed their advantages and disadvantages.(2)An effective a tourists' preferences models based on complex network for tourism destination management is proposed.(3)Performance analysis of the proposed model and validation of tourists' preferences.

The rest of this paper is organized as follows. Section 2 discusses related works, followed by the model for tourists' preferences models based on complex network designed in Section 3. Section 4 shows the research results based on proposed model, and Section 5 concludes the paper with summary and research limitations.

2 Related works

2.1 Research progress of tourists' preferences

The research of foreign tourism preference has a wide range of contents, which not only involved the direct preference analysis of tourist destination and reception facilities preference, special group preference, but also involved the relationship between tourism preference and tourists' needs, gender, the choice of career of tourism majors, the work preference of tourism agencies, etc[1]. In terms of research methods, foreign scholars attach great importance to the application of quantitative techniques and methods for empirical analysis, in addition to the use of multiple logit models (MNL, multinomial logit model) [2], conditional logit model (CLM, conditional logit model) [3], nested logit model (NL, nested logit model) [4] and mixed logit model (ML, mixed logit model) [5]are used to analyze tourists' or residents' preferences for tourism, leisure, recreation and reception facilities. And some researchers also made full use of the multidimensional scaling approach [6], the TAT (thematic apperception test) [7], and fuzzy set theory. And The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [8], and decision tree analysis [9].Such as analyzing the preferences of tourists or residents on tourism, leisure and recreation, and reception facilities.In China, the research techniques and methods of tourism preference were mainly through questionnaire survey, and then the data obtained were quantified in the form of histogram or pie chart plus statistical data analysis to solve the problem. Based on the market survey data of American tourists in six hot tourist cities in China, Ma Yaofeng et al. researched the law of preference of American tourists to China [10]. Through a questionnaire survey, Wu Bihu et al found the recreational behavior and preference characteristics of local residents around Hangzhou [11]. Taking Xi'an inbound tourists as an example, Bai Kai et al researched tourists' shopping preferences and behavior rules [12]. Chen Nan et al analyzed the relationship between tourism risk perception and tourism behavior preference through empirical investigation [13].Sun Gennian established a regression model to reflect international tourism payments, revealing the correspondence between tourism payments at various levels and GNP per capita [14]. Tian Xiangli et al. found the tourists' cognitive and selection preferences in the Qinling Scenic Spot in Shaanxi based on the perspective of differences in the source groups [15].

2.2 Research on the application of complex network in Tourism

Network analysis and tourism were both young subjects, and the application of network theory and analysis methods in tourism research started late. Among them, foreign scholars applied network analysis to tourism research earlier, and first started to study the relationship between recreational resources and users in the 1980s. Since then, with the in-depth development of network theory, network analysis method has been paid attention to in tourism research and has been applied by more and more scholars [16]. After entering the 21st century, the increasingly complex interaction between various factors in the tourism system further promotes researchers to perspective tourism phenomenon from the perspective of social network, which made network analysis becoming a hot field in tourism research. Scholars at home and abroad began to pay attention to the application of network science, especially the social network analysis method in tourism research. From the perspective of the specific research progress of social network in the field of tourism, the research content mainly focuses on the analysis of tourism destination government [17], tourism policy network [18], tourism cooperation research [19,20], characteristics of tourism route network [21,22], tourists' behavior [23], tourism enterprises [24,25], other related issues [26,27].

Generally speaking, the application of complex network analysis in the field of tourism mainly focused on two aspects: one was to use the network analysis method to research the development of business network and the internal relationship between organizations; the other was to apply it to tourism policy, including the public-private relationship among all stakeholders and the structure of Tourism governance. The scope of research is still expanding. The types of network structure included communication network, virtual network and cooperative network. At the same time, network measurement methods also used density method, structural hole, strong / weak tie method, clustering method and efficiency method [28].

3 Research method

3.1 Platform selection

This paper chose Sanya city as the research sample, it's one of the most popular city in China. As the largest online travel website in China, Trip has a total of 250 million users and 5 million daily active users in the community. Trip has collected 30 million real users' travel notes and hotel reviews. Therefore, this paper chose Trip.com Group (www.trip.com) as the source website of travel notes of sample city.

3.2 Construction of tourist preference content mining model

3.2.1 The construction of preference complex network node

First of all, we took the whole text set as the object to build a dictionary of all the nouns in the text set, that was, all the non repeated nouns in the text set are encoded. Secondly, calculated the word frequency of each noun in each dictionary. In the field of text topic mining, words with too high or too low word frequency did not have topic differentiation. Therefore, this paper eliminated the words with the highest and the

lowest word frequency of 5% in the dictionary. Finally, we extracted the top 50% of words to build a new dictionary, which is the initial node of complex network. The dictionary that defines the construction of noun words after word frequency screening is Dic , $Dic = \{0: w_1; 1: w_2; \dots; t: w_t\}$ $t = 0.5 * T$, After the highest and lowest 5% of words are proposed for the text set, the top 50% of the non repetitive nouns in the word frequency list. Finally get $V(G) = \{v_1, v_2, \dots, v_n\}$, $v_1 = w_1, v_2 = w_2, \dots, v_n = w_n$.

3.2.2 The construction of preference complex network edge

It mainly solves the construction “ $E(G) = \{e_1, e_2, \dots, e_m\}$ ” of edge between nodes and the assignment algorithm of edge weight.

The definition distance is: the number of other sentences between the sentences where two words are located. If two words appear in the same sentence, that is, co-occurrence, the distance between the two words is 1. By analogy, if there are n sentences between the sentences where the two words are located, then according to the definition of this paper, the distance between the two words is $n + 1$.

The text set of all Internet travel notes $D = (d_1, d_2, \dots, d_n)$, the text of part i is composed of k sentences, namely, $d_i = \{s_1, s_2, \dots, s_k\}$, R_{di} is the degree of relevance in Chapter v_i, v_j , As shown in the algorithm.

$$R_{di} = \sum_{t=1}^n N_t * H_t = \sum_{t=1}^n N_t / t \quad (1)$$

Where is the number of times that the nodes v_i, v_j appear in the first travel in the way of the distance t, H_t is the weight of the node v_i and v_j appears once in the form of the distance t, which is $1/t$ in this paper. That is, if the nodes v_i, v_j appear three times with the distance 3, the degree of association is the same as the nodes v_i, v_j .

After calculating the degree of association of each node v_i, v_j , the association degree of all the texts in the text set is summed, and the total correlation degree R of the nodes v_i, v_j is obtained, as shown by the algorithm.

$$R(v_i, v_j) = \sum_{i=1}^n R_{di} \quad (2)$$

Find the correlation matrix M_R of all nodes, as shown in the matrix. If the element $R(v_i, v_j) \neq 0$ in the affinity matrix, where $1 \leq i, j \leq n$, then an edge is constructed between node v_i and v_j , and the weight of the edge is set to $R(v_i, v_j)$. At this point, the construction of a preference complex network is completed.

$$M_R = \begin{matrix} & R(v_1, v_1) & \cdots & \cdots & R(v_1, v_n) \\ & \vdots & & \ddots & \vdots \\ & \vdots & & \ddots & \vdots \\ & R(v_n, v_1) & \cdots & \cdots & R(v_n, v_n) \end{matrix} \quad (3)$$

3.2.3 Preference content mining based on Louvain community mining algorithm

After building a preference complex network, we need to further explore the community from this complex network. The community mined is the set of nodes in the complex network constructed above. These sets are various preferences. In this section, based on Louvain community mining algorithm, we use complex network to mine the tourists's preference to get the preference content of tourists. Louvain algorithm is a community discovery algorithm based on the module degree. The main goal is to divide the community continuously so that the module degree of the whole network increases continuously. The algorithm uses module gain to calculate the ΔQ change of complex network module degree caused by adding a new node to the community. The algorithm is shown as follows.

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (4)$$

Where \sum_{in} represents the sum of all edge weights in community C , \sum_{tot} represents the sum of the weights of the edges of all points in community C , k_i represents the sum of the weights of the edges of the nodes i , and $k_{i,in}$ represents the sum of the weights of the edges between Node B and all points in Community C . m represents the sum of the weights of all the edges in the complex network.

Step 0 : think of each point as a separate community.

Step 1 : select a community as the starting point, divide adjacent points into the community and calculate the modularity at this time, judge whether the difference between the modularity before and after the division is a positive number, if it is a positive number, accept the division this time, if it is not a positive number, cancel the Division this time.

Step 2 :repeat the first and second steps, the modularity will not increase any more, that is to say, it will be optimal.

Step 3 :take each community mined in step 2 as a new point, build a new network, and continue to carry out step 1 and step 2 to get the community level. As shown in the Fig. 1 below.

Set the weight threshold value of the edge of the preferred complex network to prune the edge of the preferred complex network. If the weight of the edge is less than the threshold value, then prune the edge. If all the edges of a node are pruned, so that the point is isolated from the complex network, then prune the point.

After pruning, we use Louvain community mining algorithm to mine the preference complex network.

3.2.4 Construction of Preference Content Classification System Based on Image Perception Elements of Tourism Destinations

By combing the relevant theories of tourist destination image perception [29,30], this paper constructs the preference content of tourists as shown in Fig 2. Among them, the preference content, the preference composition, the preference element correspond to the feelings in the tourism location perception theory, the preference community and the preference content correspond to the community mining result, and the preference sub-element firstly prefers the preference community element according to the preference element from top to bottom. After categorization, combined with domain knowledge and experience, the bottom-up summary of the preference elements is a link in the whole preference content category system. Whether the preference elements have sub-elements depends on the richness of preference elements.

Due to the complexity of the tourist's perception of the destination, the preference elements in a preferred community may belong to multiple preference sub-elements, and one preference sub-element may also contain multiple preference communities.

3.2.5 Model validation

In the field of data structure, in an arrangement, if the front and back positions of a logarithm are opposite to the size order, that is, the front number is larger than the back number, then they are called a reverse order, and the total number of all the reverse orders in an arrangement is called the reverse order number of this arrangement. For example, if the standard order is 12345, the reverse order number of 21345 is 1. We introduced the concept of reverse order into the accuracy test of preference model, and defined the accuracy, as shown in the algorithm.

$$V = 1 - \left(\frac{B}{T} \right) \quad (5)$$

In which, B is the reverse order number of scenic spots mined according to the preference model, and t is the total order number of scenic spots sorted according to the number of favorable comments.

4 Results

4.1 Preference analysis based on structured data of network travel notes

In addition to text data, there were other structured data of tourists' travel in this paper. Through direct mathematical statistical analysis, we can get the tourism preferences of domestic tourists to Sanya. In this section, we mainly researched the following preferences of tourists based on the structured data of the crawled travel notes: travel time preference, partner preference, play preference, cancellation preference and preference and gender preference.

In the table 1, 17239 pieces of effective travel data were crawled. Each travel record included a trip of tourists. From this data table, we mainly analyzed four kinds of structured data: travel time, companionship, play methods and consumption. In the user information table, we crawled the user information of 13900 valid users, and we analyzed the gender data.

Table 1 Table of preference analysis based on structured data

Gender analysis		
Gender	Number	Proportion
Women	9359	0.6733
Man	4541	0.3267
Days		
Day	Number	Proportion
1-2	8740	0.5943
3-5	3874	0.2634
6-8	1535	0.1044
9-14	312	0.0212
Over 15	245	0.0167
Way of company		
Way of company	Number	Proportion
One	565	0.0968
With friends	2005	0.3437
Lovers	636	0.1090
Spouse	805	0.1380
Parents	522	0.0895
Parenting	1301	0.2230
Types of Tourism		
the way of playing	Number	Proportion
Delicious food	1185	0.3334
Shopping	281	0.0791
Shoot	1165	0.3278
Self driving	378	0.1064
Tourist group	107	0.0301
on foot	99	0.0279
Riding	33	0.0093
Poor travel	306	0.0861
Consumption		
Consumption range /Yuan	Number	Proportion
Below 500	271	0.0515
500-1000	220	0.0418
1000-3000	1457	0.2768
3000-5000	2237	0.4250

5000-7000	586	0.1113
7000-9000	201	0.0382
9000-15000	230	0.0437
15000-20000	31	0.0059
Over 20000	31	0.0059

4.2 Preference analysis based on online travel notes

Pruning threshold settings for complex networks are not supported by theory and experience. Therefore, this paper used the experimental method to perform multiple simulation calculations and obtains the optimal pruning threshold of 20. As shown in Fig. 3 below.

In this paper, we used Python's community library to implement the Louvain community mining algorithm for the preference complex network, and construct the preference content classification system according to the classification method discussed above. The results of community mining are shown in the Fig. 4, in which different communities are represented by different gray levels.

4.2.1 Preference mining results based on online travel preference model

The results of preference mining are shown in the table 2, where the value of preference degree is accurate to two decimal places.

Table 2 Statistics table of preference mining results

Community	Element	Sub factor
Community3	Tourist attractions 4.38 attraction 7.66 Kam Mu Kok 6.56	Kam Mu Kok 6.20
Community3	Buddha 5.35 mountain road 6.65 Petals 9.35 Nanshan Temple 5.69 worship Buddha 3.22 incense 1.87 Buddhist culture 8.52 Exhibition hall 9.36 Nanshan 4.75 South China Sea 8.29 Yuan 8.56 Avalokitesvara 9.56 Buddhism 9.86 realm 6.33 Gate 8.69 town 4.69 Guanyin 9.56 Jinyu 9.87 Strength 6.88 Zhai 4.92 element 7.29 Temple 8.63 Guanyin Pavilion 8.42 Zen 11.65 wisdom 9.75 Living beings 6.98 Avalokitesvara 7.99 Top 10.23 plants 8.65 Vegetation 9.88 method 9.01 Taoist 8.66 battery car 3.56 Merit 4.56 blessing 13.03	Nanshan Cultural Tourism Zone 8.37

Lotus 7.89 Bodhisattva 10.35
 Furudonghai 12 burning incense 9.35
 Man made 6.56 Temple 7.96
 King Kong 8.35 incense 9.88
 Hall 9.01 shoubinan mountain 9.85
 throne 10.75 Buddha 9.88
 Whisper 9.53 Nantian 12.89
 Crossing Jianglong 9.38 Taoism 10.86
 Holy way 9.11 dragon 9.56

Community3	The remotest corners of the globe 10.01 Tianya 9.35 walk 8.79 often live 8.03	The remotest corners of the globe 9.05
Community3	Dadonghai 9.23 white beach 8.86 winter swimming 5.69 free 9.51 egret Park 8.69 crescent 9.86	Great East China Sea 8.64
Community3	Sanya Bay 2.69 20 Li 2.26 coconut Gallery 6.98	Sanya Bay 3.98
Community3	Luhuitou 8.89 walk 7.32 scenic spot 8.36 ticket-2.10 Park 8.98 statue 9.69 platform 7.65 love 9.42 tourist area 10.65 deer buy ticket-0.12 culture 7.69 tourist bus 9.19	Luhuitou scenic spot 7.22
Community3	Xiaodongtian 8.56 Changshou 4.26 tickets 3.56 fare 1.25 charge -0.16 stone 6.89 Holy Land 9.89 bus 5.46 state guesthouse 7.26	Big and small cave 5.30
Community3	Yanoda 8.13 water 8.69 Admission 6.89 Liuquan 9.75 Black dragon 6.10 physical strength 7.56 BC 9.19 site 7.63 Tree 8.26 Valley 9.46 cable bridge 10.13 Shuqi 5.65 dragon blood tree 8.38 Oxygen bar 10.65 rainforest 8.10 see the tree 9.33	Yanoda 8.37
Community3	Fenghuangling 2.68 lover 3.98 hundred flowers 4.23 amorous feelings 2.44 fantasy 1.39 paradise 7.23 infatuation 3.54 set tickets 1.89 movie 3.45 Hall 1.33 suspension bridge 3.01 couple 4.89 circle 1.59 landscape 7.23 Tingtao 5.89 dialect 3.14 Fenghuangling 2.68 lover 3.98 hundred flowers 4.23 amorous feelings 2.44 fantasy 1.39 paradise 7.23 infatuation 3.54 set tickets 1.89 movie 3.45 Hall 1.33 suspension bridge 3.01 couple 4.89 circle 1.59 landscape 7.23 Tingtao 5.89 dialect 3.14	Fenghuangling haishishan League scenic spot 3.75
Community3	GeYou 6.65 tropical rain forest 7.23 cultural area 4.12 field 5.13 medicated meal 6.89	Tropical Paradise Forest Park 7.40

	paradise on earth 7.21 treasure land 8.38 waterway 6.62 Park 7.02 orchid 7.12 Forest Park 7.01	
Community3	Yulin port 1.98 military 4.50	Yulin 3.24
community6	Island 8.99 tourist 1.39 coral 5.23 diving 7.98 West Island 4.09 motorcycle 7.65 Island 3.54 speedboat 3.98 wharf 5.39 swimsuit 6.01 gesture 6.12 surfboard 8.82 specialty 2.31 whole body 8.05 visibility 5.13	West Island 5.48
community6	Wuzhizhou Island 12.87 boat diving 8.12 motorboat 9.25 boat 7.23 fish 8.22 Mazu Temple 7.56 deep sea 10.73 sightseeing 9.12 Diving suit 9.80 oxygen bottle 9.59 tropical fish 8.88 equipment 7.36 item 9.77 experience 10.03 glass 9.76 Wuzhizhou Island 12.87 boat diving 8.12 motorboat 9.25 boat 7.23 fish 8.22 Mazu Temple 7.56 deep sea 10.73 sightseeing 9.12 Diving suit 9.80 oxygen bottle 9.59 tropical fish 8.88 equipment 7.36 item 9.77 experience 10.03 glass 9.76 Price 5.98 amusement 10.55 coral reef 9.93 staff 11.78 surfing 11.03 party 9.24 weaving sister 9.83 Angkor 9.95 giant turtle 10.00 exploring the sea 9.51 golden turtle 8.99 Hometown 10.09 day watching 9.78 lover 8.34 turtle 9.67	Island of Centipede 9.43
community6	Jiejiezhou 4.56 rock 7.45 wind and waves 7.01 sailboat 8.82 sea fishing 3.89 Art Museum 5.56 Island 8.12 Pavilion 3.65 port 7.35 yacht 7.28 depth 6.12	Boundary Island 6.34
community6	Monkey Island 9.09 cable car 8.33 one way 6.12 boat ride 8.90 monkey 9.69 full package 11.30 high altitude 6.22 little monkey 11.93 rock 9.09 bridge crossing 6.26 slide 9.14 monkey 9.32 plank bridge 7.26 Chain bridge 8.99	monkey island 8.69
community11	Seven cents ridge 9.36 Hot spring 8.21 Water temperature 8.98 small fish 6.12 pool 7.02 bus 1.23 spring 8.12 bubble 8.26 bubble pool 9.31 garden 8.23 fish therapy 7.03 fish pool 7.98	Seven cents ridge 7.49
community11	Nantian 6.62 hot spring 5.29 hero slope 3.99	Nantian hot spring 5.30
community19	Longxingtianxia 9.23 Tianchi 9.10 cliff 7.52	Yalong Bay 8.62

community38	Arcade 7.73 building 7.85 snack street 9.23 landmark 5.98 Nanyang 6.29 old street 7.21 bell tower 8.09	Arcade 7.49
community39	Gongshan 3.05 gross 2.49	Mao Gong Shan 2.77
community16	Tiancheng 4.22 warehouse - 34500 Quanhe 8.36 forum 9.56 Asia 4.32 conference site 8.13 Boao 7.21 Yudai beach 11.03 town 4.23 Pinglang 2.11 Qionghai 3.26	Venue of Boao 5.36
Community0	Women's army 4.03 Museum 3.26	Memorial Park of the detachment of women 3.65
Community2	National Shell Museum 4.58	Shell Museum 4.58
Community2	Theme - 0.21 tourist 1.23 children's Park - 8.36 Playground - 6.37 roller coaster 2.13	Theme Park -2.32
Community2	Butterfly valley -5.12	Butterfly Valley -5.12
community7	Government 9.23 performing arts 10.32 eternal feelings 11.35 world 1.03 langlangshui Park 6.98 Song cheng-5.32 song and dance 12.26 two person 8.78 peacock-9.23 animal 2.13 zoo 11.30 Theater 9.56 pigeon 6.26 color 9.11 ice snow 10.36 stage 8.23 history 8.70 actor 6.59 theater 9.56 ice sculpture 12.03 opening 8.24 audience 5.22 recruit son-in-law 7.59 Chengfu 7.68	Songcheng eternal love 7.00
community60	Yazhou 7.77 Zhifu 8.23 ancient city 5.96 master Jianzhen 4.32 east crossing 5.60	Ancient city of Ya Zhou 6.38
community74	Miao Village 4.21 Li village 4.93	Li Cun Miao Village 4.57
Community3	Totem 9.85 nationality 8.25 village 2.98 tattoo 4.26 women 7.23 ancient rhyme 3.56 exotic 2.78 implication 1.06 boulder 7.56 yanbo-3.98 Miao 6.45 Li 7.13 waterfall 6.59 sculpture 6.98 Li Miao 5.55 cliff 2.79 guild hall 3.89 folk culture 9.19 grandma 3.85 Grand Master 4.32 stone carving 7.21 Li family 6.29	Ancient coconut Village 5.17
community86	College 6.32 Dongpo 8.71	dongpo academy 7.52

4.3 Model validation

In the captured scenic spot information data table, the number of favorable comments field can accurately reflect tourists' preferences for scenic spots. Therefore, this paper compared the preference sub elements (scenic spot name) contained in the scenic spot preference elements mined by the model with the preference information analyzed in the scenic spot information table to verify the accuracy of the model built in this paper. Because the information data table captured in this paper only contains the data of all scenic spots in Sanya City, only the scenic spots belonging to Sanya City in the text mining are compared.

We took the ranking of scenic spots in the statistical table of scenic spots' favorable comments as the standard order, then according to the definition of the reverse order number, the reverse order number of scenic spots in the table of scenic spots' preference elements and preference degrees, the total order number is T , and the accuracy of preference mining model is $V = 1 - \left(\frac{B}{T}\right) = 0.8579$.

According to the test results of the model, the accuracy of the proposed preference model is 85.89%, which can accurately mine the tourists' preferences.

5 Results and Discussion

Based on the vast amount of online travel notes accumulated on the Internet, this paper introduces the theory and technology of text mining into the study of tourists' preferences, which provide a new method for accurately mining tourists' preferences from the massive online travel notes. It has certain reference significance both in theory and in practice. First, based on the theory of complex network and the theory of tourism destination perception, this paper proposed distance algorithm and relevance algorithm, constructs the model of tourists' preference content mining, and realized the preference content mining of massive text data; second, based on the theory of emotion analysis, through the construction of emotion dictionary and degree adverb dictionary, the model of tourists' preference degree is established, and tourists' preference degree is accurately mined. The degree of preference for various preferred content. (5) take Sanya City as an example to verify the model and fully verify the reliability of tourists' preference model.

Table 3 Statistics table of sights praise

Name of scenic spot	Amount of favourable comments
Wuzhizhou island	11766
The remotest corners of the globe	10897
Tropical Paradise Forest Park	9474
Yalong Bay	9367
Nanshan Cultural Tourism Zone	8681
Luhuitou scenic spot	7193
The scene area of song city in Sanya	6265
West Island	2852
Big and small cave	2225
The Gulf of Sanya	1822
Great East China Sea	1821
Nantian hot spring of Pearl River	1149

Fenghuangling haishishan League scenic spot	884
Yalong Bay Shell Museum	73
Ancient city of Ya Zhou	43
Kam Mu Kok	22
Butterfly Valley	10
Li Cun Miao Village	5
Yulin	3

Table 4 Sight class preference table

Name of scenic spot	Degree of preference
Island of Centipede	9.43
The remotest corners of the globe	9.05
Great East China Sea	8.64
Yalong Bay	8.62
Nanshan Cultural Tourism Zone	8.37
Tropical Paradise Forest Park	7.40
Luhuitou scenic spot	7.22
Song city in Sanya	7.00
Ancient city of Ya Zhou	6.38
Kam Mu Kok	6.20
West Island	5.48
Big and small cave	5.30
Nantian hot spring in Zhuhai	5.30
Yalong Bay Shell Museum	4.58
Li Cun Miao Village	4.57
The Gulf of Sanya	3.98
Fenghuangling haishishan League scenic spot	3.75
Yulin	3.24
Butterfly Valley	-5.12

This paper also has some limitations. (1) the classification of preference content in the preference model is not automated. In the follow-up research, machine learning domain theory can be introduced into the classification of preference content to build the automatic classification model of preference content and further improve the efficiency of tourists' preference mining. (2) affective computing uses the basic rule-based method. Although the accuracy of mining meets the requirements, the expansibility of the model is insufficient. If we want to mine the preferences of other city tourists, we need to rebuild the affective computing dictionary. The follow-up research can improve the emotion calculation method and increase the expansibility of this model.

Acknowledgements

The author disclosed receipt of the following financial support for the research: This study was funded by National Natural Science Foundation of China- A Placeality-Based View of Destination Brand DNA Selection (71673015). Beijing Social Science Foundation -Research on the Mechanism of Cultural and Tourism Integration in the Construction of National Cultural Center(18JDGLA014). Multinational People's Representation and Multi-spatial Folding of Beijing: An Interpretation of Big Data, Beijing Social Science Foundation(18JDGLB013).Strategy of Beijing International Tourism Brand Based on Big Data. Beijing International Studies University Research Fund.

List of Abbreviations

There're no abbreviations.

Authors' contributions

LZ is the lead author of this paper. She made substantial contributions on the research idea, research design, data acquisition, and responded to the comments for revision. BD contributed to this paper in drafting, writing and revising manuscript. JR contributed to this paper on theory development and data analysis of this paper. LD contributed to this paper structure and direction of this paper, edited the manuscript, and responded to the comments for revision. HW and HK contributed to this paper on data cleaning. All authors read and approved the final manuscript.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Tourism Sciences, Beijing International Studies University, Chaoyang District, Beijing, China. ²Institute for Big Data Research in Tourism, Beijing international studies university, Chaoyang District, Beijing, China. ³Centre of Applied Informatics, College of Engineering and Science, Victoria University, Footscray, Australia. ⁴Tourism College of Hainan University, Meilan District, Haikou, China. ⁵College of Business and Economic, Australia National University, Canberra, Australia

References

1. Crouch G I, Louviere J J. A review of choice modeling research in tourism, hospitality, and leisure[J]. *Tourism Analysis*, 2000:67-86.
2. Louviere J J, Hensher D A. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event[J]. *Journal of Consumer Research*, 1983, 10(3):348-361.
3. Oh C O, Draper J, Dixon A W. Comparing resident and tourist preferences for public beach access and related amenities[J]. *Ocean & Coastal Management*, 2010, 53(5-6):245-251.
4. Dissanayake D, Morikawa T. Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference Nested

- Logit model: case study in Bangkok Metropolitan Region[J]. *Journal of Transport Geography*, 2010, 18(3):402-410.
5. Laing J H, Crouch G I. Myth, adventure and fantasy at the frontier: Metaphors and imagery behind an extraordinary travel experience[J]. *International Journal of Tourism Research*, 2009, 11(2): 127-141.
 6. Dong J K, Kim W G, Jin S H. A perceptual mapping of online travel agencies and preference attributes[J]. *Tourism Management*, 2007, 28(2):591-603.
 7. Xuan T, Ralston L. Tourist preferences influence of unconscious needs[J]. *Annals of Tourism Research*, 2006, 33(2):424-441.
 8. Tzukuang H, Yifan T, Wu H H. The preference analysis for tourist choice of destination: a case study of Taiwan.[J]. *Tourism Management*, 2009, 30(2):288-297.
 9. Kim S S , Timothy D J , Hwang J . Understanding Japanese tourists' shopping preferences using the Decision Tree Analysis method[J]. *Tourism Management*, 2011, 32(3):544-554.
 10. Ma Yaofeng, Zhao Hua, Wang Xiaofeng. An Empirical Study on the Tourism Preference of American Tourists in China[J]. *Journal of Northwest University: Natural Science*, 2006, 36(1): 137-140.
 11. Wu Bihu, Wu Jia, Dang Ning. Traveling City Residents' Peripheral Tourism Preference: A Case Study in Hangzhou[J]. *Human Geography*, 2007, 22(2): 27-31.
 12. Bai Kai, Ma Yaofeng. Research on Tourists' Shopping Preference Behavior——Taking Xi'an Inbound Tourists as an Example[J]. *TOURISM TRIBUNE*, 2007, 22(11): 52-57.
 13. Chen Nan, Qiao Guanghui, Liu Li. Research on Tourism Risk Perception and Tourism Preference of Outbound Tourists——Taking Beijing Tourists as an Example[J]. *Human Geography*, 2009(6):97-102.
 14. Sun Gennian. International tourism payment equation, payment grade and tourism preference [J]. *Geography and Geographic Information Science*, 2001, 17 (1): 50-54.
 15. Tian Xiangli, Song Baoping, Yan Junping. Study on Cognition and Choice Preference of Qinling Tourist Attractions in Shaanxi——Based on the Perspective of the Difference of Tourist Sources [J]. *Resources Science*, 2010, 32 (5): 873-879.
 16. Wang Sujie, Hu Ruijuan, Cheng Weihong. A Review of Tourism Research in Foreign Social Network Paradigms[J]. *TOURISM TRIBUNE*, 2009, 24(7): 90-95.
 17. Baggio R, Scott N, Cooper C. Improving tourism destination governance: a complexity science approach[J]. *Tourism Review*, 2010, 65(4): 51-60.
 18. Bramwell B, Meyer D. Power and tourism policy relations in transition[J]. *Annals of tourism research*, 2007, 34(3): 766-788.
 19. Scott N, Baggio R, Cooper C. *Network analysis and tourism: From theory to practice*[M]. Channel View Publications, 2008.
 20. Baggio R. Collaboration and cooperation in a tourism destination: a network science approach[J]. *Current Issues in Tourism*, 2011, 14(2): 183-189.
 21. Iñiguez, T., Plumed, M., & Martínez, M. P. L. (2014). Ryanair and Spain: Air connectivity and tourism from the perspective of complex networks. *Tourism & Management Studies*, 10(1), 46-52.
 22. Zhong L, Sun S, Law R. Movement patterns of tourists[J]. *Tourism Management*, 2019, 75: 318-322.

23. David-Negre T, Almedida-Santana A, Hernández J M, et al. Understanding European tourists' use of e-tourism platforms. Analysis of networks[J]. Information Technology & Tourism, 2018, 20(1-4): 131-152.
24. Zhang L, Zhang J. Perception of small tourism enterprises in Lao PDR regarding social sustainability under the influence of social network[J]. Tourism Management, 2018, 69: 109-120.
25. Baggio, R., & Cooper, C. (2010). Knowledge transfer in a tourism destination: the effects of a network structure. The Service Industries Journal, 30(10), 1757-1771.
26. Skokic V, Lynch P, Morrison A. Veza: An informal network of tourism entrepreneurs[J]. Annals of Tourism Research, 2019, 77: 26-37.
27. Raisi H, Baggio R, Barratt-Pugh L, et al. Hyperlink network analysis of a tourism destination[J]. Journal of Travel Research, 2018, 57(5): 671-686.
28. Tran M T T, Jeeva A S, Pourabedin Z. Social network analysis in tourism services distribution channels[J]. Tourism Management Perspectives, 2016, 18: 59-67.
29. Lina Zhong. Reconstruction of Perception Structure of Tourism Destinations——Based on Text and Complex Network Analysis[J]. Tourism Tribune, 2015, 30(8).
30. Zhong Lina, Wu Bihu, Xu Xiaobo, et al. A Review of Foreign Tourism Destination Perception Studies [J]. Human Geography, 2013, 28 (2): 13-19.

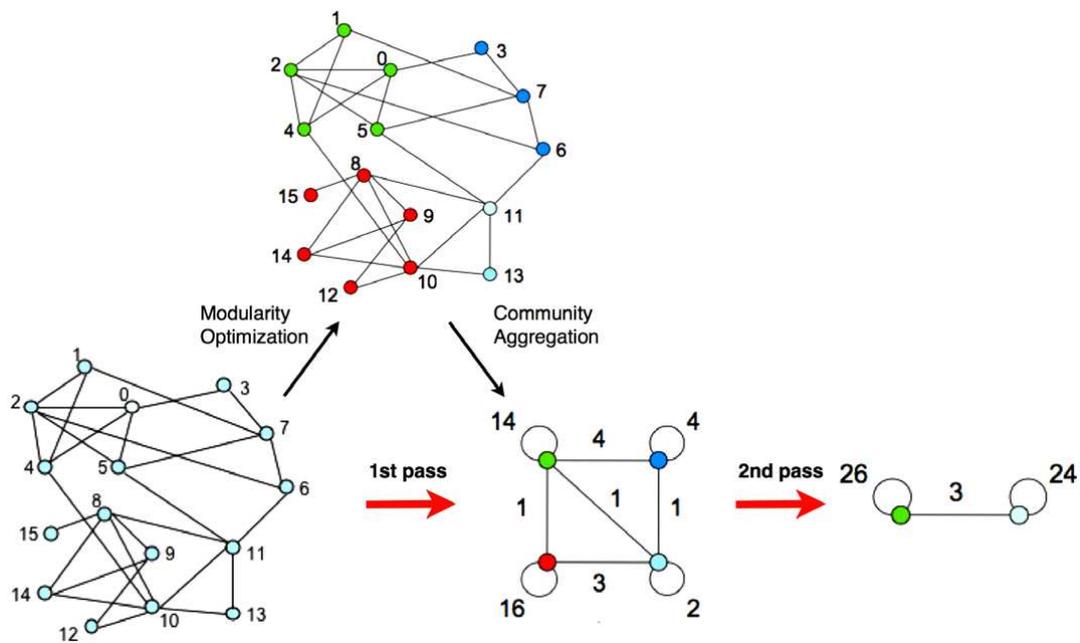


Fig. 1 Louvain community mining algorithm flow chart

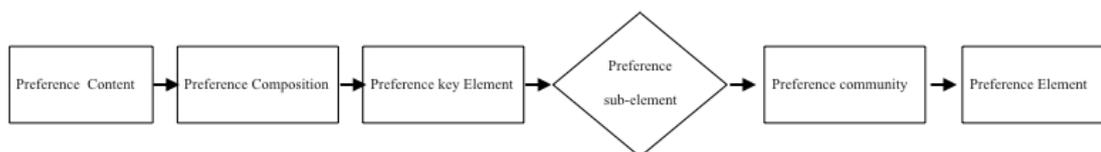


Fig. 2 Preference content classification system

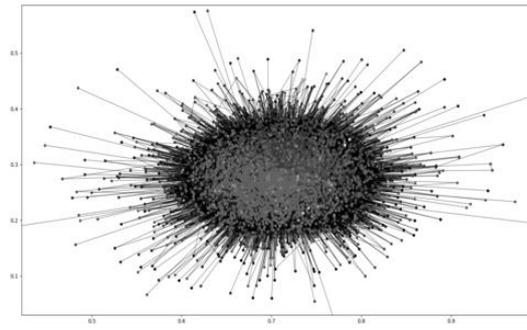


Fig.4 Preference complex network community mining results

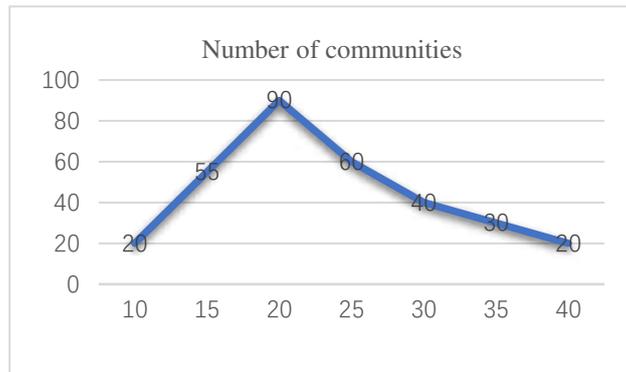


Fig. 3 Optimal pruning threshold map