Research on Evaluation of Popularity of Lijiang Scenic Area Based on Microblog Data

Bing-yang CHEN1,2, Kun YANG1,2, Jia-sheng WANG1,2,* and Meng-zhu SUN2,3

1School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
2The Engineering Research Center of GIS Technology in Western China Ministry of Education of China, Kunming 650500, China
3School of Tourism and Geographical Science, Yunnan Normal University, Kunming 650500, China

*Corresponding author

Keywords: Microblog Check-in Data, Tourist emotion, Spatial-temporal characteristics, Lijiang.

Abstract. The rapid development of LBS services and big data, massive check-in data provide a basis for the study of the popularity of scenic spots. Taking Lijiang scenic area as the research object, this paper made use of microblog check-in data, and made an evaluation on the popularity of tourist attractions by analyzing the spatial-temporal characteristics of tourist emotion and behavior. Emotion analysis, statistical analysis, spatial analysis and other methods were used to explore the spatial-temporal characteristics of tourist emotion distribution. The study finds that foreign tourists are the main force of tourism activities, in which female tourists are more emotional, and the young and middle-aged tourists are the most active. The emotional value of tourist in tourism activities is positive, give priority to with 43.82% of positive emotions. Scenic spots such as Old Town of Lijiang, Lugu Lake, Yulong Snow Mountain and other places of high popularity, while the Baisha Mural, Shigu Town attractions such as low popularity. The spatial distribution of tourist emotion is consistent with the development of key tourist areas in Lijiang.

Introduction

With the development of the Web2.0 era, more and more people like to release emotions through social platforms. Due to the huge amount of data, accurate, easy access to data and other characteristics, Sina microblog get the attention of researchers from all walks of life. Users through microblog check-in in the expression of their emotions but also retain the spatial-temporal information. A large number of spatial-temporal footmark to form the tourists' "information map"[1], it provides a valuable resource for the study of the emotional and behavioral characteristics of users, and provides a new way for the study of the popular degree of scenic spots.

With the improvement of people's living standard, people's demand is increasing. Tourists in tourism activities no longer meet the simple sightseeing, but the pursuit of a higher level of emotional experience and spiritual satisfaction. Therefore, the analysis of tourist emotion has become an important topic in the field of tourism research [2]. Emotion analysis mainly uses two methods: emotional dictionary combined with rules [3-5] and machine learning [6,7]. The difference between the microblog text and the traditional text is large, which is mainly characterized by the theme divergent, the omission of the subject and predicate and object, the omission of the context, the serious colloquial language and the text contains links [8]. With the increasing variety of data sources and the maturity of data processing methods, spatial-temporal data analysis has attracted more and more attention in the field of urban geography [9-11]. Zhang Ziang et al [12] have used the "horizontal" and "vertical" time stratification method to analyze the spatial-temporal characteristics of the tourist activities in Nanjing Zhongshan Mountain scenic area by dividing the gender and regional attributes. The research on the time distribution pattern of domestic tourists in Xi'an based on microblog check-in data by Tang Jia et al [13]. However, the study on the spatial-temporal characteristics of tourist behavior by using the social network data mostly ignores
the tourists' age attributes and the analysis of the tourists' emotions. Therefore, this study used tourists in Lijiang scenic spot a year of microblog check-in data. The analysis of tourist emotion by using the method of emotional dictionary combined with rules. Through the number of check-in to the scenic spot of hot spot analysis, while the introduction of the age attribute, statistical analysis of tourists change characteristics of emotion in different time scales. In addition, the spatial distribution of different types of tourists’ emotions was explored, and the temporal and spatial distribution characteristics of tourists in Lijiang were summarized. And then provide a mirror for tourists to select scenic spots, provide a reference for the management of scenic spots, to provide a basis for the public choose commercial and residential location.

Data Acquisition and Preprocessing

Data Acquisition

In this paper, Sina microblog user check-in information as the basis. Lijiang City, a famous tourist city with plateau lakes, snow mountain glaciers and stone forest caves, was selected as the study area. It chose 35 famous scenic spots and recorded their latitude and longitude, through calling Sina microblog open API interface, with 4km as the radius of the search. In order to ensure that all attractions were covered, it set the center point of the search radius 1.5km in attraction concentration areas, and selected from November 1, 2015 to October 31, 2016 a year time span, finally received a total of 253624 microblog data as the data source. And it collected the user's ID, gender, location (attribution), birthday, check-in spot, check-in time, microblog text and other 7 data. The statistics of relevant data are shown in table 1.

<table>
<thead>
<tr>
<th>Scenic Spots</th>
<th>Number</th>
<th>Scenic Spots</th>
<th>Number</th>
<th>Scenic Spots</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Town of Lijiang</td>
<td>25672</td>
<td>Caohai</td>
<td>5715</td>
<td>Laojun Mountain</td>
<td>2276</td>
</tr>
<tr>
<td>Lugu Lake</td>
<td>23428</td>
<td>Lijiang Songcheng</td>
<td>5185</td>
<td>Baoshan Stone City</td>
<td>2082</td>
</tr>
<tr>
<td>Yulong Snow Mountain</td>
<td>18634</td>
<td>Heilongtan Park</td>
<td>4613</td>
<td>Yufeng Temple</td>
<td>1894</td>
</tr>
<tr>
<td>Dashuihe</td>
<td>17419</td>
<td>Jade water village</td>
<td>3879</td>
<td>Guanyin Gorge</td>
<td>1802</td>
</tr>
<tr>
<td>Rigby Peninsula</td>
<td>17071</td>
<td>Spruce Meadow</td>
<td>3455</td>
<td>Ganhaizi</td>
<td>1798</td>
</tr>
<tr>
<td>Square Street</td>
<td>13766</td>
<td>Nisai Village</td>
<td>3403</td>
<td>Wenhai</td>
<td>1682</td>
</tr>
<tr>
<td>Blue Moon Valley</td>
<td>10907</td>
<td>Qingxi Reservoir</td>
<td>2812</td>
<td>Xuesong Village</td>
<td>1648</td>
</tr>
<tr>
<td>Shuhe Old Town</td>
<td>10537</td>
<td>Nvshenwan</td>
<td>2757</td>
<td>Dongba Valley</td>
<td>1542</td>
</tr>
<tr>
<td>Hutiao Gorge</td>
<td>10004</td>
<td>Glacier Park</td>
<td>2671</td>
<td>Wenbi Mountain</td>
<td>1412</td>
</tr>
<tr>
<td>Lashi Lake Wetland Park</td>
<td>8629</td>
<td>Wangulou</td>
<td>2524</td>
<td>Baisha Mural</td>
<td>1353</td>
</tr>
<tr>
<td>Daluoshui Village</td>
<td>8585</td>
<td>Baisha Old Town</td>
<td>3203</td>
<td>Shigu Town</td>
<td>1301</td>
</tr>
<tr>
<td>Mu Fu</td>
<td>8516</td>
<td>Liwubi Island</td>
<td>3198</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data Preprocessing

Due to the format of the problem, all the data are concentrated in the same column. The first is dividing data into columns. And then removing the interference: (a) In order to distinguish between tourists and ordinary users check-in data, remove the enterprise companies, schools, banks and other places where tourists check-in less likely; (b) Delete the user microblog content includes China Mobile, China Telecom, China Unicom, overtime, purchasing, promotion, recruitment, teachers, libraries, work, go home and other main body of the release of data for non -tourists. (c) Reference to the method used by Girardin[14], if the same user's first microblog and the last release of the microblog time interval of more than 30 days, it is likely to be local residents. Lastly, the user data for the location is not clear and the check-in point on the boundary of the study area is eliminated. The user check-in data for the location is not clear and the check-in point on the study area boundary is eliminated. Finally, get the 232813 check-in data.
Research Methods

The methods used in this study include emotion analysis, statistical analysis and spatial analysis, specific methods are as follows:

Emotion Analysis

Firstly, build the emotional dictionary. Make use of existing resources MBEWC, “HowNet” Dictionary (HND), Sogou Lab provides the Internet Thesaurus (SgT), the therein positive and negative emotion words are extracted, as well as merge the three parts and remove duplicates, get the emotional dictionary (ED). It can be expressed as below

\[ ED = MBEWC \cup HND \cup SgT . \]  

(1)

The entries in the emotional dictionary are composed of three tuple: emotional words (ew), part of speech (pos) and polarity value (pv):

\[ ED = \{ < ew_1, pos_1, pv_1 >, < ew_2, pos_2, pv_2 >, ..., < ew_n, pos_n, pv_n > \} \]

(2)

Then calculate the feature weight. Feature weights are used to measure the degree of importance of a feature item in the document and the strength of the differentiating ability. Lexical spatial vector model is set up aim at vocabulary sequences. Using the features of microblog word to represent each microblog \( M_j \), the features and weights of these words form the vector in the space.

\[ \left( W_{i,j}, W_{2,j}, ..., W_{n,j} \right) \]

(3)

With \( W_{i,j} \) is the weight of entry \( i \) in \( M_j \) [15], currently the most commonly used is the TF-IDF method:

\[ W_{i,j} = TF_{i,j} \times \log \frac{N}{n_i} \]

(4)

With \( TF_{i,j} \) indicates the number of feature items \( m_i \) appears in microblog \( M_i \), \( n_i \) represents the number of documents for a feature item, \( N \) is the number of all text, get feature weight calculation method by normalization:

\[ W_{i,j} = \frac{TF_{i,j} \times \log \frac{N}{n_i}}{\sqrt{\sum_{m_i \in M_j} \left( TF_{i,j} \times \log \frac{N}{n_i} \right)^2}} \]

(5)

Then, combine the chi-square test and hierarchical clustering algorithm are used to reduce the dimension. Using chi-square test select feature to preliminary dimension reduction, and then hierarchical clustering is used to reduce dimension. In order to improve the clustering effect of vector space (vs), the characteristic items are divided into positive (Pvs), negative (Nvs) and neutral (Mvs) three subspaces, and then merge together [16]. Assume that \( \text{Cluster}(X) \) is cluster vector space \( X \), it can be expressed as:

\[ \text{Cluster}(vs) = \text{Merge}\left( \text{Cluster}(Pvs), \text{Cluster}(Nvs), \text{Cluster}(Mvs) \right) \]

(6)

Assume the class cluster \( i \) is represented by \( c_i \), the average similarity between clusters is:

\[ \text{sim}(c_i, c_j) = \frac{\sum_{W_{i \in c_i}} \sum_{W_{j \in c_j}} \text{sim}(W_i, W_j)}{|c_i| \times |c_j|} \]

(7)

With \( W_{i} \) and \( W_{2} \) respectively are feature items of class cluster \( c_i \) and \( c_j \); \( \text{sim}(W_i, W_j) \) is the semantic similarity of \( W_i \) and \( W_j \) [17], \( c_i \) is the number of feature items in class cluster \( i \). Finally,
calculate the emotion value and analyze the tourist emotion.

**Statistical Analysis**

According to gender, region two tourists attributes, making the check-in data table from the month, day and age and other scales. Multiple attribute cross analysis the differences of tourists in the scenic area activities and feelings of tourists in the variation characteristics of different scales.

**Spatial Analysis**

Tourists will have different emotions in different regions, and the changes of their emotional values reflect the spatial differences of the keen areas of tourists when they express their emotions. Therefore, calculate the emotional value of tourists in each scenic spot at first, according to the method of equal number put attractions popularity degree into 5 levels. Secondly, with the help of tourists in scenic spots left the geographical coordinates, visual analysis of the distribution of tourist emotion in space through the ArcGIS10.2. From the positive, neutral and negative feelings of the three tourists to explore the different emotion types of tourists in the spatial distribution of differences. And then summarizes the characteristics of Lijiang tourists’ emotional space distribution.

**Results and Discussion**

**Tourist Emotion Analysis**

Assign each word according to the intensity of the emotional content of each word. The range of emotional words and modal words are [-10, 10], the scope of adverbs of degree is [0.8, 1.4], and the negative word is 1. The result of the emotional analysis of the tourists can be got after a series of calculations, as shown in table 2.

The limit of each microblog word is 140 words, and the value of a single word is [-10, 10], so the value of the emotional value of microblog [-1400, 1400]. As can be seen from the table, tourists in Lijiang scenic area are mainly with positive emotions. It can also be seen that tourists in the scenic area when the mood is more gentle, less extreme emotions, in line with the general expression habits.

<table>
<thead>
<tr>
<th>Summary Situation</th>
<th>Name and Emotion Range</th>
<th>Quantity</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>General situation</td>
<td>Positive emotion(0,1400]</td>
<td>102018</td>
<td>43.82%</td>
</tr>
<tr>
<td></td>
<td>Neutral emotion 0</td>
<td>89074</td>
<td>38.26%</td>
</tr>
<tr>
<td></td>
<td>Negative emotions [-1400,0)</td>
<td>41720</td>
<td>17.92%</td>
</tr>
<tr>
<td>Segmented statistical</td>
<td>Low(0,30]</td>
<td>68936</td>
<td>29.61%</td>
</tr>
<tr>
<td>results of positive emotions</td>
<td>Medium(30,100]</td>
<td>23351</td>
<td>10.03%</td>
</tr>
<tr>
<td></td>
<td>High(100,1400]</td>
<td>9732</td>
<td>4.18%</td>
</tr>
<tr>
<td>Segmented statistical</td>
<td>Low [-30,0)</td>
<td>29171</td>
<td>12.53%</td>
</tr>
<tr>
<td>results of negative emotions</td>
<td>Medium [-100,-30)</td>
<td>7543</td>
<td>3.24%</td>
</tr>
<tr>
<td></td>
<td>High [-1400,-100)</td>
<td>5005</td>
<td>2.15%</td>
</tr>
</tbody>
</table>

**Analysis of Tourists' Emotional Time Characteristic**

Form throughout the year with April, October as the spire of two "pyramid" structure can be got by Statistical analysis. The reason is due to the small holiday and climate factors the impact, tourists travel emotion is higher. As can be seen from Figure 1 in any month, period and any age range, the number of female tourists are more than the number of male tourists. Women may be more active in the tourism activities. From the details can be found, the degree of female activity is higher in the tourist season (Figure 1(a)). At the same time, it can be seen from the scale of the day that the check-in of female tourist is more intensive during the 10-18 rush hour (Figure 1(b)). This shows that female tourists are more adept at expressing their feelings. From Figure 1 (c) found that in the 50-70 age range, the number of female check-in is far more than men. Perhaps older women better grasp the use of social platforms such as microblog, more pursuit of the pace of technological era.
The results of the analysis of regional attributes from the perspective of time are shown in Figure 2. It is easy to find that foreign tourists are far more than local tourists. This phenomenon also shows the need to increase the attraction of local tourists. Daily start with scenic spot open in 6 to 10 time intervals, the number of check in began to increase. It reaches the check-in peak in the 10 to 16 time intervals, tourism activities are the most frequent. In addition to the discovery of foreign tourists in the tourist season is more active (Figure 2(a)). Indicating that the weather is warming, tourists travel enthusiasm is higher. And at this time more holidays, tourists will have plans to visit the scenic spots earlier. The entire tourism activities mainly by 20 to 50 age range of foreign tourists’ activities give priority to, accounting for 67.94% of the total number of tourists. Because the young and middle-aged tourists better physical strength, but also more adept at using scientific and technological means for the memorial, and the elderly and children in tourism activities need more care.

According to emotional value with the method of equal number to put attractions popularity degree into 5 levels, the classification results are shown in Table 3. It can be seen that although the same scenic spot will be mixed with three kinds of emotional state at the same time, but the emotional value is always positive, the positive emotion of microblog density is always higher than the same level of negative emotion. Tourists in the tourism activities will inevitably produce negative emotion because of financial loss, traffic congestion, poor infrastructure, plateau reaction and ultraviolet burn. But this form of tourism by tourists leave their permanent residence to a strange city life, enjoy the release of the pressure of work, showing visitors are more positive emotions in the tourism activities.
It can be found from Figure 3 that the distribution of positive emotions in space is similar to that of microblog check-in in Lijiang. The most concentrated positive emotions for the Old Town of Lijiang, Lugu Lake, Yulong Snow Mountain and other attractions of the check-in largest density, while the positive emotions are more sparse for the minimum check-in density of Baisha Murals, Shigu Town and other attractions. The negative emotion and neutral emotion of tourists are also concentrated in Old Town of Lijiang, Lugu Lake, Yulong Snow Mountain and other attractions. It shows that the three kinds of emotion have similar distribution in space, and the emotional intensive area is the most popular scenic spot. In addition, it is found that Old Town of Lijiang, Lugu Lake, Yulong Snow Mountain and other scenic spots are important areas for the development of tourism in Lijiang. It can be concluded that the spatial distribution of tourist emotion is consistent with the development of key tourist areas in Lijiang.

Figure 3. Emotion space distribution of Lijiang tourists.

Conclusion

This study inspired by the big data ideas and the study of tourists’ feelings, visualize microblog check-in data with the help of ArcGIS software, analysis and summary of emotional spatial-temporal distribution inside the Lijiang scenic spot of tourists, the main conclusions are as follows.

In the time level, the formation of two "pyramid" structure with April, October as the spire of, at this time the enthusiasm of foreign tourists and local tourists more obvious gap. Female tourists’ feelings are more abundant, in the tourist season and the peak period of travel time is more significant. The tourist group is given priority to with young and middle-aged. Analysis shows that foreign tourists are more willing to choose a small holiday with the appropriate temperature to play. Female is good at through microblog share their feelings because of exquisite emotion. Restricted by physical strength and physical conditions, young and middle-aged tourists are more suitable for high altitude, strong ultraviolet light to carry out tourism activities.
In the space level, the tourists' positive, negative, neutral three emotions are mainly concentrated in Old Town of Lijiang, Lugu Lake, Yulong Snow Mountain and other attractions, which accounted for a larger proportion of positive emotions. In addition, these attractions are also Lijiang important areas of tourism development. It can be concluded that the tourists of the same scenic spot may be mixed with three kinds of emotional states, but they are mainly positive emotion. The spatial distribution of tourists' emotion shows that the gathering place is consistent with the key tourism development areas in Lijiang.

In the aspect of spatial-temporal, there are some differences in the spatial-temporal variation of tourist emotion. Old Town of Lijiang, Lugu Lake, Yulong Snow Mountain and other tourist attractions for the largest differences in emotion, is positive emotion, negative emotion, "high" area, but also the most popular attractions. Baisha Murals, Shigu Town and other attractions tourists relatively low emotion. The tourists' feelings of each scenic spot are related to tourism facilities, services and famous degree of scenic spots and so on, managers can improve the inadequacies rely on these.

This paper analyzes the popularity of scenic spots from two aspects, tourists' emotions and tourist spatial-temporal behavior characteristics with the help of microblog location data, hoping to provide a reference for the tourist selection scenic spots, scenic management planning and the public selection in the commercial and residential location. In the next study, we can try to expand the research area to analyze whether the province's key tourism development area agree with the tourists' emotional spatial-temporal distribution. In addition, the research methods of tourists' emotion should be further studied.

Acknowledgement
This research was financially supported by the National Natural Science Foundation of China (41501436).

References


