Research on Spatial-Temporal Behavior Characteristics of Tourists Based on Digital Footprints: A Case Study of Guilin, China

Yao Jiansheng1,2, Liu Yanling1,2*
1 College of Tourism & Landscaper Architecture, GuiLin University of Technology, Guilin,541004, China
2 Institute of Guangxi Tourism Industry, Guilin, 541004, China
*Corresponding author’s e-mail:2018014@glut.edu.cn

Abstract. Taking Guilin as an example, this paper constructs a digital footprint database based on online travel notes, and analyzes the temporal and spatial behavior characteristics of Guilin tourists from the aspects of time, space and network structure by using the methods of season concentration index, tourism preference index, nearest neighbor index and visualization technology. The results show that: (1) there is a certain seasonality in the tourist time, and the stay days are mainly 3-4 days; (2) the spatial distribution of Guilin scenic spots is unbalanced, and the visiting preference is uneven. The popular scenic spots include Li River, Yulong River, West Street, Two Rivers and Four Lakes, and Longji Terrace. Basis on this, this paper discusses the image promotion, path planning and marketing of the scenic spots Finally, the paper puts forward the optimization measures of Guilin tourism development.

1. Introduction
Tourists are not only the main body of tourism activities, but also the service object of tourism industry. Only by mastering the behavior characteristics of tourists and understanding their internal mechanism, can we strengthen the construction of tourism destination and improve the level of tourism service [1]. Tourists' spatiotemporal behavior (mobility) is the most important embodiment of tourists in tourism activities. The study of tourists' spatiotemporal behavior can understand tourists' tourism activities better and it is of great significance to the analysis and prediction of tourism market.

American scholar Girardin et al. [2] put forward the concept of "digital footprint" for the first time, and pointed out that "tourism digital footprint" is the information and call records sent by tourists during the tour, and the text and pictures published on the Internet after the tour. Li Junyi [3] studied the connotation and characteristics of tourism digital footprint earlier in China, and pointed out that tourism digital footprint is the electronic traces with positioning information or geographical labels left by tourists on the Internet, communication network, Internet of things and other information facilities in tourism activities, and these electronic traces can reflect the temporal and spatial movement trajectory of tourists.

In recent years, researchers have used different data sources, such as GPS [4], mobile phone data [5], geotagged photos [6-7], travel notes [8-10], travel notes + Photos [11], and Internet reviews [12], to study the mobile model of tourists [5], Internet image perception [6], and spatiotemporal behavior characteristics [9, 11] and tourist satisfaction [12]. Guilin, as an international tourist attraction, has few relevant studies, which is not in line with its status as a famous tourist city. Therefore, this paper takes
Guilin city and its subordinate cities and counties as the case, and takes online travel notes as the data source to study the temporal and spatial behavior characteristics of tourists to Guilin, so as to provide scientific basis for tourism planning, image promotion and marketing of Guilin.

2. Digital footprint database case

2.1. Case location
Guilin has enjoyed the reputation of "the best landscape in the world" since ancient times. It is an international tourist resort approved by the state, an international tourism city open to the outside world, a national tourism innovation and development pilot area, and a demonstration area for China to become a world-class tourism destination. It is also the permanent host of the International Forum on tourism trends and prospects of the United Nations World Tourism Organization / Asia Pacific Tourism Association, and one of the world tourism organizations recommended by the world tourism organization One of the best tourist cities in China. In 2019, Guilin received 138.3366 million domestic and foreign tourists and realized a total tourism consumption of 187.425 billion yuan. Tourism is the pillar industry of Guilin, so it is of great significance to study the tourism demand of Guilin.

As of 2019, Guilin has 6 districts, 10 counties and 1 county-level city under its jurisdiction, with rich tourism resources, including 4 AAAAA scenic spots, 37 AAAA scenic spots, 29 AAA scenic spots and many non-a-level well-known scenic spots, such as Impression Liu Sanjie, Guilin Eternal love, Tianhu eco-tourism resort, SanQianLi Landscape cultural Resort, etc.

2.2. Digital footprint database
This paper chooses Ctrip platform as the data source of this study. Ctrip is the largest online travel service provider in China, with a large number of registered users. The online travel information is relatively complete and updated quickly, so it is suitable to be used as a sample source library. Therefore, the web crawler technology based on Python collects 5840 travel notes samples from January 1, 2014 to December 31, 2018, and uses the pandas Library of Python language to clean and preprocess the travel notes.

Firstly, the data source is cleaned and the travel notes are filtered according to the standard: ① with tour paths; ② with tour days, and the tour days are less than 20 days; ③ the main tourism is Guilin, that is, a large number of records of non Guilin scenic spots in the tour paths are deleted. Finally, 2057 valid samples were obtained.

| Unified name      | Merge name                                                                 |
|-------------------|---------------------------------------------------------------------------|
| Li River          | Yangshuo Li River, X- Li River Guilin section, Guilin Li River, Li River   |
|                   | cruise boat, Li fishing fire, Li River Essence Tour, nine horse painting   |
|                   | mountain, yellow cloth reflection, etc.                                   |
| Yulong River      | Yulong River drifting, Yangshuo, Yulong River, Yulong bridge              |
| West St           | Yangshuo West Street, Laoxi street, Yangshuo County West Street, Guilin    |
|                   | Yangshuo West Street, Ginzhuang village, Longji Ginzhuang village, Longji  |
|                   | Ginzhuang village terraces, Ginzhuang village, Dazhai, Longji Dazhai,     |
|                   | Dayao village, Ping'an Village, Ping'an Zhuang village, Ping'an Zhuang      |
|                   | village, Longji Ping'an, Huanghuo Yao village, the world's first Changfa    |
|                   | village, Qixing companion moon, Jinkeng terraces, Ping'an Village,         |
|                   | Huangluo Yao village, Jinzhu Zhuang Village, Xiangshan scenic area,        |
|                   | Xiangbei mountain, Xiangshan, Xiangbei mountain, rubber mountain in Guilin |
| Jingjiang Palace  | duxiufeng palace scenic area, Jingjiang palace, duxiufeng, duxiufeng      |
| Seven Star scenic | Seven Star Rock, Seven Star Park, x-guilin seven star rock, seven star rock |
| spot              | archway square, Guilin Seven Star Park                                    |

Then the data is preprocessed and the digital footprint database is constructed, ① 2057 tourist routes are extracted, with a total length of 19254 scenic spots, with an average length of 9.4, including 985 different scenic spots after de duplication; ② different names of the same scenic spot and different points
of interest (POI) in the same scenic spot are merged, as shown in Table 1; ③ non Guilin scenic spots are deleted (many tourist route records are very detailed, starting from the starting point to the destination) ④ count the number of times that the scenic spots appear in the tour path as the tour frequency of the scenic spot; ⑤ use Python language to call Gaode map API to obtain the longitude and latitude of the scenic spot. Guilin and Yangshuo appear in a large number in the whole tour route, but they are tourist destinations, including many scenic spots. In the statistical analysis, Guilin and Yangshuo are removed, and 79 scenic spots in Guilin are obtained. The above information is stored in SQLite database to build a digital footprint spatiotemporal database.

3. Research methods
(1) Seasonal concentration index (I)
Seasonal concentration index, also known as seasonal time intensity index (seasonal intensity index), can quantitatively analyze the seasonal concentration degree of online travel notes, so as to indirectly analyze the seasonal concentration degree of Guilin tourism:

\[ I = \frac{\sum_{i=1}^{12} (x_i - \bar{x})^2}{12} \]

Where: \( \bar{x} \) is the average value of the proportion of the number of Online Travel Notes published in each month to the total number of Online Travel Notes published in the whole year. Assuming that it follows the average distribution, then \( \bar{x} = 1/12 \approx 0.833 \). \( x_i \) is the proportion of the number of Online Travel Notes published in each month to the total attention of the whole year. The larger the index value is, the higher the seasonal concentration degree and the greater the seasonal difference of the amount of online travel notes is; on the contrary, the more the index value tends to 0, the lower the seasonal concentration degree and the more uniform the seasonal distribution of the amount of online travel notes is.

(2) Monthly change index
The change index is equivalent to standardizing the amount of online travel notes, which also reflects the change degree of the amount of online travel notes in the statistical period, but can better observe the change of data:

\[ S_i = \frac{x_i}{\frac{1}{12} \sum_{i=1}^{12} x_i} \times 100\% \]

Where: \( S_i \) (0≤\( S_i \)≤1) is the change index of the number of Online Travel Notes published in the i-th month, and \( x_i \) is the number of Online Travel Notes published in the i-th month of the year. The closer the value \( S_i \) is to the benchmark value of 100%, the more uniform the monthly distribution of network attention is and the weaker the seasonality is; on the contrary, the stronger the seasonality of network attention is.

(3) Tourism preference index
There are many ways to quantitatively analyze tourism preference. This paper uses the tourism preference index proposed by Sun Gennian to analyze tourists’ preference for Guilin scenic spots:

\[ P = \frac{x_i}{X} \times 100\% \]

Where: \( x_i \) is the number of times that the ith scenic spot appears in the tour path; \( X \) is the total number of times that all scenic spots appear in the tour path; the value range of \( P \) is between 0-1, and the closer the value is to 1, the stronger the tourist preference for the scenic spot.

(4) Nearest neighbor index
There are three types of spatial distribution of point elements: aggregation, uniformity and randomness. The point elements with uniform distribution have the largest nearest distance, followed by random distribution, and the aggregation distribution is the smallest. The nearest neighbor distance is a geographical indicator of the degree of mutual proximity of point objects in geographical space,
which can well reflect the spatial distribution characteristics of point elements. The calculation method is the ratio of the actual nearest neighbor distance and the theoretical nearest neighbor distance (i.e. the theoretical value of random distribution):

\[
R = \frac{r}{r_E} = 2\sqrt{D \times \frac{1}{A}}
\]

(4)

\[
\frac{r_E}{2} = \frac{1}{\sqrt[4]{\pi A}} = \frac{1}{2\sqrt{D}}
\]

(5)

Where: \( R \) is the nearest neighbor index, \( r \) is the average distance between the nearest points, \( r_E \) is the theoretical nearest neighbor distance, \( D \) is the point density, \( A \) is the area, and \( n \) is the number of research objects. When \( R = 1 \), the point elements distribute randomly; when \( R > 1 \), they distribute uniformly; when \( R < 1 \), they distribute coherently.

4. Experiment and result analysis

4.1. Time analysis

According to the statistics of the average number of papers published in each month from 2014 to 2018 in the digital footprint database, the seasonal concentration index is 0.012, and the inter monthly variation index is 0.065 to 0.095, which is not obvious. From the numerical point of view, Guilin tourism does not show seasonal concentration. However, when the inter month change index of the number of travel notes is visualized (Figure 1), it can be observed that Guilin's tourism is seasonal. The peak season is from April to October, and the off-season is from January to March and from November to December, which is consistent with the actual situation of Guilin's tourism. This is because the difference between the highest value and the lowest value of the inter monthly variation index is only 0.03 (0.095-0.063), so there is no seasonality in the numerical value, and the visualization technology partially magnifies this interval of the y-axis, and then small differences are observed. Therefore, there is a weak positive correlation between the number of Guilin travel notes and the number of Guilin tourists.

Figure 2 shows the statistics and visualization of tourist days and distribution density in the digital footprint database. The number of days of tourists to Guangxi is power-law distribution, in which the main stay is 3-4 days, followed by 5 days, 1 day, 2 days, 6 days and 7 days. Other stay days account for a relatively small proportion, with an average stay of 4.1 days. The proportion of overnight stay in Guilin is relatively large, and the proportion of tourists staying more than 7 days is relatively small, which indicates that the deep tourism in Guilin is in the primary stage, and the short-term sightseeing tourism and leisure tourism account for a relatively large proportion in the overall tourism of Guilin.

Fig 1. Monthly change index of the number of travel notes                    Fig 2. Distribution of tour days
4.2. Spatial analysis

4.2.1 Spatial distribution

Extracted 79 Guilin scenic spots, longitude and latitude information and geometric data of Guilin City area from digital footprint database, and calculated the nearest neighbor index of Guilin scenic spots as 0.3778 < 1, indicating the aggregation characteristics of Guilin scenic spots. Figure 3 the distribution map of tourist attractions in Guilin city is visualized based on Gaode map. Guilin scenic spots are long and narrow in North and South, mainly along the river, starting from Ziyuanbajiao Village scenic spot in the north, Liwan scenic spot in Lipu in the south, Longji terrace scenic spot in the East, and Ling River and Guilin Merryland Resort in the West.

The distribution of scenic spots is uneven, showing a spatial distribution pattern of "large dispersion and small aggregation". The main gathering areas are in Guilin city and Yangshuo, and on both sides of Li River and Guiyang highway from the city to Yangshuo. Lingchuan, Longsheng, Ziyuan, Xing'an and Lipu are scattered. The scenic spots in Yongfu, Guanyang, Quanzhou, Gongcheng and Pingle are on the track of travel notes There is no specific statistics.

4.2.2 Preference of scenic spots

According to the number of times that scenic spots appear in the path of the tour track in the digital footprint database, as the simulation value of the number of times that scenic spots are visited, according to the formula (3), calculate the scenic spot visit preference index and the cumulative value of scenic spot visit preference (Figure 5). From the cumulative proportion, 21.52% of the popular scenic spots accounted for 80.58% of the visits. It can be seen that the preference of Guilin scenic spots basically conforms to the law of 28 (Pareto Law). It can be observed directly from the contour map that the visiting heat from north to south is concentrated in Longsheng, Guilin City, Li River and Yangshuo respectively. The top five hot spots are Li River, Yulong River, West Street, Longji terrace and two rivers and four lakes, which are the cream of Guilin tourism. "Guilin's landscape is the best in the world, and Yangshuo's landscape is the best in Guilin". Li river connects Guilin and Yangshuo, which is a typical representative of Guilin's landscape. Yulong River is known as Xiaoli River, which is famous for its idyllic scenery, forming a huge contrast with the city's fast-paced life. In recent years, the popularity of sightseeing has increased year by year. Compared with the natural scenery of Yulong River, West Street is a historical and cultural block with a history of more than 1400 years It is one of the most important cultural tourist attractions in Yangshuo. Longsheng terraces, with a history of at least 2300 years, is an important agricultural cultural heritage in the world. It can be called the hometown of terraces in the world. It is a perfect fusion of national customs and natural scenery. The water system of two rivers and four lakes around the city reproduces the prosperity of Guilin, known as "Oriental Venice".

5. Conclusion

(1) From the time distribution characteristics: because Li River may have a dry season in winter, there is a weak seasonal concentration of tourists to Guangxi; limited by time and cost, tourists stay mainly for 3-4 days, which is relatively short, and can not achieve deep tourism.

(2) From the perspective of spatial distribution characteristics, the spatial distribution of Guilin scenic spots is uneven. Geographically, Guilin scenic area has three strong connected communities, including the urban scenery belt with two rivers and four lakes as the core, the Riverside scenery belt with Li River as the core, and the Yangshuo scenery area with Yulong River as the core. Most tourists only visit the core scenic spots, resulting in a relatively low visit rate of scenic spots in the surrounding counties and cities.

(3) Policy recommendations: Guilin municipal government is suggested to solve the water problem of Li River by building reservoirs and other water conservancy projects; to create winter tourism projects for Northern tourists by making use of the climate and ethnic cultural differences between the north and the South; to use new media marketing strategies to package non-core scenic spots and create a new "net
red” punch in place, so as to make Guilin tourism develop to multi-core and enhance the overall market competitiveness.

(4) Research limitations. First, the data source is single, only choose Ctrip; second, the data type is unitary, only extract space-time information from the text; third, only one case is used in this paper. In the future, we will consider integrating multi-source heterogeneous data of multiple OTAs on mobile and PC terminals, integrating multiple data types such as pictures and videos, selecting different types of tourism destinations, and deeply studying GIS spatio-temporal data mining technology, so as to analyze and summarize the basic rules of tourists' behavior and guide the development of regional tourism in a universal way.

Fig3. POI distribution of Scenic spots in Guilin

Fig4. Contour of visit preference of scenic spots
Acknowledgments
The research reported in this paper is supported by the Guilin Scientific Research and technology Development Project No. 20180102-2, the Guangxi Young and Middle-aged Teachers Basic Ability Enhancement Project No. 2019KY0292, and the Research Foundation of Guilin University of Technology No. GUTQDJJ2017141, and Key research and development plan of Guilin scientific research and technology development plan project No. 20170220, and Key R & D plan of Guangxi science and technology program No. GuikeAB17195028.

References
[1] Hannam K, Butler G, Paris C M. Developments and key issues in tourism mobilities[J]. Annals of tourism research, 2014, 44(jan.):171-185.
[2] Girardin F, Blat J, Calabrese F, et al. Digital Footprinting: Uncovering Tourists with User-Generated Content[J]. IEEE Pervasive Computing, 2008, 7 (4):36-43. [4] Li Junyi. Tourism digital footprint: revealing tourists' time and space trajectory online [J]. Ideological front, 2013, 39 (3): 103-107

[3] Li Junyi. Tourism digital footprint: revealing tourists' time and space trajectory online [J]. Ideological front, 2013, 39 (3): 103-107
[4] Zheng W, Zhou R, Zhang Z, et al. Understanding the tourist mobility using GPS: How similar are the tourists?[J]. Tourism Management, 2019, 71:54-66.
[5] Zhao Xi, Lu Xiaoni, Liu Yuanyuan. Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an. Tourism Management. 69:368-383
[6] Zhang K, Chen Y, Li C. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing[J]. Tourism Management, 2019, 75:595-608.
[7] Fan Mengyu, Chen Yining, Zhang Hui. Spatial pattern and movement trajectory of tourist POI in Hulunbeier: An Empirical Analysis Based on geographical marker photos [J]. Resources and environment in arid areas, 2019, 33 (6): 203-208
[8] Zhang Xianxian, Li Jinghan, Zuo Ying, et al. Analysis of spatial and temporal behavior characteristics of tourists based on digital footprint: a case study of Nanjing City [J]. Economic geography, 2018, 38 (12): 226-233
[9] Liu Zhen, Dai ZEV, Lou Jiajun, et al. Spatial and temporal characteristics of urban recreation behavior based on digital footprint: a case study of Shanghai [J]. World geographic research, 2019, 28 (5): 95-105
[10] Ma Xiaolu, Xiang Yucheng, Zeng Xueqing. A study on the spatial structure of Xiachuan Tibet scenic spot in tourism digital footprint [J]. Tibet studies, 2019, Vol. 1: 64-72
[11] Luo Qiuju, Liang Sixian. Spatial and temporal characteristics of self driving tourism passenger flow based on digital footprint: a case study of Yunnan Province [J]. Journal of tourism, 2016, 31 (12): 41-50
[12] Zhao Y, Xu X, Wang M. Predicting Overall Customer Satisfaction: Big Data Evidence From Hotel Online Textual Reviews [J]. International Journal of Hospitality Management, 2019, 76(1):111-121.