Exploring Tourist Experience of Island Tourism Based on Text Mining: A Case Study of Jiangmen, China

Rucong Mo*, Huangxiong Qi

School of Economics and Management, Wuyi University, Jiangmen, China

Abstract: Island tourism is an important part of the development of the marine economy. Understanding the tourist experience of island tourism is conducive to promoting the development of marine tourism. This study takes the main island tourism attractions in Jiangmen, China, as a case, and analyzes the tourist experience of island tourism through a text mining method based on the text reviews of tourists on Ctrip. The study shows that beach, hotel, attraction, seafood and seawater are the main discourse system core of tourists’ evaluation of island tourism. Tourists’ sentiment evaluation of island tourism is generally positive. Nice, convenient, clean, cheap and comfortable are the main sentiment characteristic words of tourists. The results of LDA topic model analysis show that tourists island tourism experience is mainly divided into four categories: coastal scenery, seafood cuisine, beach environment and entrance service.

1 INTRODUCTION

The ocean has important significance and value to the world’s political and economic peace, stability and development. In recent years, China has been paying more and more attention to the development of marine industry. Jiangmen’s 14th Five-Year Plan and the outline of the 2035 Vision also clearly put forward the need to actively expand the blue development space and vigorously develop the marine economy, which provides new opportunities for the high-quality development of Jiangmen’s economy. Island tourism is one of the important industries of the marine economy, and the development of island tourism will certainly give a strong impetus to the sustainable development of the marine economy. The current research of scholars on island tourism mainly focuses on the development strategy of island tourism [1-3], the environment of island tourism destination [4-5], the value of island tourism resources [6], etc. However, the current academic research on island tourism is generally not deep enough, and few scholars have studied island tourism from the perspective of tourists experience.

With the development of the experience economy, tourists are increasingly valuing the specific experiences they obtain in the tourism process that meet their individual psychological needs and preferences [7]. Tourist experiences are also increasingly valued by researchers in rural tourism, ecotourism, black tourism, outbound tourism and other aspects [8-11]. Research methods are also becoming more and more diverse, from the traditional interview method and questionnaire survey method to the analysis and utilization of online text data [8-11]. The booming of online tourism industry not only has an impact on the way of tourists consume, but also changes the way of travel information is disseminated. With the rapid development and maturity of online travel agency (OTA), more and more tourists order travel products and share their experiences through OTA platforms such as Ctrip and Qunar. A large amount of user-generated content (UGC) has become an important source of information and decision-making basis for consumers. Compared with traditional questionnaires, this kind of unstructured open-ended reviews can better reflect tourists real experiences and concerns about travel experiences from different perspectives, and more truly explore tourists experience sentiments and demands. Big data analysis also provides companies with the opportunity to capture consumer preferences and identify market trends. Research on big data in tourism has become a hot topic of academic attention today [7,10-11]. However, the current academic research in the field of island tourism is not deep and sufficient, and there are few studies on the tourist experience of island tourism, and even fewer using web texts to analyze tourists experience perceptions of island tourism.

In view of this, this paper selects three representative island tourism attractions in Jiangmen, namely Shangchuan Island, Xiachuan Island and Langqin Bay-Naqin Peninsula, as the case study, collects the tourists’ reviews on Ctrip of these three attractions, and explores the tourists island tourism experience through word frequency analysis, sentiment analysis, experience topic classification and other text mining methods. The purpose is to provide effective theoretical support and decision-making basis for the market segmentation of island tourism and the development and management of Jiangmen island tourism destination, and to promote the high-quality development of Jiangmen’s marine economy.

* Corresponding author: morucong@126.com
2 METHODOLOGY

2.1 Case Selection

Jiangmen is located in the western part of the Pearl River Delta of China. It has a large sea area, a long coastline and a large number of islands, which has a total of 561 islands, ranking second in Guangdong province in terms of number, and is rich in marine resources. According to statistics, the total marine economy of Jiangmen has grown steadily in recent years, accounting for about 16% of the city’s GDP. Among them, Shangchuan Island, Xiachuan Island, Langqin Bay-Naqin Peninsula are the three main island tourism areas in Jiangmen, which are also national 4A scenic spots, and they are the places that receive the most tourists among Jiangmen’s islands every year. Therefore, this study selects Shangchuan Island, Xiachuan Island and Langqin Bay-Naqin Peninsula as the cases to study tourist experience of island tourism, which has certain representativeness and reference value.

2.2 Data Collection

This study use the Ctrip as the main source of review texts for the following reasons: firstly, Ctrip, as the largest OTA website in China, ranks first in the list of online travel websites in terms of monthly active users [12], and in 2019, Ctrip accounted for 43.35% of monthly active users on online travel platforms, so the review data on Ctrip is a good representative; what’s more, Ctrip has a “Travel Guides” section, which is conducive to the publication of tourists’ reviews on destinations and the collection of text data. Based on this, the study uses the “Travel Guides” section on Ctrip to search for reviews of scenic spots including Shangchuan Island, Xiachuan Island, Langqin Bay-Naqin Peninsula, and writes a relevant crawler program to obtain relevant tourists’ reviews by Python programming language. Finally, a total of 4621 reviews and 243234 words are collected. The deadline for data collection is September 2021. The specific data collection situation is shown in Table 1.

Table 1: Tourist Reviews Data Set of Island Tourism

<table>
<thead>
<tr>
<th>Attraction name</th>
<th>No. of reviews</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shangchuan Island Feisha Beach Resort</td>
<td>814</td>
<td>4.4</td>
</tr>
<tr>
<td>Xiachuan Island Wangfuzhou Tourist Area</td>
<td>2200</td>
<td>4.4</td>
</tr>
<tr>
<td>Chuandao Tourist Resort</td>
<td>294</td>
<td>4.4</td>
</tr>
<tr>
<td>Langqin Bay</td>
<td>162</td>
<td>4.0</td>
</tr>
<tr>
<td>Langqin Peninsula Geological Ocean Park</td>
<td>1151</td>
<td>4.5</td>
</tr>
<tr>
<td>Total</td>
<td>4621</td>
<td>4.3</td>
</tr>
</tbody>
</table>

Note: The specific names of the attractions on Ctrip’s “Travel Guides” that include reviews of Shangchuan Island and Xiachuan Island are Shangchuan Island Feisha Beach Resort, Xiachuan Island Wangfuzhou Tourist Area, and Chuandao Tourist Resort; the specific names of the attractions that include reviews of Langqin Bay-Naqin Peninsula are Langqin Bay and Langqin Peninsula Geological Ocean Park.

2.3 Data Analysis

Before conducting text analysis, data pre-processing is required, because data pre-processing has an extremely important impact on the data analysis results. Firstly, the data are screened to remove duplicate, meaningless and irrelevant information to improve the quality of text mining, and finally 4260 valid reviews are obtained, totaling 236461 words. Then the code program is written with PyCharm software to call the jieba package for word segmentation and stop word removal of the collected data.

After pre-processing the data, PyCharm is used to compile related programs to analyze the word frequency of tourists’ reviews. Then, the ROST EA software is used to sentiment analysis on the overall reviews to analyze the sentiment characteristics of tourists on the island tourism experience. Finally, LDA model is used to explore the topic of tourists island tourism experience.

3 RESULTS AND DISCUSSION

3.1 Frequency Analysis

Using PyCharm software to write the relevant programming language, the word frequency statistics of nouns and adjectives are conducted separately for the review text data, and the results are shown in Table 2. High-frequency words play an important role in understanding the semantic space of user reviews and the visitor experience [13]. Limited by the length of the article, Table 2 only lists the top 20 high-frequency words of nouns and adjectives.

As can be seen from Table 2, “beach” (935 times), “hotel” (670 times), “attraction” (576 times), “place” (545 times), “scenery” (533 times), “seafood” (482 times), “seawater” (447 times), “view” (430 times), “environment” (390 times), “ticket” (339 times) and other terms appear more frequently, reflecting tourists’ concerns about island tourism are mainly focuses on the special scenery of the seashore, marine specialty food, and accommodation conditions. The most frequent adjectives include “nice” (1409 times), “convenient” (604 times), “clean” (364 times), “cheap” (239 times), “comfortable” (214 times), “common” (189 times), “beautiful” (140 times), etc. These high-frequency adjectives reflect the main emotional experience of tourists towards island tourism.

Table 2: Word Frequency List

<table>
<thead>
<tr>
<th>No.</th>
<th>Noun</th>
<th>TF</th>
<th>Adjective</th>
<th>TF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>beach</td>
<td>935</td>
<td>nice</td>
<td>1409</td>
</tr>
<tr>
<td>2</td>
<td>hotel</td>
<td>670</td>
<td>convenient</td>
<td>604</td>
</tr>
<tr>
<td>3</td>
<td>attraction</td>
<td>576</td>
<td>clean</td>
<td>364</td>
</tr>
<tr>
<td>4</td>
<td>place</td>
<td>545</td>
<td>cheap</td>
<td>239</td>
</tr>
<tr>
<td>5</td>
<td>scenery</td>
<td>533</td>
<td>comfortable</td>
<td>214</td>
</tr>
<tr>
<td>6</td>
<td>seafood</td>
<td>482</td>
<td>common</td>
<td>189</td>
</tr>
<tr>
<td>7</td>
<td>seawater</td>
<td>447</td>
<td>beautiful</td>
<td>140</td>
</tr>
<tr>
<td>8</td>
<td>view</td>
<td>430</td>
<td>best</td>
<td>104</td>
</tr>
<tr>
<td>9</td>
<td>environment</td>
<td>390</td>
<td>interesting</td>
<td>101</td>
</tr>
<tr>
<td>10</td>
<td>ticket</td>
<td>339</td>
<td>big</td>
<td>89</td>
</tr>
<tr>
<td>11</td>
<td>queue</td>
<td>296</td>
<td>clear</td>
<td>71</td>
</tr>
<tr>
<td>12</td>
<td>walkway</td>
<td>288</td>
<td>fast</td>
<td>67</td>
</tr>
</tbody>
</table>
3.2 Sentiment Analysis

This study uses ROST EA to perform sentiment analysis on the overall reviews, and calculates the sentiment score corresponding to each review. The results of the study show that there are 3218 reviews with positive sentiment, accounting for 75.54% of the total reviews; 24 reviews with neutral sentiment, accounting for 0.56% of the total reviews; and 1018 reviews with negative sentiment, accounting for 23.90% of the total reviews. This shows that tourists’ evaluation of the island travel experience tends to be positive.

The segmental statistics of positive sentiment show that there are 1005 reviews with general positive sentiment, accounting for 23.59% of the total reviews; 885 reviews with moderate positive sentiment, accounting for 20.77% of the total reviews; 1328 reviews with high positive sentiment, accounting for 31.17% of the total reviews. It can be seen that tourists’ evaluation of the island tourism experience is generally high, indicating that tourists are more satisfied with the Jiangmen island tourism. Through the statistical analysis of word frequency of positive reviews, it is found that adjectives such as “good”, “convenient”, “clean”, “comfortable”, “cheap” and “beautiful” appear more frequently, which are the main positive sentiment words of tourists' experience of island tourism.

In addition, the statistical results of the negative sentiment segmentation show that the general negative sentiment accounts for 19.48% of the total reviews, which has 830 reviews; the moderate negative sentiment accounts for 2.58% of the total reviews, which has 110 reviews; the high negative sentiment accounts for 1.83% of the total reviews, which has 78 reviews. The results of statistical data analysis show that the negative sentiment of tourists on Jiangmen island tourism experience are basically mild, and there are few extremely negative sentiment. Among the negative reviews on Jiangmen island tourism, “inconvenient”, “dirty”, “hard”, “expensive” and “boring” are the main negative sentiment words.

3.3 Identified Topics in Online Reviews

This part mainly uses the LDA model for topic mining of tourists’ reviews. First, the nouns and adjectives in the reviews are extracted through the jieba package of Python, and words with high frequency but no reference meaning, such as “Taishan”, “Ctrip”, etc., are removed, a dictionary is established, and word frequency statistics are performed. Then, called the gensim package and used the topic coherence measure to calculate the topic coherence score. The higher the topic coherence score, the more relevant the topic words in the topic, and the less the ambiguity within the topic [14]. The results are shown in Figure 1, and the optimal number of topic can be determined as 4 based on the maximum coherence score.

Among them, topic 1 contains the words “scenery”, “trestle”, “beach”, “wind” and “seascape”, etc., reflecting tourists’ attention to the coastal scenery. Topic 2 involves “seafood”, “cost-effective”, “price”, “fresh”, “discount”, “delicacy”, “affordable”, etc., which reflects the attention of tourists to seafood delicacies and their prices. Topic 3 mainly includes “beach”, “environment”, “seawater”, “scenery”, “seaside” and “sand”, which reflect tourists’ attention to the beach and its related beach environment. Topic 4 contains keywords such as “attraction”, “queue”, “hour”, “ticket”, and “parking”, etc., highlighting tourists’ attention to the ticketing service at the entrance of attraction. It is worth noting that “queue” and “hour” appear more frequently, which indicates to a certain extent that the convenience of ticketing service at the
entrance of scenic spots is crucial to the satisfaction of tourists experience. At the same time, parking service is also an important element of concern to tourists. At present, tourists who travel to Jiangmen’s islands are mainly from Guangdong province, and self-driving tourists account for a considerable part. Therefore, the quality of parking lot service in the attraction is also the concern of tourists.

4 CONCLUSION

Tourists’ attention to island tourism mainly focuses on the seaside characteristic scenery, marine characteristic food, accommodation conditions and so on. Whether the scenic environment (including beach, seawater, etc.) of the island tourism is clean or not, whether the price is affordable, whether the scenic environment and scenery are beautiful, whether the surrounding hotel accommodation conditions are good, whether the special seafood catering and food is hygienic, and whether the price is reasonable will greatly affect the quality of tourists experience of island tourism. Therefore, the island tourism management department should strengthen the supervision and management of island tourism destination in terms of scenic spot and hotel environment, cost performance, marine food, etc., optimize the marketing plan, improve the cost performance of island tourism consumption, optimize the scenic spot environment, and improve the construction of hotels and other surrounding facilities in order to optimize the tourists island tourism experience and form a good reputation. In addition, tourists’ evaluation of island tourism experience is generally positive, but at the same time, the negative impact of negative reviews cannot be ignored. Destination marketers should understand the problems reported by tourists through negative reviews, and deal with them in time.

Tourists experience of island tourism can be divided into four categories: coastal scenery, seafood cuisine, beach environment, and entrance service. Therefore, the island tourism management department should focus on the development of island tourism from the above four aspects. First of all, in terms of coastal scenery, traditional and popular island tourism elements should be fully combined to create a characteristic island tourism IP. Secondly, we should enrich the types of island tourism projects and products, extend the industry chain of island tourism, and provide tourists with diverse and affordable marine products. Thirdly, optimize the beach environment and entertainment facilities, keep the beach environment clean all the time, and provide a good rest and amusement environment for tourists. Finally, it is necessary to further optimize ticket purchasing and ticket checking procedures, and do a good job in guidance of tourists.

ACKNOWLEDGMENT

This paper is the phased achievement of Guangdong Province Joint Training Graduate Demonstration Base Project (Project No. 503170060260), Undergraduate Teaching Quality and Teaching Reform Engineering Construction Project of Guangdong Province (Project No. GDJX2020006), Jiangmen Social Science Planning Project (Project No. JM2021B08), and Wuyi University Graduate Education Innovation Project (Project No. YJS-SFJD-20-05).

REFERENCES
