

Research on the Evaluation of Island Tourism Experience Based on LDA Model and Text Mining: Taking 1300 Tourist Reviews as an

Example

Weiming Yang*

Belarusian State University, Minsk 220004, Belarus

*Author to whom correspondence should be addressed.

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In this paper, 1300 tourist reviews from Ctrip.com are selected with the help of Python, and then text mining-related techniques are used to analyze the island tourism experience. The analysis shows that natural scenery, cultural experience, seafood quality, and accommodation cost-effective harvest more praise, while transport services, catering price transparency, and service quality need to be further improved. Based on the above, a series of suggestions on traffic, catering, accommodation, and cultural resources are proposed to support the optimization of scenic spots at the data level.

Keywords: Text mining; Tourism management; Island tourism; Review data

Online publication: March 26, 2025

1. Introduction

In recent years, online travel reviews have become a very important way for tourists to express their travel experiences and influence the decisions of others. On the one hand, these reviews reflect the evaluation of tourists on the service quality of scenic spots, and on the other hand, they also give the managers of scenic spots a basis for improvement. In this study, we use word frequency analysis, semantic network, LDA topic model, and sentiment analysis to explore the key points that tourists pay attention to and their sentiment tendency.

2. Study background

2.1. Tourism development and the importance of online reviews

In recent years, with the rapid development of the tourism industry, tourists' evaluation of tourist destinations has become an important indicator of the service quality and attractiveness of scenic spots ^[1]. Especially in the digital era,

Online Travel Reviews (OTR) have become the primary way for tourists to obtain information and share their experiences. These reviews not only influence the decision-making of potential tourists, but also provide the basis for scenic spot managers to improve services and optimize management.

2.2. Data value of online travel reviews

In China, online tourism platforms such as Ctrip, Mare's Nest, and Where to go have accumulated a large amount of tourists' review data on different scenic spots, which cover a wide range of aspects such as tourists' satisfaction, visiting experience, service quality, price perception, and so on. Text mining technology based on big data makes it possible to analyze these reviews, thus revealing the core concerns of tourists, extracting valuable information, and providing a scientific basis for scenic spots to improve service quality and tourism experience ^[2].

2.3. The special characteristics and research value of island tourism

As a typical coastal tourism destination, the development of the sea island of Tandang Town faces common problems such as traffic constraints and the balance between ecological protection and tourism development. By analyzing this case, not only can it provide a basis for local tourism upgrading, but its research methodology and conclusions are of reference significance for similar island tourism destinations. Especially in the context of the current rapid development of marine tourism, this kind of research is particularly important.

3. Research design and methodology

3.1. Research object and data source

In this study, the island of Tandang Town was selected as the research object, and the relevant data were obtained from the tourists' comments on Ctrip.com. These data largely contain the comments given by tourists on the scenic environment, traffic conditions, accommodation conditions, catering services, and the overall visiting experience.

3.2. Data processing and analysis methods

- (1) Word frequency analysis: Extracting high-frequency words to identify the main content of tourists' concerns.
- (2) Semantic network analysis: Semantic network analysis, focusing on the analysis of keyword interconnections, so as to clearly reveal the structural characteristics of the tourist experience.
- (3) LDA theme model analysis: This can be used to identify the core themes of the reviews and summarize the main aspects discussed by the visitors.
- (4) Sentiment analysis: Calculating the sentiment tendency of the comments, and identifying the positive and negative comments of tourists on the scenic spots.

3.3. Research process

The main process involved in this study covers the aspect of data collection, as well as data pre-processing, in addition to data storage mining analysis, and other related operations ^[3]. The process is shown in **Figure 1**.

3.3.1. Data collection

Visitor reviews on Ctrip.com were selected as the data source for the study, and a specific timeframe was first identified, and relevant information about the reviews was extracted, such as review content, ratings, and posting time. Data acquisition methods included crawler technology or other data collation means to ensure the comprehensiveness and representativeness of the data.

3.3.2. Data processing

Raw data may contain "noise" information, so data cleaning is required to remove irrelevant characters, special symbols, etc. Subsequently, word separation technology is used to split the content of the comments into independent words, which facilitates subsequent analyses.

3.3.3. Data storage

The processed text data is collated and stored in a database, while word frequency statistics and co-occurrence analysis are performed to identify high-frequency words and their associative relationships. The purpose of this session is to lay the foundation for topic extraction and sentiment analysis.

3.3.4. Result analysis and output

The study uses LDA topic modelling to classify the topic of the review content and combines it with sentiment analysis methods to explore tourists' tendency to evaluate different aspects ^[4]. Ultimately, based on the results of the data analysis, suggestions for optimizing the tourism experience are put forward to provide data support for the service improvement of scenic spots.



Figure 1. Research process

4. Python-based comment text data analysis

4.1. Data processing

4.1.1. Data extraction and pre-processing

The 1300 comment data were selected from Tandang Town Island sourced from Ctrip travel, and the online comment data of cultural tourism attractions were collected using a data collector and saved in Excel format.

4.1.2. Data cleaning

The invalid data was cleaned, such as empty comments, default comments, comments composed

entirely of numbers or letters, duplicate content, and other logically improper invalid comments; after cleaning, there were 1002 valid comments remaining.

4.1.3. Filtering deactivated words and customized participles

Subsequently, 2743 deactivated words such as time words, auxiliary words, English, and symbols were imported, and 7 customized participles of scenic spot names were added to make the word splitting more scientific.

4.2. Data mining analysis

Based on **Table 1**, "nice" was the most frequent word, with the majority of visitors rating the experience or product positively. Words such as "seafood," "view," and "scenery," which are closely related to the food and natural beauty of the travel experience, also show a high level of satisfaction. Words such as "diving," "hiking," "beach," and "island" are closely related to specific tourist activities. References to "Wentian" and "Lingdingyang" also reflect that tourists recognize their culture. Hong Kong's "Xiangzhou" and "Zhuhai" describe its geographical location. In terms of transport, "ferry tickets" and "seasickness" have become a major problem for tourists. On the whole, the satisfaction level of tourists' evaluations is relatively high.

Table 1. Word frequency									
Words	Word frequency	Words	Word frequency	Words	Word frequency	Words	Word frequency	Words	Word frequency
Not bad.	329	Ship Tickets	70	climb a mount ain	48	beautiful	40	Interestin g	32
seafood	240	Value for Money	69	price	47	fun	39	Diving	31
scenery	183	Hong Kong	64	clean	47	market	38	graceful	31
scenery	137	Processi ng	63	Holida y	46	Beautiful.	38	comforta ble	31
worthwhile	136	Xiangzh ou	55	Seasic kness	43	fresh	37	Wen Tianxian g	29
Zhuhai	133	Clear	54	Recom mende d	43	Seascape	37	Service	29
Sandy beach	132	Accom modatio n	53	Beautif ul	41	Cheap	36	Excursio ns	29

2025 Volume 3, Issue 2									
Sea	112	Boating	52	Island	41	Sea Breeze	35	Lindingy ang	27
Island	90	Swimmi ng	52	Favour ite	41	Sea	33	B&B	27
Hotels	79	Seaside	49	Leisur ely	40	Experienc e	32	Tourism	27

4.3. Social network (co-occurrence semantic network analysis)

In order to prevent errors in word cutting, the nodes are able to present the relationship between the word and the object [5]. Word frequency analysis can only be reflected by the frequency, but not a complete evaluation of the whole passage. In the covariance network, the thickness and length of the line can reflect the frequency of co-occurrence, that is, the strength of the relationship, as shown in Figure 2.



Figure 2. Social network

The semantics of the graphic discovery is based on "nice" as the core word, and aspects such as seafood, scenery, and beach all received very good ratings, which, together with "nice" constitute positive ratings. In the case of "seafood," there are synonyms with "processing," "beach," "value for money," and "worth it." Looking at the relationship network constructed with "beach" as the main term, the words "sea water," "clear," "seafood," and "island" form a synonym, which suggests that the situation of sea water management in the neighborhood is relatively good. Then the relationship network mainly based on "boat ticket," which links seafood, seashore and Hong Kong to Heung Chau, which is related to the geographical location, and comprehensively, through the word frequency analysis, it can be seen that the seafood, scenery, hotels, and other aspects of the place are recognized by tourists.

4.4. Sentiment analysis of word-of-mouth topics based on the LDA model

The LDA theme model is mainly used to identify the implicit themes from a large amount of text data. These themes can be regarded as the abstract concepts of recurring vocabulary in the document, each theme is composed of a set of vocabulary with a certain probability distribution ^[6]. The LDA model will try to rely on unsupervised learning in this

way, from the given set of documents to discover these latent themes, and at the same time, the distribution of the documents and the themes and the distribution between documents and topics, as well as between topics and words, are estimated accordingly.

Table 2 shows the results of the LDA topic model for analyzing tourists' comments, mainly to see what the most discussed topics are. From the data, "scenery" and "transport" are the most discussed topics, accounting for 34.758% and 24.830%, indicating that tourists are most concerned about what the scenery looks like and whether the transport is convenient. In contrast, "seafood" and "services" accounted for a slightly lower percentage, 21.146% and 19.266% respectively, but they were still an important part of tourists' discussions.

Topics	Theme weight	Occurrences	Percentage
Transport	24.830%	487	48.549%
Seafood	21.146%	125	12.485%
Service	19.266%	76	7.589%
Scenery	34.758%	314	31.377%
Total	100%	1002	100%

Table 2. Topics in the LDA model

On the whole, the attention of the four themes varies from high to low, but it will not be said that one completely overpowers the others, and the discussion is still relatively balanced. This also shows that tourists' evaluation is multi-faceted, not only concerned about the beauty of the scenery and smooth traffic, but also talks about how the food is and whether the service is good. This analysis can help the scenic spot to optimize in a more targeted way, for example, focusing on improving the transport issues that tourists discuss most, and at the same time doing more fine-tuning in seafood and service, so as to make the overall experience more upgraded.

4.5. Sentiment analysis

In this study, LDA topic model and SPSSAU text sentiment analysis were used to conduct in-depth mining of tourists' reviews with the aim of identifying core themes in the reviews and analyzing the sentiment tendencies expressed by tourists. Text sentiment was classified into "positive," "biased positive," "biased negative," and "negative" four categories in order to understand visitors' attitudes towards different aspects more systematically. "Positive" indicates that the comments are overall positive, conveying emotions of recognition and satisfaction; "biased positive," although generally positive, may contain some minor suggestions for improvement; "biased negative" represents a slight dissatisfaction in the review, which is positive but still carries room for improvement, while "negative" directly expresses a clear sense of dissatisfaction. Through this classification, the emotional attitudes of tourists on different topics can be extracted more accurately, providing more valuable references for subsequent analyses, as shown in **Table 3**.

	Table 5. Sentiment analysis table							
Traffic	Sentiment analysis	Seafood	Sentiment analysis	Service	Sentiment analysis	Scenery	Sentiment analysis	
Route	Positive	Not bad.	Positive	not bad	Positive	Not bad.	Positive	
Excursions	Positive	Fresh	Positive	Enthusiastic	Positive	Worth it	Positive	
Hong Kong	Positive bias	Seafood	Positive bias	Ctrip	Positive	Island	Positive	
Xiangzhou	Positive	Affordable	Positive bias	Hotel	Negative	Wen Tianxiang	Positive	
By Boat	Negative	Tasty	Positive bias	Attitude	Positive	Clean	Positive	
Going up	Negative	Price	Positive bias	Price	Negative	Holiday	Positive	
Boat Tickets	Negative	Restaurant	Positive bias	-	-	Beautiful	Positive	
Seasickness	Negative	Processing	Positive bias	-	-	Sandy Beach	Positive bias	
Tickets	Negative	Sea urchin	Positive bias	-	-	sea	Positive	
fearful	Negative.	Not cheap	Negative	-	-	Prices	Negative	

TADIC J. Schullinging analysis laure	Table 3.	Sentiment	analysis	table
---	----------	-----------	----------	-------

In terms of transport, the route and the experience of travelling around the island received a lot of positive comments, but the boat experience had the problems of "seasickness" and "expensive ticket price." In catering, the freshness of the seafood and the processing service were positively evaluated, but not so good in terms of price. In terms of accommodation, the service attitude is very good, but the satisfaction of tourists in terms of price is relatively low. Overall, there are more prominent shortcomings in terms of transport, while the advantages in terms of nature and culture are quite significant. To sum up, the negative evaluations mainly focus on the boat experience and price in terms of transport, while the positive evaluations mainly focus on the user experience and the scenery.

4.6. Research summary

Based on 1300 tourists' review data on Ctrip.com about Tandang Town Island, this study systematically mined tourists' evaluation characteristics and emotional tendencies towards the scenic spot through data cleaning, word frequency statistics, co-occurring semantic network analysis, and LDA topic modelling. It is found that the overall satisfaction of tourists is high, and the core concerns are focused on four aspects: natural landscape, catering experience, transport service, and accommodation quality.

4.6.1. Tourists' evaluations are mainly positive

High-frequency words such as "not bad," "worthwhile," and "beautiful scenery" are dominant, indicating that the scenic spot has a high level of satisfaction in terms of natural scenery (e.g., seawater, sandy beaches) and cultural experience (e.g., Wentianxiang Scenic Spot). This indicates that the scenic spots are widely recognized for their natural beauty (e.g., sea water, beaches) and cultural experience (e.g., Wen Tianxiang spots).

4.6.2. Transportation services are the main shortcoming

Negative comments focus on issues such as "seasickness," "expensive fares," and "fewer boat trips," which directly affect tourists' experience.

4.6.3. Catering and accommodation are divided

The freshness and processing service of seafood are praised, but the price transparency is insufficient; there is room for improvement in the service attitude and cost performance of hotels.

4.6.4. Cultural resources potential to be tapped

Tourists' references to cultural symbols such as "Wen Tianxiang" and "Lingdingyang" show the feasibility of cultural tourism integration, but the related development is currently insufficient.

5. Management insights and suggestions

5.1. Optimizing the transport experience and improving visitor satisfaction

Tourists reflect "seasickness" and "expensive fares," we can consider optimizing the comfort of the boat, such as adding some anti-sickness facilities, and make corresponding adjustments to the fare strategy ^[7]. At the same time, meeting the needs of different tourists can reduce the negative experience that tourists may have.

5.2. Strengthening seafood and catering management

"Seafood" and "processing" are highly rated. However, it is important to ensure that prices are transparent and hygiene standards are strictly met to prevent negative feedback such as "not cheap." It is beneficial to introduce a special seafood set menu to further strengthen the restaurant's service quality training. The management can also use the power of market supervision or cooperation with merchants to stabilize price fluctuations and improve the overall dining experience ^[8].

5.3. Improving accommodation and service quality

In view of the existence of "hotel prices" and "service attitude" of the negative evaluation of the situation, there is a need to optimize the service process and the introduction of different types of rooms, to further improve the hygiene conditions, supporting facilities, and the introduction of digital services to promote efficiency ^[9].

5.4. Digging deep into the potential of natural and cultural resources

Island tourism development needs to balance ecological protection and economic development to avoid excessive commercialization affecting the ecological environment ^[10]. Combined with the "Wen Tianxiang," "Lingdingyang,"

and other cultural symbols to create cultural attractions or cultural and creative products, it is important to dig into the potential of cultural resources to enhance the attractiveness of the destination ^[11]. In addition to strengthening the ecological protection of publicity, it is necessary to show the results of environmental protection (such as clean beach action) and consolidate the "beautiful" "clean" tourism image.

Disclosure statement

The author declares no conflict of interest.

References

- Sun X, Cao C, 2022, Research on the Image of Wuhu Nighttime Tourism Destination from the Perspective of Tourists' Experience—An Approach based on Network Comments. Journal of Longdong College, 33(02): 134–138.
- [2] Bai W, Chen X, 2023, Research and Analysis of 'Intelligent Tourism' Text based on Data Mining Technology. Journal of Zhaoqing College, 44(02): 45–54.
- [3] Zhao H, Yang X, Zhao X, et al., 2022, Big Data and AI Technology Empowering High-Quality Integrated Development of Culture and Tourism. Technology and Industry, 22(09): 388–394.
- [4] Xu L, Zhang Y, Yao L, 2024, Text Mining of Travel Online Reviews based on BiLSTM and LDA Models. Modern Computer, 30(22): 130–134 + 140.
- [5] Ma J, Xu J, 2024, Analysis of Technology Development Trend based on Thesis Semantics and Citation Network. Software Guide, 23(11): 100–106.
- [6] Chu X, Shen S, Wang M, et al., 2024, The Experience of World-Class Universities in Exploring the Specification of Generative Artificial Intelligence Application and the Inspiration to Our Country—Text Mining based on LDA Topic Model Analysis. Modern Distance Education, (03): 38–47.
- [7] Gao S, Li Z, 2023, Analysis of Tariff Reform of Passenger-Roller Transport in Qiongzhou Strait. Water Transport Management, 45(08): 1–4 + 7.
- [8] Wen Y, Sun Y, 2025, Research on the Optimization Strategy of Cold Chain Logistics in Seafood Aquaculture Industry—Taking the Case of Xianren Island Sea Area. Commercial Exhibition Economy, (05): 135–138.
- [9] Yu Z, He Q, Shi X, et al., 2024, Research on Mechanism and Path of New Productivity Enabling High-Quality Development of B&B Industry. Journal of Hebei Tourism Vocational College, 29(03): 1–9.
- [10] Wen Y, Lai Q, Xu Z, 2021, Construction and Empirical Evidence of Evaluation Index System for Territorial Tourism Development in Yangjiang City. Journal of Leshan Normal College, 36(05): 46–53.
- [11] Zhu S, Li Z, Jiang Y, et al., 2023, A Study on Dimensional Differentiation of Island Tourism Destination Image based on Big Data Analysis—Taking Hailing Island in Yangjiang, Guangdong as an Example. Business and Exhibition Economy, (07): 51–54.

Publisher's note

Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.