

Article

Not peer-reviewed version

Tourist Satisfaction in World Heritage Sites Based on Natural Language Processing

[Xuerui Gai](#)^{*}, [Ru Liu](#)^{*}, [Yue Song](#)^{*}

Posted Date: 16 January 2026

doi: 10.20944/preprints202601.1227.v1

Keywords: natural language processing; LDA; West Lake and Li River; tourist satisfaction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Tourist Satisfaction in World Heritage Sites Based on Natural Language Processing

Xuerui Gai *, Ru Liu * and Yue Song *

School of Tourism Management, Shenyang Normal University, Shenyang 110034, China

* Correspondence: gaixuerui@163.com (R.G.); 15184761071@163.com (R.L.); 15604149479@163.com (Y.S.)

Abstract

Monitoring tourists' satisfaction has always been a concern for World Heritage Sites. Due to its ability to truly reflect tourists' travel experience and satisfaction evaluations, online review texts, after being mined and analyzed, can be promptly transformed into intuitive and highly valuable information, thus attracting increasing attention. This paper proposes a research method combining the Latent Dirichlet Allocation (LDA) topic model (hereinafter referred to as LDA) and tourist satisfaction responses, using user online reviews on Ctrip about the World Heritage Site of the West Lake Cultural Landscape of Hangzhou (hereinafter referred to as Hangzhou West Lake) and the Lijiang River Scenic Area in Lingchuan, Southern China Karst (hereinafter referred to as Guilin Lijiang River) as the original data. Ctrip provides two channels for tourists to evaluate their travel experience: satisfaction scoring and text comments. The Chinese comments are classified into themes through LDA in natural language processing, and the themes that tourists are most concerned about are identified as service, tourist attractions and travel experience. Through the 5-level satisfaction scoring (1 to 5) of tourists, a statistical analysis is conducted on tourist satisfaction and the themes they are most concerned about. The results show that the satisfaction under the service theme is the lowest, with nearly 11.5% of tourists at Hangzhou West Lake and nearly 21.1% of tourists at Guilin Lijiang River expressing dissatisfaction. Although there is a gap in tourist satisfaction between the two World Heritage Sites, both need to be significantly improved.

Keywords: natural language processing; LDA; West Lake and Li River; tourist satisfaction

1. Introduction

China is a major country in terms of world heritage. As of 2024, the total number of Chinese world heritage sites listed in the World Heritage List by UNESCO (United Nations Educational Scientific and Cultural Organization) has reached 59, ranking among the top in the world [1]. World heritage site tourism not only plays a significant role in public cultural services and lifelong education but also serves as an important source of economic income for these sites. As an important criterion for measuring the social responsibility and economic function of world heritage sites, tourist satisfaction and its influencing factors have always been a core issue of academic concern [2,3]. Currently, the assessment of tourist satisfaction mainly relies on questionnaire surveys [4,5]. Before the era of big data, questionnaire surveys were necessary and feasible, providing good guidance for the development of scenic spots. However, due to the pre-assumption of data dimensions and influencing factors by researchers, the data collection through questionnaire surveys is inevitably subjective and pre-assumed. To overcome these drawbacks, online text analysis methods have gradually emerged [6–8]. These methods are widely used in the hotel and homestay sectors, but there are relatively few studies on the analysis of online comments of world heritage sites in China. Jinglong Li and Haitao Wang [9] selected Chinese world heritage scenic spots as research objects by searching for keywords on the Baidu Index platform. Through correlation analysis and the application of online text analysis methods, they explored the differences in attention and experience characteristics between online users and real tourists towards Chinese world heritage scenic spots,

and identified and addressed some factors affecting tourist satisfaction. Xiaobin Ma et al. [10] studied the image perception of the world cultural heritage, the Suzhou Gardens, from the perspective of foreign tourists. These research methods either overly rely on the quality of dictionaries, with limited universality, or fail to uncover deeper patterns within the data, resulting in relatively simplistic conclusions. To better address these issues, this paper attempts to propose a satisfaction research method based on LDA. By clustering the themes and specific comments in online reviews using LDA, it aims to effectively extract the factors contributing to dissatisfaction in world heritage sites, providing a basis for enhancing their influence and competitiveness.

2. Research Design

2.1. Overview of the Study Areas

This paper selects two world heritage sites - Hangzhou West Lake and Lijiang River in Guilin as the study areas. The West Lake Scenic Area in Hangzhou is one of the first batch of national key scenic spots approved by the State Council and a national 5A-level tourist attraction. The West Lake is surrounded by mountains on three sides and contains clear water, covering an area of approximately 60 square kilometers. On June 24, 2011, at the 35th World Heritage Conference held in Paris, France, the West Lake Cultural Landscape of Hangzhou was inscribed on the World Heritage List. The West Lake is an outstanding example of landscape design under the Chinese aesthetic theory of "harmony between man and nature" and "expressing emotions through mountains and waters", showcasing the artistic style of "poetic and picturesque" in Chinese landscape design since the Song Dynasty [11].

The Lijiang River Scenic Area in Guilin was among the first batch of national-level scenic spots approved by the State Council and is also a national 5A-level tourist attraction. On June 23, 2014, at the 38th World Heritage Conference, the "South China Karst II" project, which includes Guilin in Guangxi, Shibing in Guizhou, Jinfoshan in Chongqing, and Huanjiang in Guangxi, was approved and inscribed on the World Natural Heritage List. The karst landform in Guilin is located within the Lijiang River Scenic Area. The Guilin karst holds a unique position in the world's karst landforms, especially the karst peak forest landform. World karst experts unanimously recognize the Guilin karst as the pearl on the crown of the world's karst [12].

2.2. LDA Model

LDA, as a topic model based on Bayesian learning, is an extension of latent semantic analysis and probabilistic latent semantic analysis. It was proposed by Blei et al. in 2002 [13]. LDA is widely used in fields such as text data clustering, computer vision, and bioinformatics processing. Essentially, LDA is a probabilistic graphical model. Figure 1 shows the plate notation of LDA as a probabilistic graphical model. In the Figure, nodes represent random variables, solid nodes are observed variables, and dashed nodes are latent variables; directed edges represent probabilistic dependencies; rectangles (plates) indicate repetition, and the numbers within the plates indicate the number of repetitions.

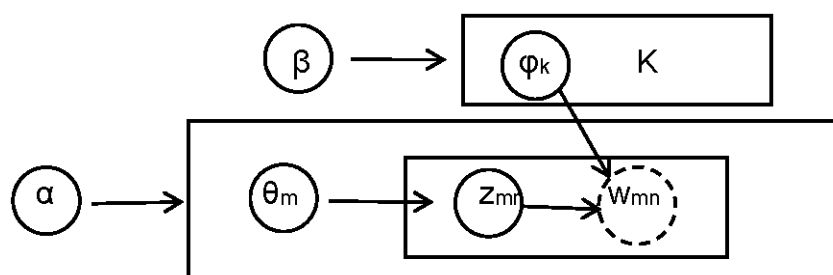


Figure 1. Block representation of LDA probabilistic graph model.

Here, K represents the number of topics, and M is the total number of online texts. β is the conjugate prior parameter of the word distribution for each topic, and α is the conjugate prior parameter of the topic distribution for each online text. Z_{mn} is the topic of the n th word in the m th online text, and W_{mn} is the n th word in the m th online text [14]. For a collection of online texts, the observable known variables are W_{mn} . The prior parameters α and β are given based on experience. All other variables are unknown latent variables that need to be learned and estimated based on the observed variables.

3. Data Sources and Data Preprocessing

3.1. Data Sources

Ctrip is a well-known tourism brand in China, dedicated to providing one-stop global travel services for travelers, including accommodation reservations, transportation ticket reservations, vacation reservations, and business travel management services, covering the entire travel process from before, during, and after the trip, as well as destination services. Through a transaction network composed of an APP, website, and 24/7 customer service center, it achieves a perfect connection between users and products, providing travel services to hundreds of millions of travelers each year. By searching on the Ctrip website, it was found that the tourist review data is detailed and of high quality. Therefore, this article uses this as the data source for analysis. Using Octopus, online review data of tourists for Hangzhou West Lake and Guilin Lijiang River was obtained, covering a period of nearly 10 years from October 2015 to January 2025, with a total of 7,491 reviews (4,051 for Hangzhou West Lake and 3,440 for Guilin Lijiang River). Before data processing, it is necessary to reduce noise to minimize interference with topic clustering. The collected reviews need to be cleaned: removing duplicate reviews, redundant punctuation, and meaningless text, etc. Eventually, 6,759 reviews were obtained (3,501 for Hangzhou West Lake and 3,258 for Guilin Lijiang River).

3.2. Data Preprocessing

The most crucial step for LDA topic clustering to achieve good results is data preprocessing. This study adopts two steps for data preprocessing. Firstly, the open-source Python library Jieba is used for Chinese text segmentation. Jieba uses two segmentation methods: dictionary-based segmentation and Hidden Markov Model-based segmentation. After segmentation, a total of 191,308 words were obtained. Secondly, stop words are removed. This study uses the stop word list released by Harbin Institute of Technology, which contains 767 stop words. These 767 stop words serve as the basic stop word list for this study. Based on the preprocessing results of this study, the basic stop word list needs to be further expanded. For example, the high-frequency words such as "啊啊", "好好", and "赞赞赞" that appeared in this study are of no help for the topic clustering of Hangzhou West Lake and Guilin Lijiang River. Therefore, the stop word list needs to be expanded to better remove stop words. After the above data preprocessing steps, the quality of the data used for topic classification has been greatly improved.

4. Results

4.1. Topic Classification

This study conducts topic clustering on the vocabulary of online reviews of Hangzhou West Lake and Guilin Lijiang River on Ctrip website based on LDA. The initial values of the topics are specifically set as: $\alpha = 0.5$, $\beta = 0.1$, $K = 10$; the number of iterations for Gibbs sampling is set to 50. During the topic clustering estimation process, the number of topics K is sequentially set to: 10, 9, 8, 7, 6, 5, 4, 3, 2, 1. After multiple estimations and dynamic parameter adjustments, and combined with the observed pyLDAvis visualization effect analysis, the optimal number of topics $K = 3$ is finally obtained. The pyLDAvis visualization result is shown in Figure 1.

In Figure 2, the circles represent the clustered topics. The three circles do not overlap and are far apart, indicating that the topic clustering effect is very good. When topic 1 is selected, the

corresponding circle will turn red, and the keywords in this topic will also be displayed (Note: Figure 2 shows the keywords when no topic is selected), as shown in the right column list. The frequency of the keywords is indicated by the length of the red bars. The pyLDavis visualization graph can help the operators of Hangzhou West Lake and Guilin Lijiang River to view the distribution of keywords under each topic very intuitively.



Figure 2. Visualization of topic clustering for online comments by (a) Hangzhou West Lake(above) and (b) Guilin Lijiang River tourists using PyLDavis.

Previous studies have shown [3,4] that statistical analysis based on keyword clustering has significant practical value. It helps to understand the logical relationship of the context of online text

content and cluster the inherent themes of online texts, and is more conducive to precisely exploring the specific problems that need to be solved in World Heritage Sites. Based on the specific circumstances of this study and drawing on the empirical classification method of Xiaobin Ma and Jinhe Zhang [10], LDA topic classification was first adopted, followed by manual content analysis for adjustment and refinement. The combination of the speed of artificial intelligence and the experience of researchers can fully leverage the advantages of natural language processing and human initiative, enabling the acquisition of optimized data quickly and effectively. On the basis of LDA topic clustering, 10 tourism experts and 10 tourism graduate students were selected to name the topics, analyze, discuss and adjust the keywords. The researchers ultimately determined 3 topics and their contents: Topic 1 - Services, Topic 2 - Tourist Attractions, and Topic 3 - Travel Experience, as shown in Tables 1 and 2.

Table 1. Topic classification and corresponding topic words for online comments from Hangzhou West Lake tourists.

Topic1: Services	Topic2: Tourist Attractions	Topic3: Travel Experience
Bicycle 0.008	West Lake 0.080	Suggestion" 0.007
Sightseeing Cruise 0.004	Broken Bridge 0.009	Worthwhile 0.006
Travel-friendly 0.003	Su Causeway 0.008	Not bad 0.006
Service 0.003	Bai Causeway 0.006	Enjoy 0.005
Hotel 0.002	Leifeng Pagoda 0.006	Beauty 0.004

Table 2. Topic classification and corresponding topic words for online comments from Guilin Lijiang River tourists.

Topic1: Services	Topic2: Tourist Attractions	Topic3: Travel Experience
Tour Guide 0.012	Li River 0.067	Not bad 0.018
Sightseeing Cruise 0.012	Guilin Scenery 0.027	Worthwhile 0.012
Dock 0.012	View 0.013	Promotion 0.005
Ctrip 0.011	Yangshuo 0.011	Negative Feedback 0.004
Service 0.007	Bamboo Raft 0.007	Beauty 0.002

Analyzing Table 1, it can be found that the clustering logical relationship of Topic 1, Topic 2 and Topic 3 is very clear: the service-related topics mainly include bicycles, sightseeing cruise and travel-friendly, etc.; the tourist attraction-related topics include typical tourist attractions in the West Lake scenic area of Hangzhou such as the West Lake, Broken Bridge, Su Causeway and Leifeng Pagoda; the tourist experience-related topics mainly include feelings and evaluations of the Hangzhou West Lake, such as suggestion, worthwhile, enjoy and beauty, etc.

Analyzing Table 2, it can be found that the clustering logical relationship of Topic 1, Topic 2 and Topic 3 is very clear: the service-related topics mainly include tour guide, sightseeing cruise, Ctrip and dock, etc.; the tourist attraction-related topics include typical tourist attractions in the Lijiang River scenic area of Guilin such as the Lijiang River, Guilin Scenery, Yangshuo and bamboo rafts, etc.; the tourist experience-related topics mainly include feelings and evaluations of the Lijiang River in Guilin, such as not bad, promotion, negative feedback and beauty, etc.

This paper extracts, classifies and analyzes the content of online comments from tourists visiting the Hangzhou West Lake and the Lijiang River in Guilin by topic, obtaining 723 topic words for services, 1199 topic words for tourist attractions and 1250 topic words for Travel experience in the Hangzhou West Lake; 1112 topic words for services, 774 topic words for tourist attractions and 1114 topic words for Travel experience in the Lijiang River in Guilin. From the frequency of topic words, it can be seen that the most attractive aspect for tourists in the Hangzhou West Lake is Travel experience, followed by tourist attractions and then services; in the Lijiang River in Guilin, it is Travel experience, services and tourist attractions respectively.

Service is the most fundamental element for tourists in tourism activities and runs through the entire process of tourism activities. From the perspective of tourists, the most direct and effective way to enjoy natural scenery is to see and touch nature, that is, various tourist attractions in the Hangzhou West Lake and the Lijiang River in Guilin such as the Three Pools Mirroring the Moon and Elephant Trunk Hill. These tourist attractions have strong appeal and tourists have a strong interest in them. With the help of tourist attractions and the yearning for World Heritage Sites, experience of the Hangzhou West Lake and the Lijiang River in Guilin will be formed in the minds of tourists, thereby forming travel experience. Tourist experience is the most important activity for tourists, which requires contact, perception and evaluation before forming a final judgment on tourism satisfaction.

4.2. Satisfaction Analysis

Ctrip provides two channels for tourists to evaluate their travel experience: satisfaction scoring and text comments. Online text comments from tourists have implicit characteristics and require natural language processing, that is, LDA chosen in this study for topic classification, which belongs to overall clustering; while satisfaction scoring has explicit characteristics, and tourists can directly evaluate their Travel experience through a 5-point scoring system (1 to 5 points), which belongs to specific evaluation. In this paper, the Travel experience evaluated directly are divided into five topics: services, tourist attractions and Travel experience. In this study, if tourists give a score of 4 or 5, they are classified as satisfied, while a score of 1 to 3 indicates that tourists are dissatisfied with the tourism service. By drawing the theme classification response chart of tourists' online comments on Hangzhou West Lake and Lijiang River in Guilin (as shown in Figure 4) based on the extracted themes, it is possible to intuitively understand which theme contents tourists are satisfied or dissatisfied with.

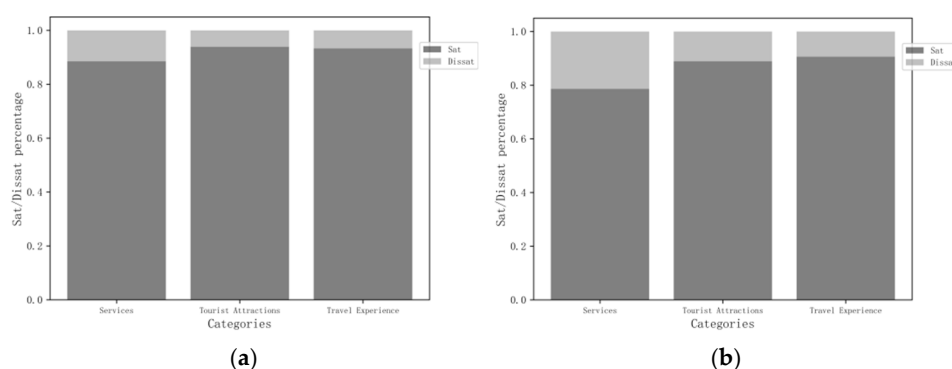


Figure 3. Satisfaction response chart of online topic classification for (a) Hangzhou West Lake and (b) Guilin Lijiang River tourist comments.

As can be seen from Figure 3, the lowest satisfaction of tourists for both Hangzhou West Lake and Lijiang River in Guilin is with the service, accounting for 88.5% and 78.9% respectively. For Hangzhou West Lake, the remaining two items of satisfaction from low to high are tourism experience and tourism attractions, accounting for 93.3% and 93.9% respectively; for Lijiang River in Guilin, the remaining two items of dissatisfaction from low to high are tourism experience and tourism attractions, accounting for 90.1% and 94.8% respectively. For tourist attractions, the themes of dissatisfaction are more significant for improving the quality of the attractions. By observing the original text of the service theme, there are nearly a thousand pieces of data involving themes such as service, convenience and management. For example, Comment A: "It was so crowded. I went there on February 16th, not during a holiday, but the traffic was terrible. Hangzhou, a quasi-third-tier metropolis, still has a two-way single-lane road around the lake. The traffic was terrible. There wasn't even a bus lane, and private buses were all squeezed together. It took several hours to walk three kilometers. If you have a tight schedule, don't come here, or choose to visit the area within walking distance. Warning." Comment B: "The rowing boats were replaced by big boats, and no price

difference was refunded. The customer service couldn't be found." Comment C: "Tourism in Guilin is very disappointing. The Guilin Tourism Bureau is poorly managed. The 5A scenic area of Lijiang River has no official account to promote to tourists. The streets and alleys of Guilin are full of small individual tour operators. Independent tourists don't need to join a tour group. They can buy tickets at the entrance of Mopanshan Scenic Area. The ticket price is 270 yuan in the peak season, plus 30 yuan for meals, totaling 300 yuan. The boat tickets are not divided into first and second floors and are randomly issued. The guide who picked us up persuaded us to upgrade to the second floor, charging an extra 50 yuan per person. We were tricked. I'm so sad! I'm not happy! Moreover, the guide didn't give us the tickets until we arrived at the scenic spot. When we got the tickets, we had to wait for two hours. It would have been better to buy the tickets ourselves. The whole city of Guilin lives on tourism. When they see tourists, it's like seeing Tang Seng's meat. Everyone wants to have a bite. If you ask passers-by how to tour Lijiang River, they either say they don't know or directly introduce you to a travel agency. Even the doorman is like this. Also, the travel agencies trick you. They say that if you don't book today, there won't be any tickets tomorrow. You won't be able to resist! Be very careful of the travel agencies using the half-day tour of Lijiang River to trick you. You must buy the tickets at Mopanshan Scenic Area yourself. Any tour of Lijiang River that costs less than 300 yuan is a watered-down version. Comment D: "The tourism market in Guilin is too chaotic. I booked the Lijiang River boat ticket and the direct shuttle bus. They said the driver would contact us about the pick-up time and place the day before, but no one contacted us until midnight. I called the customer service of Ctrip the next morning, and they said they couldn't contact the travel agency either. It was past 8 o'clock when someone called me and told me to take a taxi to the pier to get the ticket. I arrived at the pier less than ten minutes before the departure. It was so frustrating! The shuttle bus fee wasn't refunded, and the taxi fare wasn't settled for a long time! There's no credibility at all!" Comment E: "The scenery is fine, but the transportation is inconvenient. The scenic area is poorly managed, and the staff's quality is uneven. Their attitude is not friendly, showing a particularly disappointing phenomenon. I won't go there again." Comment F: "This is not what I imagined. It's too commercialized now. The staff in this scenic area have a bad service attitude. I couldn't understand what they were saying. It's very unprofessional." These review data are all over the original text, indicating that the management and service levels inside and outside the scenic areas need to be improved. The local tourism bureau and scenic area managers should take corresponding measures to increase tourists' satisfaction.

5. Discussion

This study proposes a new method for analyzing tourists' satisfaction. Previous methods for online text satisfaction research mostly used sentiment analysis to replace satisfaction [16,17], or constructed a tourist satisfaction evaluation index system [16]. In the tourism industry, emotions are generally divided into two opposite dimensions: positive emotions (positive affection) and negative emotions (negative affection), and tourists' satisfaction or dissatisfaction is inferred based on these two types of emotions. The traditional view holds that positive emotions have a positive impact on tourists' satisfaction, while negative emotions have a negative impact. Many studies have proved that emotions are important factors affecting consumer satisfaction [18–20]. Oliver [21] believes that positive and negative emotions of consumers have different impacts on consumer satisfaction. Tourists who experience positive emotions during their travel will have higher satisfaction, while those who experience negative emotions will have lower satisfaction. Sentiment analysis mainly focuses on analyzing the positive or negative emotional tendencies expressed in online text opinions, that is, positive or negative emotions [22]; while satisfaction is the result of tourists' perception of a series of events that occur during their interaction in the tourist destination, and these perceptions come from the tourists' personalities and the vacation goals they consider important [23]. Through the analysis of sentiment and satisfaction, it is questionable to simply equate the positive and negative of sentiment analysis with the high and low of satisfaction.

This study adopts the method of natural language processing (Natural Language Processing), which is an important direction for future online text analysis. Natural language processing is widely applied in search, web, documents, and auto-completion, etc. The natural language processing in documents can be further divided into types such as product review classification, knowledge extraction, and truth verification [24]. The natural language processing method adopted in this study is based on statistical machine learning, which has greatly improved in accuracy and stability compared to rule-based methods. Statistical machine learning is currently an important method for text classification [25], and the LDA model is a major research method for text classification. In traditional tourism research, when studying the content that tourists are concerned about, predefined research paradigms are often used, mainly relying on some mature scales for questionnaire surveys and statistical analysis. Even when using mixed research methods - qualitative and quantitative research, the research is still carried out within the framework of predefined models. Obtaining non-predefined Travel experience and concerns has always been a difficult point in traditional tourism research. This study contributes a beneficial methodological supplement to the tourism research paradigm.

This study provides research findings that can be used by tourism attraction operators to solve specific problems. By deeply exploring and addressing the problems existing in the attractions, it can enhance tourists' travel experience and satisfaction. Based on the online reviews on Ctrip, it can be directly observed which themes tourists are concerned about and the specific contents included in these themes. This is the practical application value of online review texts. More importantly, tourism attraction operators can use the specific contents included in these themes to make targeted improvements and solve related shortcomings and problems. Taking the Hangzhou West Lake and the Lijiang River in Guilin as examples in this study, the themes analyzed are: service, tourist attractions, and travel experience. Although tourists show over 90.0% satisfaction with tourist attractions, travel experience, and service, the dissatisfaction rate with service is as high as over 11.0%. This indicates that there are significant problems in service quality and tourism transportation in the Hangzhou West Lake and the Lijiang River in Guilin, which need to be urgently addressed.

This study also has some limitations: the number of online travel agencies selected needs to be expanded. This study only selected Ctrip, and there are other similar online travel websites such as Meituan and Dianping. Whether the research results are limited by the category of the website, the characteristics of users, and the reviews still needs to be further verified. Secondly, the satisfaction is directly measured by tourists' online ratings, and the usefulness of online reviews still needs to be further verified. Finally, this study takes the Hangzhou West Lake and the Lijiang River in Guilin as examples. The Hangzhou West Lake and the Lijiang River in Guilin only represent ecological attractions. Different types of attractions are complex and diverse with strong uniqueness. Whether the methods discussed in this study are applicable to other types of attractions is also a direction worth paying attention to.

6. Conclusions

The LDA model was used to classify the themes of tourists' online reviews and obtain the frequency of theme words. Through the obtained hot theme words of online reviews, it is initially judged that the themes affecting the satisfaction of tourists in Hangzhou West Lake and Lijiang River in Guilin are: Theme 1: Service, Theme 2: Tourism Attraction, and Theme 3: Travel Experience. The frequency of the key words shows that the travel experience are the most attractive to tourists. From the perspective of scenic spots, the "golden signboard" of World Heritage Sites can greatly enhance the reputation of the scenic spots and is an important means to attract tourists [15]. From the perspective of tourists, World Heritage Sites have great appeal. Visiting famous World Heritage Sites is a lifelong wish for many people. Based on the theme classification extracted and combined with the tourists' satisfaction scores, the theme classification and tourists' satisfaction response diagrams of online comments on Hangzhou West Lake and Guilin Li River were drawn. Tourists are highly satisfied with the tourist attractions, travel experience and service of Hangzhou West Lake and Guilin

Li River. Tourists' feelings towards tourist attractions such as the Three Pools Mirroring the Moon and Leifeng Pagoda in Hangzhou West Lake are aesthetic and affectionate emotional tourism experiences, while those towards tourist attractions such as Elephant Trunk Hill and Peach Blossom River in Guilin Li River are beautiful and interesting emotional tourism experiences. The final tourism experience is worth it, beautiful, good and recommended for others to visit and enjoy. Although tourists are very satisfied with the above theme contents, for the contents under the service theme, nearly 11.5% of tourists in Hangzhou West Lake and nearly 21.1% of tourists in Guilin Li River are dissatisfied. In Hangzhou West Lake, the main problems are the traffic outside and inside the scenic area, which are extremely congested, and during peak hours, tourists have to walk a long way to get a taxi. In Guilin Li River, the main problem is the high ticket price and the frequent occurrence of overcharging. As the most important travel experience of World Heritage Sites - Hangzhou West Lake and Guilin Li River, they should not be negatively affected by service issues on the "golden signboard" of World Heritage Sites. This requires the high attention of local governments and scenic spot managers.

Author Contributions: Conceptualization, X.G. and R.L.; methodology, R.L.; software, Y.S.; validation, R.L., Y.S. and X.G.; formal analysis, R.L.; investigation, R.L.; resources, Y.S.; data curation, Y.S.; writing—original draft preparation, X.G.; writing—review and editing, X.G.; visualization, X.G.; supervision, X.G.; project administration, X.G.; funding acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Liaoning Federation of Social Science Project(No.2024lslybkt-104).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to privacy concerns, the data provided in this study can be obtained from the corresponding author.

Conflicts of Interest: The authors declare that they have no potential conflicts of interest.

References

1. Chinese Government Website.Super Complete and Stunning! 59 World Heritage Sites in China: Where Do You Want to Go? Available online: https://www.gov.cn/yaowen/liebiao/202408/content_6966788.htm (accessed on 06 August 2024).
2. Yu, J.; Liu, M. Research hotspots and frontiers of public service satisfaction: A bibliometric analysis. *Journal of Shenyang Normal University (Social Science Edition)* 2023, 47, 73-78.
3. Zhao, C.; Chen, M. Analysis of influencing factors of tourist satisfaction based on web text mining. *Statistics & Decision* 2019, 35, 115-118.
4. Guo, J.; Li, W.; Bu, Y. Investigation and database construction of urban green spaces based on RS and GIS. *Science Technology and Engineering* 2007, 7, 3877-3880.
5. Cai, Q.; Gao, C. Influencing factors and mechanism of tourist satisfaction in garden expo parks: An empirical study based on structural equation modeling. *Chinese Landscape Architecture* 2016, 32, 58-64.
6. Zhou, X.; Zheng Y. Influencing factors of tourist satisfaction and verification of brand loyalty in the Three Gorges region. *Statistics & Decision* 2016,182-184.
7. Liu, Y.; Bao, J.; et al. Emotional characteristics of Chinese tourists to Australia: A big data-based text analysis. *Tourism Tribune* 2017, 32, 46-58.
8. Zhang, R.; Zhang, J. Perceived tourism image of Shanghai Chenshan Botanical Garden: An analysis based on web text and IPA model. *Chinese Landscape Architecture* 2019, 35, 83-87.
9. Li, J.; Wang, H. Online attention and satisfaction of China's world heritage sites based on UGC data. *Journal of Southwest University (Natural Science Edition)* 2023, 45, 138-150.
10. Ma, X.; Zhang, J. Image perception of Suzhou Gardens (a world cultural heritage site) by foreign tourists: A multi-granularity analysis based on natural language processing. *China Ecotourism* 2022, 12, 550-565.

11. West Lake Scenic Area Administrative Committee. Overview of West Lake Scenic Area. Available online: <https://westlake.hangzhou.gov.cn/col/col1598539/index.html> (accessed on 22 May 2019).
12. Guilin Lijiang Scenic Area Administrative Committee. Introduction to Guilin Lijiang Scenic Area. Available online: https://lgw.guilin.gov.cn/jqjs/201606/t20160620_579380.htm (accessed on 20 November 2024).
13. David, M. B.; Andrew Y. N.; Michael, I. J. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 2003, 3, 993-1022.
14. Li, H. *Machine Learning Methods*, Tsinghua University Press: Beijing, China, 2022; pp. 324-348.
15. Xue, L.; Wu, B.; Qi, L. Value transformation of China's world heritage and implications for communication concepts. *Econ. Geogr* 2010, 30, 844-870.
16. Feng, K.; Yang, Q. CHANG Xinyi, LI Yanlai. Customer satisfaction evaluation of fresh produce e-commerce based on online reviews and stochastic dominance criteria. *Chinese Journal of Management Science* 2021, 29, 205-216.
17. Wu, W.; Gao, B.; Yang, H.; Sun, H. Impact of review texts on hotel satisfaction: A sentiment analysis-based approach. *Data Analysis and Knowledge Discovery* 2017, 1, 62-71.
18. Wang, J.; Yan, F. Sentiment analysis of microblog comments based on LSTM + Word2vec. *Journal of Shenyang Normal University (Natural Science Edition)* 2024, 42, 138-144.
19. Sun, B.; Ao, C.; Wang, J.; Zhao, M. Satisfaction evaluation of ecotourism based on web text mining. *Operations Research and Management Science* 2022, 31, 165-172.
20. Lai, X.; Xie, J.; Wang, X. Quality elements of rural tourism attractions and their impact on tourist satisfaction: A Kano model analysis based on explainable neural networks. *Tourism Science* 2025, 39, 120-139.
21. Oliver R L. Cognitive, affective, and attribute bases of the satisfaction response. *J. Consum. Res.* 1993, 20, 418-430.
22. Liu, B.; Translated by Liu K.; Wang, X. *Sentiment Analysis*, 2nd ed.; China Machine Press: Beijing, China, 2022.
23. Ryan C.; Translated by Li M.; Wang L. *Research Methods in Tourism*, Tourism Education Press: Beijing, China, 2012.
24. Lane H; Howard C; Max H; Translated by Shi,L.; Lu, X.; Tang K. et al. *Practical Natural Language Processing*, Posts & Telecom Press: Beijing, China, 2021.
25. Che, W.; Dou, Z.; Feng, Y.; et al. Natural language processing in the model era: Challenges, opportunities, and developments. *Science China Information Sciences* 2023, 53, 1645-1687.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.