International Journal of Science and Engineering Applications
Volume 14-Issue 10, 157 – 160, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1410.1025

Tourist Sentiment Analysis of Scenic Spots Based on Textual Big Data: A Case Study of the Chengdu Research Base of Giant Panda Breeding

Ziyang Liu
School of College of Urban Construction
Yangtze University
Jingzhou, China

Abstract: With intensifying competition in the tourism industry, refined analysis of visitor experience has become increasingly critical for effective destination management. Taking the Chengdu Research Base of Giant Panda Breeding as a case study, this research leverages visitor reviews from online travel platforms to explore emotional tendencies through text mining and sentiment analysis techniques. Using a self-developed Python web crawler, approximately 2,500 valid reviews were accurately collected and rigorously screened from the Ctrip platform. Methodologically, this study innovatively integrates SnowNLP sentiment polarity analysis with Latent Dirichlet Allocation (LDA) topic modeling to systematically examine visitor experience from three perspectives: overall sentiment tendency, evaluation of specific landscape elements, and correlations with management factors. The findings indicate a generally high level of visitor satisfaction, with panda exhibits and natural environments identified as primary sources of positive feedback. Nonetheless, issues such as inadequate infrastructure and suboptimal service quality were also evident. Based on these insights, targeted recommendations are proposed, including optimizing transportation organization, upgrading infrastructure, and enhancing service training to improve overall visitor experience. The methodological integration and detailed analysis provide robust empirical support and practical guidance for tourism management decision-making.

Keywords: Tourism evaluation; Sentiment analysis; Online reviews; Textual big data; Chengdu Research Base of Giant Panda Breeding

1. Introduction

With the rapid development of the tourism industry, service quality and visitor experience have become critical factors influencing destination competitiveness. Tourist sentiment is a direct indicator of experience quality. Traditional survey methods, though valuable, are limited by sample size, subjectivity, and high cost, making it difficult to fully capture authentic visitor perceptions (Alaei et al., 2023)[1]. In recent years, with the rapid advancement of online information technologies and social media, textual big data analysis based on user-generated content (UGC) has emerged as an important tool in tourism research (Li et al., 2019)[2]. By mining visitor reviews shared on platforms such as Ctrip and Qunar, researchers can gain more accurate and intuitive insights into visitor emotions and demands, providing a solid foundation for scientific decision-making and management optimization (Xiu & Qiguang, 2024)[3]; (Huang & Chelliah, 2024)[4].

Sentiment analysis, a major branch of natural language processing (NLP), aims to identify and extract emotions, attitudes, and opinions expressed in text (Jim et al., 2024)[5]. In tourism studies, sentiment analysis has been widely used to assess visitor satisfaction regarding facilities, service quality, environmental hygiene, and other aspects (Rahadian, 2024)[6]; (Fu & Pan, 2022)[7]. Common approaches include lexicon-based methods and machine learning-based methods (Yu et al., 2024)[8]. Lexicon-based methods rely on predefined sentiment dictionaries to calculate sentiment scores by matching sentiment words in the text, while machine learning methods train classifiers such as support vector machines (SVM), Naive Bayes, or deep learning models like long short-term memory networks (LSTM) to automatically identify sentiment polarity (Fu & Pan, 2022)[7]; (Jim et al., 2024)[5].

This study utilizes textual big data to analyze tourist sentiment at the Chengdu Research Base of Giant Panda Breeding, including overall sentiment evaluation, sentiment toward specific landscape elements, and the relationship between sentiment and management factors. Data collection and processing procedures follow established methodologies for UGC segmentation, frequency computation, and sentiment polarity analysis (Zhou & He, 2025)[9]. The research covers positive, neutral, and negative sentiment classification, sentiment evaluation of facilities and services, and recommendations for management optimization.

By applying web crawling to collect review data from Ctrip and other platforms, and conducting sentiment and topic modeling analyses in Python, this study innovatively:

1. Leverages massive UGC data to overcome limitations of traditional survey methods;
2. Integrates sentiment polarity analysis with LDA topic modeling for in-depth exploration of visitor experience from both sentiment and thematic perspectives.

The findings enrich theoretical and methodological research on tourism sentiment analysis and provide practical optimization suggestions for the Chengdu Research Base of Giant Panda Breeding.

2. Study Area and Data Source

2.1 Study Area

The Chengdu Research Base of Giant Panda Breeding, located in Chenghua District, Chengdu, Sichuan Province, was established in 1987 and covers approximately 100 hectares. It is a globally renowned center for giant panda conservation and research, with a rich natural ecological environment, diverse plant species, and high green coverage, providing an ideal habitat for pandas and other rare animals. In recent years, driven by socioeconomic development and increasing tourism demand, the number of visitors has steadily grown, making it a prominent tourist attraction in Chengdu. The base features multiple areas including the cub nursery, adult panda viewing zones, red panda ecological exhibition areas, and a science education center, focusing on wildlife conservation, public education, and eco-tourism development

2.2 Data Source

Data were primarily collected from online travel platforms such as Ctrip using a self-developed Python web crawler, with "Chengdu Research Base of Giant Panda Breeding" as the core keyword. The crawler was configured with filters for time range, rating levels, and other criteria to ensure data diversity and representativeness, initially capturing over 3,000 reviews

2.3 Data Preprocessing

Considering potential noise in raw reviews, a multi-stage cleaning process was employed. First, irrelevant advertisements, duplicates, and off-topic reviews were automatically removed using regular expressions. Second, manual verification by the research team ensured the authenticity and validity of the data, resulting in approximately 2,500 high-quality reviews used for analysis. This strict preprocessing guarantees accuracy and reliability for subsequent sentiment analysis and topic modeling.

3. Methodology

3.1 Textual Sentiment Analysis

SnowNLP, a tool specifically designed for Chinese text, was used for sentiment analysis. It provides sentiment scores

3.4 Title and Authors

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For three authors, you may have to improvise.

3.5 Subsequent Pages

For pages other than the first page, start at the top of the page, and continue in double-column format. The two columns on the last page should be as close to equal length as possible.

ranging from 0 to 1, effectively capturing positive, neutral, and negative expressions in Chinese text. Sentiment scores were categorized as follows: positive (≥0.7), neutral (0.3–0.7), and negative (≤0.3), ensuring clear differentiation and interpretability.

3.2 LDA Topic Modeling

Latent Dirichlet Allocation (LDA) was employed to explore specific topics of interest within visitor reviews. Chinese word segmentation was performed using the Jieba library, and stop words were removed to reduce noise. The text was converted into a bag-of-words format using the Gensim library.

Optimal topic number was determined by evaluating perplexity and coherence scores for topic numbers ranging from 3 to 10. Results indicated that 5 topics achieved the lowest perplexity and highest coherence, which were then used for model training (num_topics = 5, passes = 500, random_state = 42).

non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.3 Visualization

Matplotlib was used to generate visualizations including pie charts and histograms for overall sentiment distribution, radar charts for sentiment toward specific landscape elements, and word clouds for negative review keywords, enhancing interpretability and management relevance.

4. Results and Analysis

4.1 Overall Sentiments

Analysis of Ctrip reviews revealed that positive sentiment accounted for 64.3%, negative sentiment 17.0%, and neutral sentiment 18.7%, indicating generally positive visitor feedback with some negative experiences.

High sentiment scores clustered near 1.0, confirming high overall satisfaction, while small number of low-score reviews (<0.1) suggested areas needing improvement.

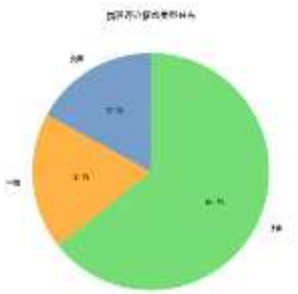

Figure. 1 Distribution of Sentiment Types in Ctrip Visitor Reviews.

International Journal of Science and Engineering Applications
Volume 14-Issue 10, 157 – 160, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1410.1025

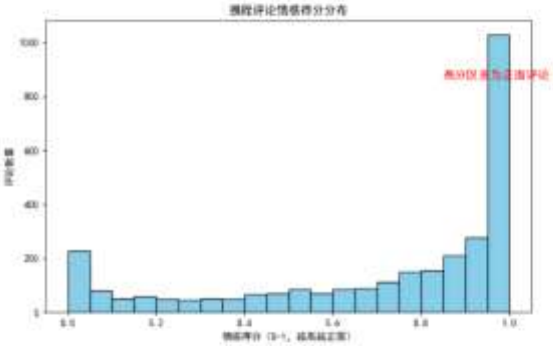

Figure. 2 Detailed Distribution Histogram of Sentiment Scores in Ctrip Visitor Reviews

4.2 Sentiment by Landscape Element

Visitors rated panda exhibits and natural environments highest, while infrastructure and service quality received lower scores, highlighting areas for improvement in facility development and service management.

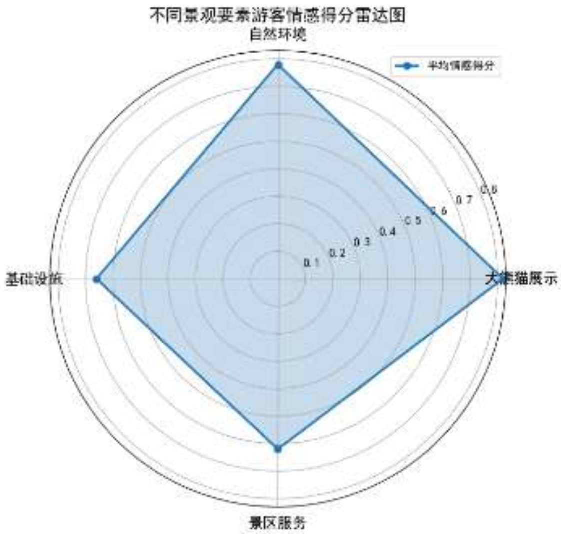

Figure. 3 Radar Chart of Visitor Sentiment

4.3 Sentiment and Management Factors

Negative emotions were concentrated around keywords such as "queue," "lack," "crowded," "traffic congestion," and "service," reflecting dissatisfaction with service quality, facility management, and crowd control.

Figure. 4 Word Cloud of High-Frequency Keywords

5. Discussion

5.1 Sources of Positive and Negative Sentiment

Positive sentiment largely stemmed from close encounters with pandas and the scenic ecological environment, which enhanced visitor experience. Negative sentiment arose from management issues, including traffic congestion, outdated infrastructure, and staff service attitudes.

5.2 Recommendations for Facility and Service Optimization

To address traffic issues, real-time monitoring systems and additional shuttle services are recommended. Infrastructure upgrades, including rest area and restroom expansion, are suggested, referencing recent projects at Beijing Zoo. Regular staff training programs are recommended to improve service quality, similar to practices at Shanghai Disneyland.

5.3 Implications for Management

Fine-grained sentiment analysis identifies specific satisfaction and dissatisfaction sources, guiding targeted management interventions to enhance visitor experience and destination competitiveness.

5.4 Future Research

Future studies should expand data sources to include platforms like Weibo, Xiaohongshu, and Douyin, and apply advanced deep learning models such as BERT for improved sentiment accuracy. Longitudinal monitoring of visitor sentiment could provide real-time feedback for adaptive management.

6. Conclusion

This study applied textual big data analysis to visitor sentiment at the Chengdu Research Base of Giant Panda Breeding, revealing generally positive sentiment while identifying key management issues. Panda exhibits and natural environments drive positive feedback, whereas infrastructure and service management contribute to negative sentiment. The integration of SnowNLP and LDA models provides robust insights from both sentiment and thematic dimensions, offering empirical support for management optimization. Limitations include relatively narrow data sources and room for methodological improvement. Future research should expand data coverage and leverage advanced analysis techniques to further support refined tourism management practices.

7. References

[1] Alaei A, Wang Y, Bui V, et al. Target-Oriented Data Annotation for Emotion and Sentiment Analysis in Tourism Related Social Media Data. *Future Internet*, 2023, 15(4): 150.

[2] Li Q, Li S, Zhang S, et al. A Review of Text Corpus-Based Tourism Big Data Mining. *Applied Sciences*, 2019, 9(16): 3300.

[3] Shu X, Lv Q. Tourism Service Management for Nanchuan Shenlong Gorge Scenic Area: Insights from Web Text Analysis. *Proceedings of Business and Economic Studies*, 2024, 7(6): 170–179.

[4] Huang X, Chelliah S. Attributes Influencing Tourist Satisfaction: Sentiment Analysis and Topic Modeling of Online Reviews. *Journal of China Tourism Research*, 2024: 1–20.

[5] Jim J R, Talukder M A R, Malakar P, et al. Recent Advancements and Challenges of NLP-based Sentiment Analysis: A State-of-the-Art Review. *Natural Language Processing Journal*, 2024, 6: 100059.

[6] Fachrur A R. Sentiment Analysis of Hotel Reviews in Anyer Beach Tourism Area: A Lexicon-Based Text Mining Approach. *JELAJAH: Journal of Tourism and Hospitality*, 2024, 5(2).

[7] Fu M, Pan L. Sentiment Analysis of Tourist Scenic Spots Internet Comments Based on LSTM. *Mathematical Problems in Engineering*, 2022, 2022: 1–9.

[8] Yu Y, Chen J, Mehraliyev F, et al. Exploring the Diversity of Emotion in Hospitality and Tourism from Big Data: A Novel Sentiment Dictionary. *International Journal of Contemporary Hospitality Management*, 2024, 36(12): 4237–4257.

[9] Zhou R Y, He W. Sentiment-Based Classification of Urban Park Landscapes Using UGC Data. *China Urban Forestry*, 2025, 23(02): 61–67

.