Research on Guilin Tourism Image Based on LDA

Changli Huang[1,2(✉)] and Yanling Huang[1,2]

[1] College of Tourism and Landscape Architecture, Guilin University of Technology, Guilin, Guangxi, China
huangchanglihq@163.com
[2] Institute of Guangxi Tourism Industry, Guilin, Guangxi, China

Abstract. The image of a tourist destination is the core factor in the formation of a city's tourist image, and is also a key factor in the decision-making of tourist attractions managers. This research takes Guilin tourist city as the research object, takes travel notes crawled by Ctrip travel website as the research data, uses LDA theme model and sentiment analysis method, based on the two-dimensional model of "cognition-emotion" constructed based on the image of tourist destination, to discuss Cognitive Image and Emotional Image of Guilin Tourism Purpose. The results show that: according to the LDA model, six major topics, namely, tourism landscape, local cuisine, entertainment activities, tourism services, commercialization degree, and experience perception of Guilin tourist city are the main image attributes perceived. The positive emotional tendency of tourists' comments accounted for 61%, indicating that tourists' overall perception of the image of Guilin tourist destinations is satisfied and recognized.

Keywords: LDA model · sentiment analysis method · tourism perception image · "cognition-emotion" model

1 Introduction

The image of a tourist destination is the sum of tourists' impressions, ideas, beliefs and perceptions of a tourist destination [1]. It is one of the keys to the precise marketing of the destination and also the core factor in the formation of a city's tourist image. With the development of the web2.0 era, many tourists use social media to share their travel experience and generate a large number of real and effective comments and travel notes. This type of User Generated Content (UGC) effects the tourist's perception of the tourist destination, and provides a new perspective for studying the image of tourist destinations.

Most researches on the perceived image of tourist destinations are based on text analysis methods such as word frequency analysis and semantic analysis based on network data. Although the number of samples of such methods is large, they are insufficient for in-depth mining. Based on this, this research takes Guilin as an example, uses python to crawl online travel notes, and adopts the LDA theme model to explore the tourism image from the perspective of tourists' perception.

Z. Zhan et al. (Eds.): EIMSS 2022, AHCS 7, pp. 78–84, 2023.
https://doi.org/10.2991/978-94-6463-024-4_10

2 Literature Review

At present, combined with natural language technology in the field of artificial intelligence, in-depth exploration of the image of tourist destinations has become a current research hotspot [2]. Zhang and Shu (2019) used KHCoder software to explore the tourist image of inbound tourists to the ethnic villages in Qiandongnan Prefecture on the TripAdvisor website [3]. Wong et al. used tourist review data on TripAdvisor to reveal the evolution of Macau's destination image from 2005–2013 [4]. Peng et al. (2019) used the hornet's travel notes from 2011 to 2017 as the data source, and used the content analysis method to analyze the tourism image statically and dynamically [5]. Lu and Liao (2019) use text analysis and grounding to analyze the "cognition-emotion" three-dimensional model of Nanyue Hengshan [6]. To sum up, in terms of research objects, existing research focuses on small-scale tourism image analysis such as scenic spots, and lacks research on large-scale image perception of tourist cities. However, the topic classification technology in deep learning can effectively classify irregular UGC according to topic types and improve the reliability of topic classification [7].

3 LDA Topic Model

3.1 LDA Topic Modeling

Firstly, this research uses "Guilin" as the search term and uses python to crawl the travel notes of the Ctrip travel website from 2010 to 2020 as the data source. Secondly, we must use regular expressions to filter the noise data, remove the interference information that affects the results of text mining, and retain the text data needed for the research. Then, through the visualization of the topic confusion, the inflection point of the graph change is selected as the optimal topic solution. Finally, the preprocessed documents are substituted into the LDA topic model to calculation, and the topic probabilities and feature word probabilities are obtained respectively.

3.2 The Basic Theme of LDA

LDA is an unsupervised machine learning technology that overcomes the limitations of other topic modeling methods such as latent semantic analysis (LSA) and probabilistic latent semantic analysis (PLSA) [8]. The idea of the theme is: a document is formed by selecting a certain topic with a certain probability, and selecting a certain word from the topic with a certain probability. It means that a document represents a probability distribution composed of several topics, and each topic is in turn Represents a probability distribution composed of several words. The LDA topic model is shown in Fig. 1.

Among them, M represents the total number of texts in the corpus, N represents the total number of words in the text, Z represents the topic, W represents the word vector of the text, θ represents the topic distribution, α is the Dirichlet hyperparameter of θ, and Ψ represents the word distribution, β is the hyperparameter of Dirichlet where Ψ is.

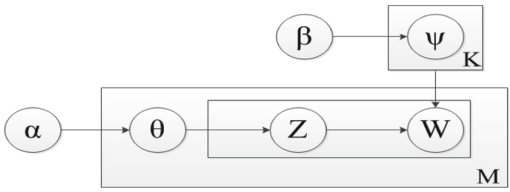

Fig. 1. LDA model

Table 1. STATISTICS OF THE TOP 35 HIGH-FREQUENCY WORDS

Feature words	frequency	Feature words	frequency	Feature words
Guilin	15469	Bamboo raft	6449	Beer fish
Yangshuo	12785	Hour	6337	pier
Lijiang	12367	Xingping	6183	travel
hotel	11673	Dragon's Back	5765	rice flour
West Street	10562	Yulong River	5639	landscape
Scenic spot	10408	night	5109	Inn
Terraces	8541	landscape	5013	gallery
frequency	Feature words	frequency	Feature words	frequency
4925	beer	3923	Liu Sanjie	3584
4808	garden	3842	Eternal Love	3495
4587	journey	3822	Yinziyan	3431
4476	drifting	3794	Tickets	3416
4319	Liangjiang	3782	like	3246
4034	minute	3720	recommend	3032
3993	afternoon	3701	impression	2834

4 Data Analysis

4.1 Data Acquisition and Preprocessing

This study uses Python to crawl the travel notes about Guilin on Ctrip, and obtains 1745 travel notes. Then, the travel notes were filtered and segmented to obtain 1238 effective travel notes and 16379 high-frequency words (frequency $>$ 5). This study only listed the first 35 high frequency words (see Table 1).

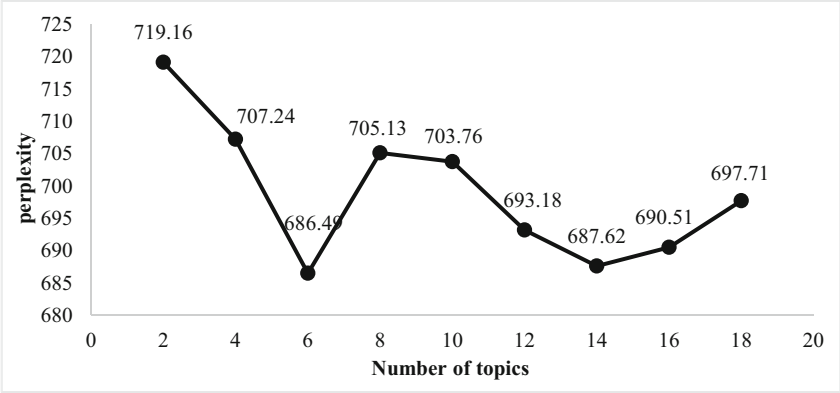

Fig. 2. Line chart of perplexity under different K values

4.2 LDA Subject Classification

This research calls the ida. perplexied() function to obtain the degree of perplexity, and performs LDA topic modeling on the travel notes text after data preprocessing. When the number of topics is finally determined to be K = 6, the degree of perplexity is the smallest (see Fig. 2). The topic distribution results of the travel notes text is are shown in Table 2 (this article only lists 10 characteristic words).

It can be seen from Table 2 that the LDA theme clusters are relatively clear, which reflects tourists' perception of Guilin's tourism image (see Table 2). Topic1 is the tourist landscape, namely Yangshuo, scenery, Lijiang River, Yinziyan and West Street, etc. It mainly reflects the tourists' perception of the tourist landmarks and scenic spots in Guilin; topic2 is local delicacies. Local delicacies such as rice noodles, beer fish, roast duck are popular among tourists; topic3 is entertainment activities, in which words such as magnificent and characteristic express tourists' perception of entertainment activities, and words such as "Sister Liu", "bamboo raft", and "e-donkey" reflect the tourists' choice of entertainment activities; topic4 is the degree of commercialization. The words "bar", "lively" and "a street" vividly reflect the lively and commercialization of Yangshuo West Street; topic5 is the experience and feeling. Tourists' travel in Guilin echoes their psychology, which creates collision and psychological induction; topic6 is tourism services, mainly designing related vocabularies such as scenic spot tickets, tourism consumption, staff services and local residents' ways.

4.3 Sentiment Image Analysis

This study uses the sentiment analysis function in the ROST CM6.0 software to analyze the sentiment tendency of the sample's emotional vocabulary. The result is shown in Table 3. It can be seen from the table that the frequency of positive emotion words is as high as 61%, and the frequency of negative emotion words is 14%. Positive emotions include words such as "satisfaction, good, like, shock" and other words, showing the tourists' sense of identity with the tourist destinations in Guilin. Neutral perception include exclusive vocabularies such as "fair, okay, almost, that's it". Negative perception

Table 2. THE DIMENSION DIVISION OF EACH TOPIC FEATURE WORD

Topic1	Probability	Topic2	Probability	Topic3	Probability
Yangshuo	0.013	Yangshuo	0.014	Yulong River	0.011
Lijiang	0.009	delicacy	0.008	Liu Sanjie	0.009
Ancient town	0.005	rice flour	0.008	Bamboo raft	0.007
longji	0.005	West Street	0.007	Magnificent	0.006
Yulong River	0.004	good to eat	0.005	Village	0.006
West Street	0.003	recommend	0.005	Eternal Love	0.005
Liang jiang	0.003	Beer fish	0.004	Cruise ship	0.004
pier	0.002	Photograph	0.003	EDonkey	0.003
Yinziyan	0.002	night	0.003	feature	0.002
garden	0.002	Roast duck	0.002	pier	0.001
Topic4	Probability	Topic5	Probability	Topic6	Probability
commercialize	0.009	feature	0.010	Tickets	0.010
	0.009	happy	0.009	price	0.009
a street	0.007	guide	0.008	queue	0.006
lively	0.005	like	0.008	Online	0.004
boss	0.005	misty rain	0.005	guide	0.004
Shop	0.004	Fit	0.005	Tourist	0.003
Consumption	0.003	Casual	0.004	Bus	0.003
breath	0.003	travel	0.004	sincere	0.002
Online	0.002	good	0.002	free	0.002
shopping	0.001	Photograph	0.002	Grinning	0.002

Table 3. TOURIST SENTIMENT EVALUATION WORDS

Emotional attributes	Emotional word evaluation
Positive perception (61%)	Nice, satisfied, happy, like, great, enjoy, happy, free, leisure, magnificent
Neutral perception (25%)	Generally, that's it, nothing, okay, almost
Negative perception (14%)	Bad, can't, regret, not enough, pity, boring, too many people, disappointed, wasteful, crowded

vocabularies include such as “bad, pity, noisy, confusion”, which reflects the regrets of tourists after their expectations have failed.

5 Conclusion

This research selects the most representative landscape city—Guilin, uses Ctrip’s travel notes as the data basis, and uses LDA theme methods and sentiment analysis to explore the tourism perception image of Guilin’s landscape city. The study found that the tourism landscape, local cuisine, entertainment activities, tourism services, commercialization, and experience perception of Guilin tourist destinations are the most concerned by tourists.

Tourists have a high degree of satisfaction and recognition with Guilin tourist destinations, and their emotional characteristics are polarized. Tourists’ emotional evaluations are divided into three categories: positive, neutral, and negative. Tourists’ negative perception of Guilin is mainly reflected in the weather, the flow of people, and the irregular management of the scenic spot.

This article excavated and analyzed the online comment texts of tourists, and achieved certain results, but still has the following shortcomings: this research only selects Ctrip as the data source, and does not involve travel notes from foreign travel websites and other domestic travel websites, so that the samples are insufficient.

Acknowledgments. Institute of Guangxi Tourism Industry Postgraduate Research and Innovation Fund Project “Research on the Cognitive Differences of “Host and Guest” Guilin Tourism Image Based on the Analysis of Graphic Data” (LYCY2021-22); National Science Foundation of China (NSFC) “Research on Host-Guest Online Interaction Behavior Characteristics and Its Influencing Mechanism at Ethnic Tourist Destinations in Southwest China” (72064007).

References

1. Kim SE, Lee KY, Shin SI et al (2017) Effects of tourism information quality in social media on destination image formation: the case of Sina Weibo. Inf Manag 54(06):687–702
2. Zhang K, Chen Y, Li C (2019) 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. Tour Manage 75(07):595–608
3. Zhang HC, Shu BY (2019) A study on the perception of international tourism image of ethnic villages based on online text analysis-taking Qiandongnan prefecture as an example. J Northwest Univ Natl (Philos Soc Sci Ed) 03:145–152
4. Wong CUI, Qi SS (2017) Tracking the evolution of a destination’s image by text-mining online reviews - the case of Macau. Tour Manag Perspect 23(07):19–29
5. Peng D, Huang YT (2019) Research on the imagery of Lijiang Ancient City tourist destination: content analysis based on web text. J Tour 34(09):80–89
6. Lu LJ, Liao XP (2019) Research on the image perception of Nanyue Hengshan tourist destination based on UGC data. Econ Geogr 39(12):221–229

7. Liang CC, Li RJ (2020) Analysis of Lijiang Ancient City image perception based on LDA and feature dimensions. Prog Geogr Sci 39(04):614–626
8. Brandt T, Bendler J, Neumann D (2017) Social media analytics and value creation in urban smart tourism ecosystem. Inf Manag 54(6):703–713