

Research on Service Industry Development Strategy of Guangxi Based on Data Mining

Wengui Liu*, Zhichao Li, Bowen Zheng, Haoyan Xu, Meng Huang

School of Business, Guilin University of Electronic Science and Technology, Guilin, Guangxi, China

**Corresponding Author.*

Abstract: With the rapid development of China's economy, the service industry has become an important part of the national economy, among which tourism, as an important pillar of the service industry, has a significant role in promoting regional economic development. Guangxi Zhuang Autonomous Region, relying on its unique natural scenery and rich ethnic cultural resources, has become a well-known tourist destination at home and abroad. However, in the face of increasingly fierce competition in the tourism market, the tourism industry in Guangxi urgently needs to improve its core competitiveness. Through the study using data mining technology, the development status quo of Guangxi's tourism industry is deeply analyzed, and taking Harbin tourism industry as a comparison, the shortcomings of Guangxi's tourism industry in terms of service quality, product innovation, infrastructure construction and marketing are revealed by crawling and analyzing the texts of tourism reviews of the two places. Based on this, this paper puts forward a series of targeted development strategies, including upgrading service level, innovating tourism products, strengthening infrastructure construction and utilizing modern marketing means, aiming to provide theoretical support and practical guidance for the sustainable development of Guangxi's tourism industry.

Keywords: Data Mining; Guangxi Service Industry; Development Strategy; Comparative Analysis

1. Introductory

1.1 Status and Significance of Tourism in Guangxi

As an important component of the modern service industry, tourism is playing a growing

role in promoting economic development, employment and improving people's livelihood. Under the background of globalization and informationization, the competition in tourism is not only limited to the traditional service quality and attraction of attractions, but also involves the innovation of tourism products, the adjustment of marketing strategies, and the overall improvement of tourism experience. As a country with rich tourism resources, China's tourism development has been emphasized by the government and all walks of life.

In this study, Guilin City, which is located in the northwestern part of Guangxi, is selected as an example to be analyzed. Guilin City Exploration Tour has been listed as one of the top ten tourism boutiques in Guangxi, Guilin City is a famous cultural city with a history of more than 2,000 years. From ancient times to the present day, successive generations have left a splendid cultural heritage here. Among them, the prehistoric culture represented by Zhenpi Yan, the ancient city construction culture represented by the ancient city pattern of Song Dynasty and Ming Dynasty, the ancient water conservancy science and technology culture represented by Lingqu and Xiangshidi, the landscape culture represented by the famous mountains and rock carvings, the culture of the Ming Dynasty feudal lords represented by the Jingjiang King's City and King's Mausoleum Tomb Complex, and the modern culture represented by the recent and modern revolutionary relics and historical monumental sites are the essence of Guilin's history and culture. The essence of Guilin's history and culture. There are a total of 552 cultural relics and monuments within the urban area, and 117 of them are listed as cultural relics protection units at all levels, among which there are 5 at the national level, 23 at the autonomous region level, and 89 at the city (county) level. Therefore the study of tourism in Guangxi is of great significance. The survey

about tourism in Guilin has also been a concern and research work for the relevant departments. In this study, we choose to obtain data by using online comments and mine and analyze the data obtained, in order to fill the gap in the study of tourism online comments in Guangxi.

In recent years, with the development and maturity of network technology and the continuous emergence of travel websites, there are more and more registered users of travel websites. Especially since the Prime Minister's government work report in 2015 put forward the strategy of "Internet+", "Internet+Tourism" has also developed rapidly. The website portals such as Ma's nest, Ctrip, Touniu, Where to go and so on are more and more popular among the masses, and the functions of the tourism website are also more and more perfect, and many tourists will publish their own experience and feelings on the corresponding website after the journey, and evaluate the tourist places. Due to the openness of the network, a huge amount of travel comments will be generated, and these comments contain a lot of valuable information. On the one hand, tourism enterprises and tourism-related management departments can take advantage of the feedback information from tourists' comments to continuously improve the development of the local tourism industry. On the other hand, tourists can make better decisions and journey plans through the content of existing tourists' online reviews. How to find out the potential and hidden information and value from the text content of tourists' comments has become an urgent problem for tourism websites and local tourism related departments.

Data mining technology provides great help to solve the above problems, on the one hand, it can batch access to tourists' network comment data, on the other hand, it can also process and analyze the unstructured network comments, and finally draw the corresponding conclusions, so as to achieve the research on the development of Guangxi's tourism service industry from the comments of tourists.

1.2 Current Status of Domestic and International Research

At present, domestic and foreign research on tourism network reviews based on data mining technology has achieved a lot of results, but

the research on tourism network reviews of Guangxi service industry based on data mining technology has not been found yet. Foreign related studies mainly include Robert et al. who used artificial neural network tools to analyze the content of the image of seven tourist destinations Robert et al. used artificial neural network tools to analyze the image of seven tourist destinations [1]; Choisis et al. also used the content analysis method to conduct a comparative study on the image of tourist destinations reflected in the online reviews of Macau tourism [2]; Hunter Wc conducted a comparative study on the image of tourist destinations in Seoul by analyzing the content of both online information and traditional print media [3].

Domestic related research is mainly reflected in the following four aspects: First, according to the content of tourists' online comments on tourist satisfaction analysis, such as Sun Xiaopei used content analysis as a bridge to transform online comments into tourist satisfaction, realizing the transformation from qualitative data to quantitative data [4] Zheng Jun et al. based on the tourism demand template, constructed a model for analyzing the public opinion and satisfaction of scenic spots, providing more specific satisfaction values for understanding the various related contents of scenic spots [5]; Wang Shiyi et al. used Octopus crawler software to crawl the visitor evaluation data from OTA platforms such as Ctrip, Meituan, and Mabee's Nest as the data samples, and with the help of the ROST CM6 and ROST EA software, they got the visitor satisfaction on the scenic spots in different areas and analyzed and proposed the construction suggestions. Satisfaction and analyze and put forward construction suggestions [6]; Secondly, the theme analysis of tourists' online comments, such as Zheng Xianwei and Zhang He used the extraction of theme words, semi-supervised clustering processing, and established the LDA theme model for attraction comments [7] Yaru Wu used the tourists' comments from four tourism websites and used ROST CM software to analyze the image composition, emotional image and overall image from three parts based on the rooting theory [8]. analysis [8]; Thirdly, the sentiment analysis of tourists' online comments, such as Yu Jing selected the microblogging data released by tourists in

Xi'an city, and based on the sentiment lexicon to the data for sentiment analysis, but also with the help of ArcGIS software to visualize the emotional spatial-temporal characteristics of the tourists' comments [9]. In addition, He Haiyu used a crawler to crawl a large number of users, media, official media, and big Vs in various fields of microblogging, and visualized the data with the help of SonwNLP library for sentiment analysis of the text [10]; Fourthly, based on the network public opinion caused by emergencies, some problems of the tourism industry were studied, such as the case of Phoenix Ancient City Ticket Incident of Hunan Province by Wang Chao, Luo Keren, and using the statistics of high-frequency words, social network semantic analysis and other text mining methods to analyze the network public opinion text generated by the event [11]. In addition, Wang Chao also took the "Qingdao overpriced shrimp" rip-off incident that happened in 2015 as an example, analyzed the content of the network text formed by the heated discussion triggered by the incident on the Internet, and put forward countermeasures and suggestions on tourism catering prices based on the conclusions obtained [12]. Zhou Xuefang takes the incident of female tourists being beaten and disfigured in Lijiang, Yunnan Province, as a representative case for research and analysis, which generates an analysis of the relationship between negative network public opinion and tourists' willingness to travel, and obtains suggestions and countermeasures to deal with similar incidents [13].

In summary, using data mining techniques to research on tourism network comments has gradually become the focus and hotspot of academic attention, at the same time, academic data mining techniques for network comments are also improving, but there is a lack of research using data mining techniques to specifically research on Guangxi service industry tourism network comments, especially the application of text data mining techniques in Guangxi service industry tourism network comments is still a gap in the research. especially the application of text data mining technology in Guangxi service industry tourism online review is still in the blank.

2. Guangxi and Harbin Tourism Review Text Crawling

In order to conduct data mining analysis of tourism review texts of Guangxi Guilin and Harbin, we chose online tourism platforms such as Ctrip.com, Where to go and Hornet's Nest as data sources. User reviews and ratings on these platforms can provide information about tourist attractions, hotels, restaurants, etc., which is an important reference value for studying the development status and problems of tourism in Guangxi and Harbin. Data mining was first carried out on Ctrip.com (www.ctrip.com), which is one of the well-known online travel platforms in China, providing services such as hotel booking, air ticket booking, travel and vacation. On Ctrip.com, users can evaluate and score hotels, attractions, restaurants, etc., and these comments and scoring data can be used as a data source for the study. GoWhere.com (www.qunar.com) is another well-known online travel platform that provides similar services. Users can also evaluate and score travel products on GoWhere.com, and these data can be used for research and analysis. Hornet's Nest (www.mafengwo.cn) is a platform that focuses on travel tips and travelogue sharing, where users can post their travel experiences and reviews. These travelogues and evaluations can provide more detailed and nuanced information, which can be very helpful for research.

In order to obtain the comment text data on the above platform, we use the text crawling method. The specific steps are as follows:

Determine the crawling target, we need to determine the specific objects to crawl such as attractions, hotels or restaurants, for example, a famous attraction in Guangxi or a special restaurant in Harbin. Determine the crawling rules: According to the structure of the target web page, we need to analyze the HTML code of the web page to determine where the comment text is located and the related tags. By looking at the source code of the web page, we can find the tags and attributes where the comment text is located. Writing a Crawler Program Based on the crawling rules, we can write a crawler program using a programming language such as Python. The crawler program can automatically visit the target web page and extract the comment text data by simulating the browser behavior. Data cleaning and organization, crawling down the data may contain some irrelevant information or noise,

need to carry out data cleaning and organization. The main use of Excel in the "sorting and filtering" "to remove duplicates" and other functions. For example, HTML tags, special characters and duplicate text can be removed to retain keywords and ratings and other useful information. Data storage: The cleaned and organized data can be stored in a database or text file for subsequent data analysis and mining.

3. Comparative Analysis of Tourism between Guangxi and Harbin

3.1 Characterization of Online Comments

In order to intuitively and clearly explore the features embedded in tourists' online comments about Guilin and Harbin tourism, word cloud diagrams were used for visual analysis.

Table 1. Guilin's Visualization and Analysis based on Word Cloud Map

byword	word frequency	byword	word frequency	byword	word frequency	byword	word frequency
stepped fields	753	pub (public house)	70	(mountain) peak	38	idyllic	24
scenery	436	line up	70	park (for public recreation)	38	course of events	24
River Li, Guangxi	434	magnificent sight	70	color	37	slow business season	24
Guilin prefecture level city in Guangxi	409	guidebook	69	vistas	37	recommend	24
scenic area	356	matrix	67	course (of history)	37	liable (to)	24
ridges of the dragon	352	climatic	66	picturesque	36	sino	24
Long River	341	the whole distance	66	beautiful	35	in the end	23
bamboo raft	316	prices	65	sunrise	34	sense of experience	23
landscaping	283	influential family	58	workers and peasants	34	parking fee	23
Yangshuo prefecture level city in Guangdong	261	pneumococcal	58	placement	33	elephants	23
place of interest (tourism)	209	Ctrip (Chinese company)	57	nature (the natural world)	33	boat builder	23
room	178	mountain road	56	Elephant Mountain	33	beautiful	23
timing	173	quality-price ratio	54	transportation	32	free	23
drift along or about	147	until 1959, Guangxi province	54	global	32	golden yellow	23
Elephant trunk hill in Guilin	143	photographs	53	friends	32	moods	22
Sansui, Japanese company	139	guesthouse	53	major	32	cruise ship	22

Table 2. Harbin's Word Cloud Map-based Visualization and Analysis

byword	word frequency	byword	word frequency	byword	word frequency	byword	word frequency
Harbin	906	memorial tower	47	European	27	Paris, capital of France	19
churches	492	suggestion	45	flavor	26	center	18
constructions	491	Byzantine or Eastern Roman empire (395-1453)	44	Western-style food	26	nationwide	18
Central Avenue	475	skyline	43	exoticism	26	step	18
ice and snow	220	magnificent sight	42	transportation	25	prosperous	18
line up	212	length of time	41	cultures	25	city center	17
global	190	an excellent	41	very good	25	synthesis	17
ice sculpture	183	architectural style	40	too cold	25	shopper	17
Sofia (capital of Bulgaria)	176	(euphemism) go to the toilet	39	typical case	24	prices	17
(children's) sliding board	175	alien land	38	matrix	24	merchandise	17
Harbin prefecture level city in Inner Mongolia	135	trade	38	major	24	snacks	17
car-free zone	132	ice lantern	38	Xinyang prefecture level city in Henan	24	store	17
piazza	120	busy	38	travel sight	23	functionality	16
pretty good	113	bustling with noise and excitement	37	friends	23	a domed roof	16
Georgia	113	recommend	37	scenic area	23	guidebook	16
histories	111	look around	36	snow sculpture	23	gastronomy	16
place of interest (tourism)	98	supermarkets	36	shopping streets	23	ad hoc	16
observation wheel	96	Songhua River	36	museums	22	ambitious	16
sports event	91	various birds of the species Columbidae	35	vote rigging	22	general	16
nice	91	greatest	35	kid	22	the Far East	16
sino	85	not the same	34	site developed for a group of related enterprises	22	mold-making	15
iconic	83	properly	34	acceptable	22	accompany the troops	15

quality-price ratio	76	bakery	33	delicious	22	exterior condition	15
ice lolly	76	somewhat	32	onion head	21	greener	15
artists	74	well worthwhile	32	east and west	21	climatic	15
wind force, direction etc	72	Western-style restaurant	31	pub (public house)	21	chocolates	15
cordial	71	concert	30	Southerner	21	(regional) specialty	15

According to the Ctrip tourism website, tourists' comments on the popular attractions in Harbin crawled and organized to get Table 2, we can see that most of the comments are positive, according to 2214 comments through the Python language to draw the word cloud diagram Figure 2, the word cloud diagram according to the frequent degree of occurrence of the word show.



Figure 2. Word Cloud of Valid Comments in Harbin

According to the word cloud map, you can see the popular amusements as well as famous locations in Harbin. Harbin has a number of exotic attractions, and tourists enjoy traveling in winter by visiting ice and snow, ice sculptures, popsicles, and other ice-related recreational activities and food.

3.2 Thematic Analysis

The size of the text in Figures 1 and 2 reflects the frequency and importance of the corresponding words in the comments, the larger the font size, the higher the frequency of occurrence and the higher the importance, and the smaller the font size, the lower the frequency of occurrence and the lower the importance.

In the following section, we compare the comments with Figure 1, Table 1, Figure 2 and Table 2 to analyze the characteristics of Guilin and Harbin online comments from the perspective of the six elements of tourism (i.e., food, lodging, travel, tourism, shopping, and

recreation), and analyze the characteristics of Guilin's service industry.

Guilin:

Eat: Breakfast, restaurant, buffet, cooking, Guilin rice noodles.

Living: Hotel, Room, Bathroom, Elevator, Air-conditioning, Hot-water, TV,

Walking: Parking lot, boat ride, transportation, bus station, mountain road.

Tour: attractions, pier, air, terraces, scenic, view, ticket, jinkeng, cable car, raft, landscape, guide, nature, cruise, service.

Purchase: Price, value for money, price point.

Amenities: Springs, photo shoots, river rafting.

Positive words: peaceful, convenient, beautiful, spectacular, clean, clear, beautiful, hygienic, breathtaking, world-class, famous

Pejorative: average, worthless, fee, inconvenient

Harbin:

Eat: Restaurant, Northeastern cuisine, iron pot stew, pot roast, frozen pears, popsicles.

Living: Hotel, Room, Bathroom, Elevator, Air-conditioning, Hot-water, TV

Walking: Parking, bus station, walking, bus, Middle East Railway.

Tour: Attractions, churches, air, ice, scenic spots, tickets, Ferris wheel, Daoli district, Songhua River, scenery, guided tours, European architecture.

Purchase: Price, value for money.

Amenities: Slides, photo shoots, ice sculptures, skiing.

Positives: value for money very fun exotic beautiful prosperous fun very beautiful special recommendable

Pejorative: too cold, queuing

In general, the characteristics of Guilin and Harbin have differences and similarities, should be reasonable use of their respective characteristics to promote and optimize, for Guilin, the terraced rice fields, rafting, landscape scenery, special food performance is the focus of tourism development, Harbin's

service has obvious advantages, Harbin hospitality to treat the southerners have a special friendly way to let tourists feel the service, while Guilin's service is more lack of. The service of Guilin is more lack of, for foreign tourists rather less friendly, affecting the development of tourism in Guilin.

4. Problems in the Development of Tourism in Guangxi

After years of development, Guangxi tourism industry has a basic scale, but there are still a lot of restrictions on the Guangxi tourism industry to enhance the quality and efficiency of the problem, mainly reflected in the following aspects.

4.1 Inconvenient Travel

As most of Guangxi is a terrain of interlocking mountain ranges, the terrain is relatively rugged, leading to the existence of an inadequate transportation network in some areas, especially in remote areas or around tourist hotspots. The construction of roads, public transportation, airports, and other transportation facilities may be inadequate, and the pejorative term inconvenient appears in tourists' evaluations of Guilin, reflecting the inconvenience of tourists' travel in Guilin.

4.2 Insufficient Professional Quality of Tourism Service Employees

The appearance of words such as tour guide, rafter, service and charge in the key words of Guilin reflects that the quality of tourism practitioners is closely related to the degree of goodwill of tourists. In Guilin bamboo rafting, there is the situation of boatmen receiving tips, because most of the projects of tourism practitioners are from the local boatmen, no professional training, lack of professionalism, professional tourism practitioners knowledge, business level is not high, the lack of good sense in the service attitude of raw, unable to provide good service to tourists, the quality of service is difficult to guarantee.

4.3 Lack of Rational Planning

Through the word cloud map and crawling tourists' comments, it can be seen that Guilin has fewer kinds of amusement programs, rafting, ropeways, boat trips, taking pictures, and many attractions have the same amusement programs and lack of related

personnel to promote them, so travelers lose interest after one trip. Scenic area managers do not have reasonable planning for the development of attractions, once developed and then abandoned, the development of tourism stagnation.

5. Conclusions and Implications

5.1 Findings

This paper analyzes and researches the tourism comments published by tourists in Guilin and Harbin by crawling them, and analyzes the research on the development strategy of digital transformation of Guangxi's service industry through the relevant comparisons between the participles of the comments and the word cloud diagrams, and draws the conclusions there:

Guilin and Harbin have different industrial structures and a large difference in the level of economic development, which affects the attractiveness to tourists, and Guilin has a greater potential for service tourism and lacks rational planning.

By crawling the data, data analysis, found that Guilin's keywords appear in the charge tips, etc., Harbin is almost none, through the understanding of the Guilin, found that there are a lot of indiscriminate charges, differential treatment of tourists, tourism service practitioners are not enough professional quality, resulting in Guilin's low wind rating.

Tourist satisfaction of scenic spots, through the quality of service and transportation accessibility and other impacts, different advantages and disadvantages, high satisfaction can attract more tourists and word-of-mouth publicity. There are many tourists are not satisfied with the transportation in Guilin, which is slightly inconvenient to travel.

5.2 Research Implications

By analyzing the hotness of scenic spots, the tourism characteristics of their respective regions can be identified. This will help the government and enterprises to be more targeted in their promotions, launch rational planning, emphasize their unique culture and natural resources, and achieve differentiated development.

Tourists' higher satisfaction with Harbin may be due to better service quality. Guilin can learn from Harbin's successful experience by

strengthening service training and improving the professional quality of tourism service practitioners in order to improve tourist satisfaction.

Guilin is slightly inconvenient in transportation because of the terrain influence, Guilin can analyze and improve the inconvenience in transportation according to the geographic environment and the attractions around to the tourists' path of play, to attract more tourists to Guilin.

Acknowledgments

This thesis was funded by the university students' innovation and entrepreneurship project "Research on Digital Transformation Grouping Path of Guangxi Service Industry". Project No.: 202210595037.

References

- [1] Govers R, Go F M, Kumar k. Virtual destination image: a new measurement approach Uj. *Annals of Tourism Research*, 2007, 34 (4):977-99721.
- [2] Choi S, Lehto X Y, Morrison A M. Destination image representation on the web: content analysis of Macau travel related websites. *Tourism Management*, 2007, 28 (1):118-129.
- [3] Hunter W c. The social construction of tourism online destination image: A comparative semiotic analysis of the visual representation of Seoullu *Tourism Management*, 2016, 54 (2) 221-229.
- [4] Sun Xiaopei. *A Study on Destination Tourist Satisfaction Based on Online Reviews*. Shanghai: East China Normal University, 2011.
- [5] J. Zheng, J. Lou. A method for analyzing opinion satisfaction of scenic spot evaluation data based on tourism demand template. *Computer Age*, 2017(3):62-64+67.
- [6] Wang Shiyi, Wu Zhongjun, Cao Hongli. Research on tourists' satisfaction in Li River scenic area based on big data mining technology. *Journal of Southwest Forestry University (Social Science)*, 2023, 7 (06): 81-88.
- [7] Zheng Xianwei, Zhang He. Application of LDA topic extraction model in internet travel reviews. *Industrial Control Computer*, 2014, 27:92-94.
- [8] Wu Yaru. Research on optimization strategy of theme park tourism image based on network text analysis. *Zhengzhou University*, 2022. DOI:10.27466/d.cnki.gzzdu.2022.002662.
- [9] Yu Jing. Research on tourists' emotion and spatial and temporal changes based on microblog big data. Xi'an: Shaanxi Normal University, 2015.
- [10] He Haiyu. Research and Implementation of Sentiment Analysis of Online Comments under Big Data. *Computer Knowledge and Technology*, 2023, 19(18):64-66. DOI:10.14004/j.cnki.ckt.2023.0852.
- [11] Wang Chao, Luo Keren. Research on inclusive development of tourism based on online public opinion - A case study of the Phoenix Ancient Town ticket incident in Hunan. *Economic Geography*, 2014, 34 (1):161-167.
- [12] Wang Chao, Wang Zhizhang. Research on the price supervision mechanism of tourism catering based on online public opinion a case of Qingdao's sky-priced shrimp incident. *Price Monthly*, 2017, (2):34-37.
- [13] Zhou Xuefang. Research on the influence of negative online public opinion on tourists' willingness to travel. *Yunnan University*, 2023. DOI:10.27456/d.cnki.gyndu.2021.002183.