Analysis of Online Reviews of Tourists in 5A Scenic Spots Located in Sichuan Province based on Text Mining

Qinglin Luo[1a], Sheng Zhong[1]*, Zhaocai Li[2b]

[1]*Sichuan University Business school Chengdu, China*
[2]*Chengdu University of Technology College of Management Science Chengdu, China*
[a]*luoqinglin@stu.scu.edu.cn*
[b]*1415195900@qq.com*
[*]*Corresponding author: szhong666@163.com*

Abstract

How to efficiently process and analyze online reviews of tourists has become one of the key issues for tourism operators today. This helps to identify user concerns and emotional tendencies to a large extent. More importantly, the knowledge gained from data mining can be used to improve the services of the tourist attractions and optimize the quality of the services. In this paper, we use natural language processing technology to analyze online reviews of tourists in 5A scenic spots of Sichuan Province, and determine the positive and negative polarity of tourists' online reviews. Through text mining of tourists' online reviews, the concerns and tendencies of tourists in the process of touring are analyzed. In this way, relevant suggestions and measures are proposed to scenic spot managers and tourists in order to optimize scenic spot services and improve the accuracy of tourists' tourism decisions.

Keywords*-text mining; online reviews; sentiment analysis.*

1. Introduction

Nowadays, the rapid development of information technology has made the Internet one of the standing tools in the daily life of modern people, and thus the Internet has become the main medium for people's daily operation. Tourism, as a major contributor to the tertiary sector, has gradually become an important focus of public research, and the quality of tourism destination services has become a top priority. Improving the quality of services in tourist destinations will help achieve the goal of a modern service industry that satisfies the people. At present, there exists a large amount of tourists' online review information on major tourism websites, and deep text mining of these unstructured information to explore tourists' concerns and tendencies towards scenic spots has become one of the important focuses of tourism research.

With the large-scale development of the Internet, text data has become one of the main forms of tourism big data. The main way to express the opinions of tourists is through online reviews. Therefore, text mining on these data is very helpful to the development of tourism [1]. Vinay et al. tried to study online reviews of heritage hotel customers through text mining methods. They revealed the behavior patterns and the factors of satisfaction and dissatisfaction of customers [2]. Dong et al. used text mining methods to design a tourist preference model in Sanya [3]. Combining the tourist preference classification research with the extracted features, they used the promoted emotional tendency formula to calculate the emotional tendency of each attribute. Besides, they analyzed the specific tourist preference and the corresponding satisfaction. Based on Tripadvisor.com's online reviews, Prameswari et al. used text mining methods and aspect-based sentiment analysis to obtain sentimental comments from hotel users [4]. They studied the application of the recurrent neural tensor network (RNTN) algorithm, which is usually used for sentence-level sentiment classification. In addition, the output of the algorithm can be used for evaluation to improve the quality of the hotel industry and support the tourism industry in Indonesia. Prameswari et al. combined two techniques in text mining, namely sentiment analysis and text summarization, to describe hotel conditions in tourist areas with different levels of

G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 1412–1416, 2023.
https://doi.org/10.2991/978-94-6463-042-8_205

tourism development, thereby helping to improve the quality of the hotel industry and supporting Indonesia's tourism industry [5]. Taking Wanlu Valley ecotourism area in Guangdong province as an example, Zhang collected Baidu index and mined the post-consumer evaluation texts of five tourism websites, such as Tongcheng and Ctrip, by using ROST Content Mining software [6]. By mining the high-frequency feature words in the tourism evaluation texts, a socio-semantic network matrix map was constructed, and then the tourism network attention index and tourist evaluation perception information were comprehensively analyzed.

In this paper, based on the tourists' online reviews of 5A scenic spots of Sichuan Province, we exploit natural language processing technology to perform text mining on these reviews. The valuable information implied in tourists' online reviews is mined. In addition, tourists' concerns and preference points for tourist destinations are learned, and the influencing factors of tourists' choice of tourist destinations are also explored. On this basis, tourists' concerns and preferences are presented to scenic spot managers, which is conducive to improving scenic spot products and optimizing scenic spot services.

2. Materials and Methods

In this paper, we apply the natural language processing tool ROST CM6 for text mining of online reviews of tourists in 5A scenic spots of Sichuan Province, and the main preliminary work involved is as follows:

(1) Removing duplicate corpus

Due to some subjective or objective reasons, there will be a large number of duplicate corpus of review information of some things or products. However, these duplicate corpus will have some influence on the information mining and analysis. Thus, we employ an Excel tool to filter the corpus, remove the duplicate corpus, and finally obtain more than 26,000 online reviews from tourists for subsequent analysis.

(2) Removing short sentences

In some shorter online reviews, a few less descriptive sentiments information is included. To prevent any influence on the subsequent sentiment determination, this paper uses an Excel tool to count the strings of each tourist online review and filter out the short sentences. According to the results of previous studies, the data with review length less than 30 characters are removed in this paper.

(3) Removing deactivated words

Words that do not have actual meaning in the text are called deactivated words. In the text mining process, these words can affect the experimental accuracy and word frequency analysis. In order to prevent these words from affecting the subsequent experimental operations, these words need to be screened and removed from the operation.

(4) Removing Internet words

Since online reviews of tourists come from the Internet, tourists will more or less add some emoticons and Internet phrases when posting online reviews. Some online phrases will have an impact on the subsequent information mining work. For this reason, the emoticons and non-Chinese strings in the dataset are filtered and removed in this study.

3. Results and Discussions

3.1. Overall analysis of tourists' online reviews

Sentiment is determined and counted for all tourists' online reviews using ROST CM 6 software, and the sentiment distribution of tourists' online reviews is plotted using in Figure 1, and the sentiment intensity is plotted according to the results in Figure 2.

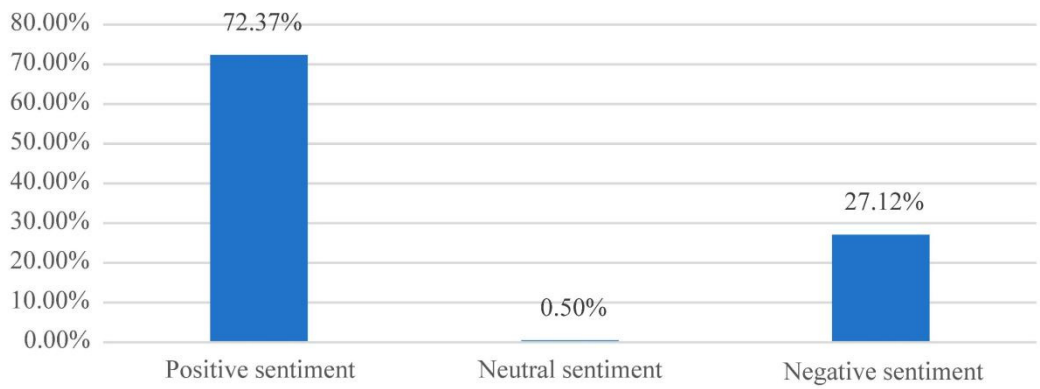

Figure 1 Sentiment distribution.

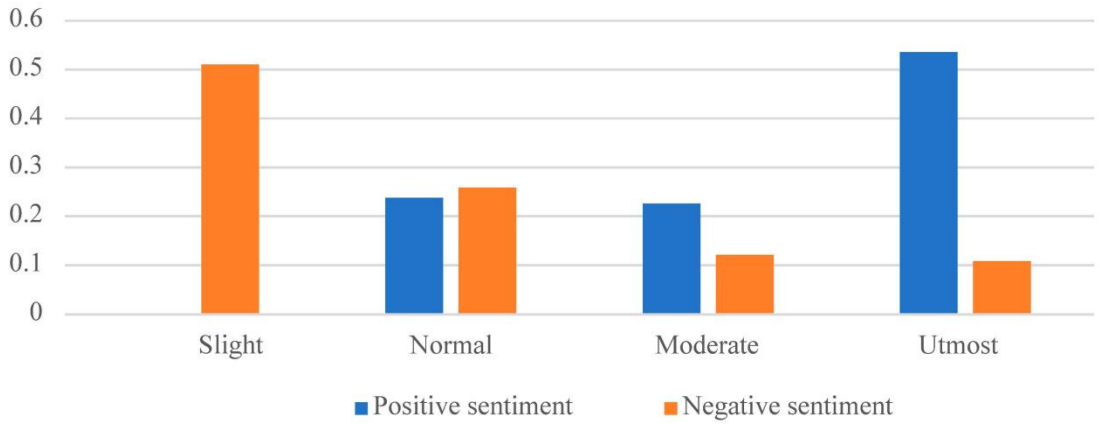

Figure 2 Sentiment intensity distribution.

From Figure 1, we can see that most of the tourists' online reviews are positive, only 27.12% of the tourists' reviews belong to negative reviews, which means that most of the tourists keep a happy mood to enjoy the whole journey during the tour. Positive emotion can make people keep agile thinking and vigorous energy, which can make tourists experience the whole tour pleasure perfectly during the tour and enjoy the whole tour perfectly. From the distribution of sentiment intensity in Figure 2, we can see that most of the positive emotions belong to the high-intensity part, most of the negative emotions belong to the mild intensity part, and the high-intensity negative emotions are only about 0.1. This means that most of the tourists can enjoy the services and products of the whole scenic spot with high-intensity positive emotions, and most of the tourists with negative emotions only have mild intensity emotions to enjoy the services and products of the scenic spot. Most of the tourists with negative emotions only enjoy the services or products of the scenic spot with slight intensity. For tourists with high intensity of positive emotions, scenic spot managers can improve their own scenic spot management system in all aspects, arrange all the projects of tourists in the scenic spot in an orderly manner, so that tourists can better experience the projects of the scenic spot and avoid affecting tourists' sense of travel experience because of order problems. For tourists with negative emotions, scenic spot managers can add more projects that drive less emotional sense in the early stage and more emotional sense in the later stage. In addition, a step-by-step guide for these tourists is also necessary. Good emotions towards will increase the tourists' sense of play experience.

3.2. Seasonal analysis of tourists' online reviews

Using ROST CM 6 software, all tourists' online reviews are divided into four seasons: spring, summer, autumn and winter for sentiment determination and counting, and then the Excel tool is used to draw the sentiment distribution of tourists' online reviews in four seasons, as shown in Figure 3, and the sentiment intensity graph is drawn according to the results, as depicted in Figure 4.

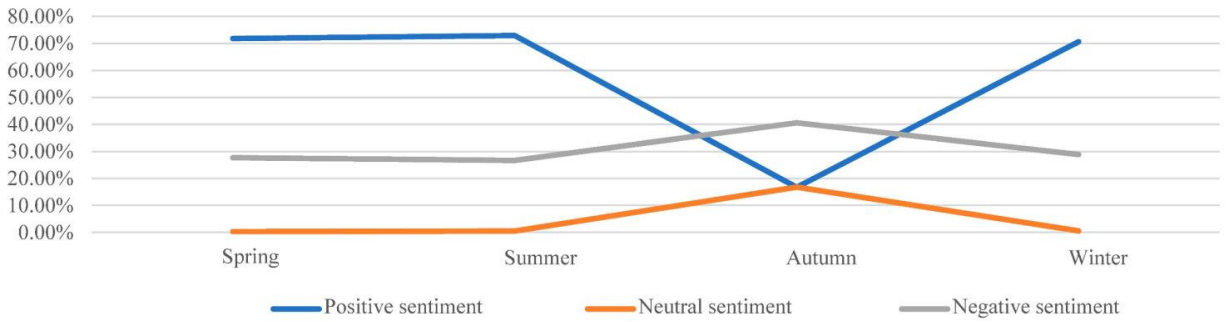

Figure 3 Distribution of sentiment in four seasons.

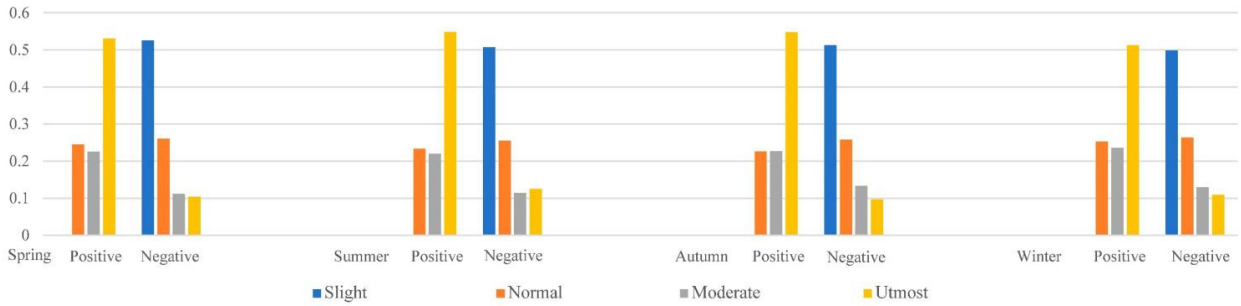

Figure 4 Distribution of sentiment intensity in four seasons.

From Figure 3, we can see that most of the positive emotions are in the two seasons of spring and summer, and most of the negative emotions are concentrated in autumn, indicating that most of the tourists who choose to travel in the two seasons of spring and summer have a good sense of travel experience during the journey, and can be better engaged in the pleasure of travel. In addition, a part of tourists who choose to travel in autumn will have a less satisfactory sense of travel experience and will become poorly impressed with a certain scenic spot. Therefore, the scenic spot manager can observe the specific tourism situation on site, summarize the reasons that will make tourists emotional, give the corresponding solution measures, and seriously implement the root cause of the problem. From Figure 4, we can find that the high intensity of negative emotions concentrated in the summer. Summer is the peak season of tourism, and the weather is hot, a little thing will stimulate the tourists' mood swings. More importantly, the high temperature can easily make tourists' negative emotions climb to the peak. Therefore, the scenic spot managers can open a few more lounges with clean environment and cool indoor air for tourists to rest and relax. Of course, the establishment of a few more scenic issues consulting points in the summer, to help tourists solve the scenic matters within their reach, can also allow tourists to better enter the state of play.

4. Conclusion

In this paper, we take the 5A scenic spots in Sichuan Province as the research object, and use the tourists' online reviews as the data source to conduct the text mining work. The tourists' concerns and focus in the process of tourism are analyzed, as well as what kind of landscape and landscape features each 5A scenic spot possesses. To this end, we summarize the following conclusions and suggestions.

(1) The scenic features of the 5A scenic spots in Sichuan Province vary and all have their own representative characteristics. Tourists should do a good job when choosing a destination, understand what they really want to see on this trip and what kind of travel experience they want to get. Clear travel goals allow tourists to save time costs while making their trip more meaningful.

(2) Among the information hidden in online reviews, convenience accounts for the largest proportion of tourists' senses, indicating that tourists pay more attention to the convenience of tourist attractions. Therefore, scenic spot managers can optimize the convenience of all aspects of things in the scenic spot. In addition, clearly communicating the convenience of the existence of scenic spots to tourists is more likely to drive tourists' travel emotions

(3) Tourists will have different emotional changes in the process of tourism, scenic spot managers can randomly investigate the factors that will cause tourists to produce emotional changes during the journey. And according to these factors, we can develop appropriate response strategies to stop or purify the tourists journey emotional triggering factors. In this way, more repeat tourists can be attracted and the sense of tourist experience is increased.

References

[1] Li Q, Li S, Zhang S, Hu J, Hu J (2019) A Review of Text Corpus-Based Tourism Big Data Mining. *Applied Sciences-Basel*, **9**.

[2] Chittiprolu V, Samala N, Bellamkonda RS (2021) Heritage hotels and customer experience: a text mining analysis of online reviews. *International Journal of Culture Tourism and Hospitality Research*, **15:**131-156.

[3] Dong L, Wang H, Zhang C, 2017. Analysis of Sanya 's visitors preference based on text mining of tourism forum. In: *Symposium on Service Innovation in Big Data Era. Hainan Univ, Serv Syst Engn Dept, Hainan, PEOPLES R CHINA*. 273-284.

[4] Prameswari P, Surjandari I, Laoh E, 2017. Opinion Mining from Online Reviews in Bali Tourist Area. In: *3rd International Conference on Science in Information Technology (ICSITech). Univ Pendidikan Indonesia, Bandung, INDONESIA*. 226-230.

[5] Prameswari P, Zulkarnain, Surjandari I, Laoh E, 2017. Mining Online Reviews in Indonesia's Priority Tourist Destinations Using Sentiment

Analysis and Text Summarization Approach. In: *8th IEEE International Conference on Awareness Science and Technology (iCAST)*. *Chaoyang Univ Technol, Taichung, TAIWAN*. 121-126.

[6] Zhang Y (2018) Research on the Promotion of Word of Mouth in Tourist Scenic Spots Based on Web Text Mining--the Case Study of Wanlu Valley in Guangdong Province. *MATEC Web of Conferences*, **173:**03060 (03064 pp.)-03060 (03064 pp.).