

Big data analysis based on the hotel online reviews

Haifei Qin^{a,*}, Xiaobo Ye^b

Chuxiong Normal University, YunNan 675000, China

^aqhfei@cxtc.edu.cn, ^byxb@cxtc.edu.cn

* Haifei Qin

Abstract: With the rapid development of network technology, self-service travel has already become more and more. Booking hotels with online reviews has become the choice of people. The paper takes the hotel's online review as the research object. From the research status of online reviews, according to the problem is put forward, analyzed and put forward technology route. Finally, it wills is a foundation that realize intelligent search for hotels and personalized recommendations for customers.

Key words: online review; semantic analysis; data cleaning

1. Introduction

With the advent of “online shopping”, “high-speed trains”, “mobile payment”, “shared bicycles” and network technology, self-guided travel has driven individual hotel guests to climb year by year and occupying half of the total hotel customers. According to the “2017 Ctrip Accommodation Data White Paper” data, from July 2016 to June 2017, the growth rate of hotel accommodation online demand was as high as 28.8%. 85% of tourists chose self-guided travel, and the national hotel room supply growth rate reached 8.2%. The number of newly added rooms exceeds 1.4 million, and its fastest growing provinces are: Qinghai, Hainan, Guizhou, Gansu, Yunnan and other western regions ^[1]. Facing the great transformation of customer's travel mode, the hotel accommodation industry has to pay more attention to the change of consumer behavior than ever before, and pay attention to their source, booking decision process, consumption preferences, occupancy experience and word-of-mouth reviews, etc. People pay more and more attention to online reviews. Taking TripAdvisor as an example, the online reviews have exceeded 500 million, increasing by nearly 30 thousand items per day, and the growth rate of Ctrip is faster.

2. Research status analysis

Online reviews are divided into digital ratings and online reviews. At present, many scholars focus on digital scores because the digital scores are relatively intuitive, easy to understand and analyze. However, the numerical scores are relatively coarse and small, and they are difficult to refine. For example, with the same hotel being rated as 5 points, customers feel different about it. Some are concerned about location, some are concerned about facilities, others are concerned about services, etc., different people have different concerns, interest points are different, and their evaluations are also different. From the perspective of related economics and market theory, products and services have multi-dimensional attributes. Due to the different preferences of consumers, the expectations of hotel functions and services are also different. That is, when users make reference to hotel reviews to make decisions, they will follow their preferences or only pay attention to some aspects of characteristics. Considering only the numerical score does not reflect the user's comprehensive and accurate evaluation of the product [2]. But online reviews can express the true feelings of the customers, and can also be referenced and trusted by the buyers. Online reviews, as a guest

experience on the internet, play an important role in other customers' purchase decisions. These experiences are also important information to be considered in the market expansion and product development plan of enterprises [2]. Therefore, online review data has been becoming more and more important.

At present, the research on online reviews is mainly based on sentiment analysis. It analyzes people's emotions and sentimental tendencies of certain product, and judge people's joy, anger, sadness, music, criticism, praise from online review, so as to identify the popularity of the product. Online review mining is an opinion mining, but it is not the same as sentiment analysis. Sentiment analysis is only part of the opinion mining. Opinion mining was first proposed by HatzivassiloglouV[3] in the ACL/EACL '97 meeting. In 2008, Pang Bo, Yao Tianfang [4][5] et al as to opinion mining research, technology, application and future challenges summarized. In 2012, Liu Bing [6] as to sentiment analysis and opinion mining in an article summarized on opinion mining involves the technology. In 2015, Ravi [7][8], Guellil et al further elaborated on the opinion mining. In 2016, Rana[9] reviewed the extraction technology. In 2017, Sun et al [10] and Li Jianhua et al [11] summarized in opinion mining on mining. In 2018, Han Zhongming et al [12] review on network comment level opinion mining methods. In view of the opinion mining, more and more people are interested in this. Online review research, as part of the point of view, has a strong emotional color. But in many studies, the research data is one of the few hotel online reviews, hotel online reviews data acquisition is the most convenient way, you can get from theTripAdvisor, Ctrip, meituan, dianping,micro-blog, WeChat and other sites.

Online reviews have become an important source of information in the consumer's decision-making process [13]. Studies have shown that if the product is recommended by others, the number of product selections will increase two times, depending on the type of source recommended [14]. Consumers increasingly seek peer experience when preparing to purchase a product or service, and more than 60% of consumers consult customer feedback before purchasing [14]. Accommodation reviews determine the hotels online word-of-mouth, sales and future income.

3.Technical analysis of online review data research

3.1 The question is raised

Data is the basis of data analysis, with the development of intelligent mobile phone and wireless technology, swarm intelligence sensing technology has become an important ways of understanding the space-time characteristics in hotel online reviews timely, users can use mobile phone publish and access to the text and image information about the hotel, and provide a good basis for their decision making. However, the online reviews data are sparsely and unilaterally. If we want to know the hotels in a comprehensive and accurate way, we must efficiently use the ability of the same group to perceive and acquire specific objects. Therefore, the perception and acquisition of hotel online review data is the basis of the online review of big data analysis.

Hotel online review data from different dimensions portrays the hotel's characteristics and customers' feelings. According to semantic mining can understand the hotel's overall characteristics and the customer's real experience, but the traditional semantic mining relies on a specific situation, lack of comprehensive application capabilities. In natural language, the analysis of text semantics is like a child learning language, starting with words, turning words into phrases, turning phrases into sentences, and then converting sentences into paragraphs, sections, etc. The semantic analysis of texts is just the reverse that it is the process of the chapter to the section, to the paragraph, to the sentence, to the phrase, to the word. The online review data does not have the typical structure of articles, chapters, sections, and paragraphs. Only sentences and even sentences are incomplete, so

the semantic recognition is difficulty. Therefore, the semantic analysis of hotel online review data is another key issue in hotel online data analysis.

At present, the system will provide time, locations, type, and price range when you select hotel online. But the demand of the customer cannot be embodied, unable to achieve intelligent and personalized recommendation real. According to the customer's past accommodation experience and hotel online reviews find out similar customer groups, and according to the similar experience to provide intelligent recommendation for similar groups of customers. Therefore, the intelligent search and recommendation of the hotel online review is another key problem in hotel online data analysis.

In summary, in the era of big data, online reviews data contains a large number of spatial information, to achieve accurate, interconnected, intelligent data, it is necessary to further the three key technologies: the hotel online reviews data perception and acquisition, semantic analysis of online hotel reviews, the intelligent search and recommendation of online hotel reviews data.

3.2 Technical route research

The hotel online reviews of big data mining proposed research program includes the hotel online reviews of big data perception and acquisition, semantic analysis and processing, intelligent search and recommendation. The general framework of the hotel online reviews as follows figure 1.

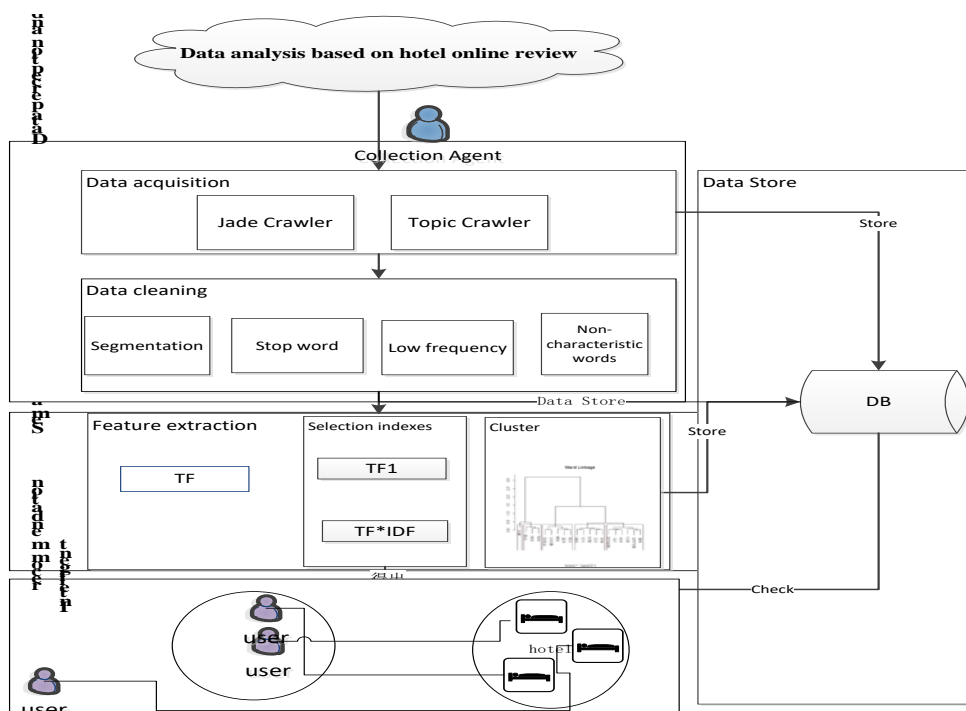


Fig. 1 Hotel online analysis architecture

Data perception and acquisition stage: the hotel online reviews of big data intelligence and can be used to obtain crawler technology. Based on Jade crawler technology and topic- crawler technology is the core technology of network data acquisition.

Semantic analysis and processing stage: includes data cleaning and feature extraction. Data cleaning is going to use text cleaning methods, segmentation, Punctuation, word stop, low frequency word, and High-frequency non-feature words. Feature extraction is based on word cloud, word frequency, word weight, document frequency, inverse document frequency, TF-IDF and other feature selection indicators. Feature selection is performed by cluster analysis.

Intelligent recommendation stage: we use decision tree, support vector machine and naive Bayes to classify the characteristics of online reviews data. Finally, we use correlation analysis to make initial matching between customer groups and hotel groups.

When a customer inquiry the hotel, according to the customer's query condition, historical residence record and interest in historical review, will quickly search out a hotel that is suitable for customers and personalize recommendation.

4. Conclusions

The paper based on the hotel online review data research. Analysis the status of the hotel online review, research technology, and put forward the method of the hotel online reviews analysis. According to the data cleaning, semantic analysis, and feature extraction. Finally, realize intelligent search for hotels and personalized recommendations for customers.

5. References

- [1]http://www.sohu.com/a/218388910_156111
- [2] Wu Weifang, Gao Baojun, Yang Haixia, Sun Hanlin. The Impacts of Reviews on Hotel Satisfaction: A Sentiment Analysis Method, *Data Analysis and Knowledge Discovery* 2017, 3(3):62-71
- [3] Ye, Q., Law, R., Gu, B. and Chen, W. "The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings", *Computers in Human Behavior*, 2011, Vol. 27 No., pp. 634-639.
- [4] Hatzivassiloglou V, Mckeown KR. Predicting the semantic orientation of adjectives. In: *Proc. of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conf. of the European Chapter of the Association for Computational Linguistics (ACL/EACL'97)*. Stroudsburg: ACL, 1997. 174-181. [doi: 10.3115/976909.979640]
- [5] Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations & Trends in Information Retrieval*, 2008, 2(1-2):1-135. [doi:10.1561/1500000011]
- [6] Yao TF, Cheng XY, Xu FY, Han W, Wang R. A survey of opinion mining for texts. *Journal of Chinese Information Processing*, 2008, 22(3):71-80 (in Chinese with English abstract).
- [7] Liu B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 2012, 5(1):1-167. [doi: 10.2200/S00416ED1V01Y201204HLT016]
- [8] Ravi K, Ravi V. A survey on opinion mining and sentiment analysis. *Knowledge-Based Systems*, 2015, 89(C):14-46. [doi: 10.1016/j.knsys.2015.06.015]
- [9] Guellil I, Boukhalfa K. Social big data mining: A survey focused on opinion mining and sentiments analysis. In: *Proc. of the 12th Int'l Symp. on Programming and Systems (ISPS)*. Washington: IEEE Computer Society, 2015. 1-10. [doi: 10.1109/ISPS.2015].
- [10] Rana TA, Cheah YN. Aspect extraction in sentiment analysis: Comparative analysis and survey. *Artificial Intelligence Review*, 2016, 46(4):459-483. [doi: 10.1007/s10462-016-9472-z]
- [11] Sun S, Luo C, Chen J. A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 2017, 36:10-25. [doi: 10.1016/j.inffus.2016.10.004]
- [12] Li JH, Liu GS, Lin X. Survey on sentiment orientation analysis and its applications. *Journal of Cyber Security*, 2017, 2(2):48-62
- [13] Li, G., Law, R., Vu, H.Q., Rong, J., Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral. *Tour. Manag.* 2013, 36 (3), 321-330.
- [14] Hong-gang Peng, Hong-yu Zhang, Jian-qiang Wang Cloud decision support model for selecting hotels on TripAdvisor.com with probabilistic linguistic information, *International Journal of Hospitality Management*, 2018, 68, 124-138