

# Exploring Customer Experience of Smart Hotel: A Text Big Data Mining Approach

Huangxiong Qi<sup>1</sup>, Rucong Mo<sup>1\*</sup>

<sup>1</sup>School of Economics and Management, Wuyi University, Jiangmen, Guangdong 529020, China

**Abstract.** In order to analyse the factors and dimensions that customers pay attention to smart hotels, this experiment selects the user reviews of five smart hotels on Ctrip as the research samples, and carries out network text big data collection, text pre-processing and topic mining through the relevant algorithms of Python programming language. The results show that customers' accommodation experience of smart hotel mainly includes five aspects: breakfast and transportation, staff service level, intelligent service, room environment, and room hardware facilities. Among them, the customer's attention to the intelligent services of smart hotels and the intelligentization of hardware facilities in guest rooms reflect the difference in customer experience between smart hotels and traditional hotels, which provides a certain reference for the optimization of hotel service levels.

## 1 Introduction

The vigorous development of online tourism industry not only has an impact on the consumption mode of tourists, but also changes the dissemination mode of tourism information. A large number of user generated content (UGC) has become an important information source and decision-making basis for consumers. Big data analysis and artificial intelligence also provide companies with opportunities to capture consumer preferences and identify market trends. Research on tourism big data has become a hot topic in today's academic circles [1-2].

With the rapid development of 5G technology, artificial intelligence, big data and other related fields, smart tourism has become a higher expectation for people to pursue leisure and entertainment activities, and it is a new trend of tourism development in the new era. Smart hotel is an indispensable part of smart tourism. Zhang et al. demonstrated that smart hotel was based on the new generation of information technology, in order to meet the personalized needs of tourists, provide high-quality and high satisfaction services, and carry out systematic, intensive and intelligent management reform on various resources, information, facilities and services in the hotel [3]. Scholars' research on smart hotel is still in the initial exploration stage. The research on smart hotel mainly focuses on customers' acceptance and satisfaction with intelligent equipment services. Some scholars also discuss the application and technology development of artificial intelligence from the perspective of hotel intelligence [4-5]. In terms of research methods, questionnaires or case studies are mainly used [6-7], but no relevant research on using big data to analyse customers' perception of smart hotel experience has been found. Creating a better

experience for customers has become a necessary means for merchants to attract customers and improve their own industry competitiveness in the era of experience economy [8]. At present, in the field of hotel research, scholars mainly focus on the research of customer experience in traditional high star hotels or ordinary hotels [1-2], while there are insufficient research on customer experience in the field of smart hotels.

In view of the importance of smart hotels in the future development of the hotel industry and the inadequacy of academic research on the customer experience of smart hotels, this study will select smart hotels as the research object and use Internet big data to study its customer experience. It aims to understand the difference between the customer experience dimension of smart hotels and traditional hotels', and provide theoretical support and decision-making basis for hotel companies to optimize and upgrade smart hotel services.

## 2 Materials and methods

### 2.1 Research design

Comprehensively considering the representativeness of the data and the feasibility of experiment, through combing the smart hotels in the current domestic hotel industry, five representative smart hotels in China are selected as the research objects, namely Hangzhou Flyzoo Hotel, Hangzhou Dragon Hotel, Shenzhen Leyeju Smart Hotel, Shijiazhuang Fenglinwan Smart Theme Hotel, Dalian Shanshui S Hotel. These hotels have a high degree of intelligent service and management, and are distributed in different cities, so they are representative. Taking "Ctrip", which has the highest active users of the travel

\* Corresponding author: [morucong@126.com](mailto:morucong@126.com)

app, as an example, this study collects tourists' comments of the above five smart hotels, so as to analyse tourists' accommodation experience of smart hotels.

## 2.2 Data collection

This paper uses Python to write relevant programs to obtain the user comment information about the above five hotels on Ctrip, including user nickname, user ID, check-in time, travel type and other basic user information, as well as the user's rating of hotel, comment content and other comment information. Considering the different opening time of each hotel, the huge difference in the cumulative amount of reviews, and the timeliness of reviews, the final decision is to select the review data from January 2019 to January 2021. The collection time is February 2021, and a total of 6939 reviews are collected.

## 2.3 Data analysis

This research uses Python programming language to write programs to process and analyse data. First of all, the data is cleaned, and the duplicate, meaningless and irrelevant information is deleted to improve the quality of text mining. Then, Jieba package is used to segment and stop the collected data. Secondly, perform descriptive statistics on the preprocessed information to illustrate the sample characteristics. Finally, based on Latent Dirichlet Allocation (LDA) model, the topic mining of comment text is carried out to analyse the dimensions that customers pay attention to the smart hotel. LDA is a topic mining model with term, topic and document three-level Bayesian probability as the core structure.

Before performing LDA topic analysis, the topic coherence measure (Formula (1)) was used to calculate the coherence score for quantitative evaluation, and the best number of topics was selected. The evaluation results are consistent with the experts' manual marking, which can better find meaningful or highly cohesive topics [9].

$$C(z; S^z) = \sum_{n=d}^N \sum_{t=1}^{n-1} \log \frac{D_2(w_n^z, w_t^z) + 1}{D_1(w_t^z)} \quad (1)$$

$S^z = \{w_1^z, \dots, w_N^z\}$  is the most representative set of high-frequency words in topic  $z$ ;  $D_1$  is the number of comments containing the word  $w_t^z$ ;  $D_2$  is the number of comments containing the words  $w_n^z$  and  $w_t^z$ . In order to avoid logarithm to zero, 1 is added to the numerator.

After determining the number of topics, use the genism package in Python for topic training, and abstract the topic words through the LDA algorithm to obtain the customer accommodation experience topic and the corresponding topic words, and dig out the dimensions of the customer's attention to smart hotel accommodation.

## 3 Results and discussion

### 3.1 Data pre-processing

Firstly, data cleaning is performed to eliminate short, repetitive, and irrelevant invalid comments to improve the quality of text mining. The number of valid comments obtained after data cleaning is 6,861, with a total of 260,099 words. Secondly, the data is processed by word segmentation, removing stop words and merging stop words.

### 3.2 Descriptive statistical analysis

The sample data analysis results show that the reviewers have a high praise rate for the five smart hotels, of which 82.2% of the reviewers gave 5 points (full score), 96% of them gave more than 4 points, with an average score of 4.8. From the perspective of travel types, business travel, family parent-child travel, couple travel and friend travel are the main types, accounting for 41.2%, 19.1%, 14.3%, and 10.0% respectively. There are also a small number of travel alone (5.6%), booking for others (2.4%) or others (7.5%).

Using Jieba package of Python to classify the part of speech of user comments, and carry out word frequency statistics, respectively analyze 25 high-frequency words of nouns and adjectives, and manually correct some words with inaccurate part of speech tagging. The results are shown in Table 1. High frequency words play an important role in understanding the semantic space of user reviews and tourists' experience [1]. It can be seen from Table 1 that high-frequency nouns such as "service", "room", "front desk", "breakfast", "environment", "facility" and "service attitude" are the most frequently used, which are the elements that tourists pay the most attention to hotel. And "good", "convenient", "clean", "enthusiastic", "hygienic", "intelligent", "comfortable" are the main feelings of tourists to the smart hotel.

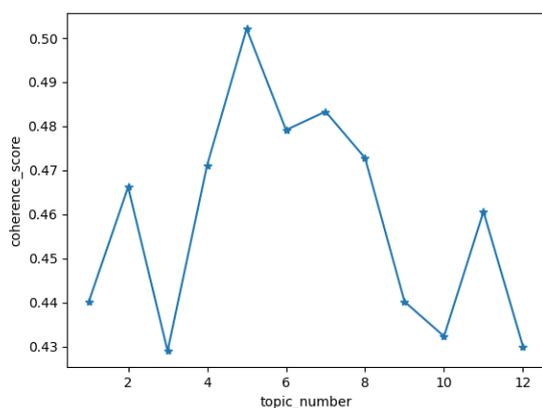
**Table1.** Statistical results of high-frequency words

No.	Noun	Frequency	Adjective	Frequency
1	service	2190	good	1949
2	room	2159	convenient	879
3	front desk	1695	clean	820
4	breakfast	1055	enthusiastic	568
5	environment	947	hygienic	404
6	facility	656	intelligent	364
7	experience	471	comfortable	323
8	service attitude	444	cozy	305
9	health	404	great	280
10	traffic	377	rich	262
11	feel	358	free	236
12	cost performance	319	very big	229
13	child	304	easy	205
14	upgrade	270	complete	177

No.	Noun	Frequency	Adjective	Frequency
15	waiter	260	warm	173
16	handle	253	general	158
17	location	242	tidy	157
18	attitude	189	well	125
19	lobby	181	spacious	92
20	washing machine	176	luxury	87
21	personnel	162	rich	82
22	price	159	patient	79
23	guest rooms	150	perfect	78
24	toilet	142	superior	72
25	train station	138	direct	66

### 3.3 Topic mining based on LDA model

This part mainly uses the Latent Dirichlet Allocation model to conduct thematic analysis of user reviews. Firstly, extract the nouns and adjective in the reviews by jieba package, and eliminate the words with high word frequency but no reference meaning, such as "hotel", "Hangzhou", "Ctrip" and so on. Secondly, build a dictionary and conduct word frequency statistics. Thirdly, call the gensim package and use topic coherence measure to calculate the coherence score. The higher the coherence score is, the more relevant the subject words in the topic are and the less ambiguous they are. The results are shown in Fig. 1. According to the maximum coherence score, the best number of topics is 5.



**Fig 1.** Topic number and coherence score

Finally, the gensim package in Python is used for topic training. Topic words are abstractly classified by LDA algorithm to obtain 5 topic categories of customer accommodation experience and the corresponding 10 subject words, which are sorted in descending order according to the posterior probability of occurrence. The results are shown in Table 2. Since the topic classification obtained from the experiment only has the topic number but no topic name, it is necessary to summarize and name each topic according to relevant topic words and logical relations. It can be seen from Table 2 that the five themes focus on different aspects, mainly including the breakfast and transportation, staff service level, intelligent service, room environment, room hardware facilities, etc.

Among them, topic 1 includes "breakfast", "location", "train station", "transportation" and other nouns, and adjectives such as "convenient", "good", "rich", which reflects customers' attention to the richness of the hotel breakfast and convenient transportation. The richer the breakfast provided by the hotel, the better the location and the more convenient the traffic, the higher the customer satisfaction. The key words in topic 2 include "service", "front desk", "environment", "service attitude" and other nouns, as well as adjectives such as "warm", "great" and "good", reflecting customers' attention to the service level of hotel staff. The better the service attitude and service level of hotel staff, the better the check-in experience of customers. In addition, the front desk of the hotel is an important place for customers to check in, as well as an important contact object during the stay. Therefore, the service level of the front desk staff has an important influence on the overall hotel service level of the customer's evaluation. Topic 3 mainly includes "child", "lovers" and other guests, as well as "facility", "robot", "theme", "price", "elf" (intelligent assistant) and other nouns, and "good" etc. adjectives, reflecting the intelligent services of the hotel. It can be seen that intelligent service devices such as robots and elves are more attractive to young people such as children and lovers, and they are more satisfied with the intelligent service. Topic 4 includes adjectives such as "clean", "hygienic", "comfortable", "convenient" and nouns such as "room" and "transportation", which is mainly reflected in customers' attention to room environmental conditions. Although topic 4 and topic 1 both mention transportation, they have different emphases. It can be seen that customers mainly value the hygiene and comfort of the room, while the convenience of transportation improves the satisfaction of customers. The keywords of topic 5 include "room", "cost performance", "intelligence", "washing machine", "facility", "toilet", "curtain", etc., which reflect the hardware facilities of the hotel room. At present, the intelligentization of hotels is mainly reflected in the hardware equipment conditions of the rooms, such as smart washing machines, smart toilets, automatic curtains and so on. Topic 5 fully reflects the customers' attention to the room hardware conditions. The more intelligent the hotel room hardware conditions are, the more attractive they are to customers.

**Table2.** Topic classification of smart hotel reviews

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
breakfast 0.114	service 0.172	good 0.224	clean 0.099	room 0.080
convenient 0.061	front desk 0.107	experien ce 0.028	room 0.098	cost performan ce 0.027
good 0.034	environm ent 0.085	child 0.028	hygienic 0.095	intelligen ce 0.026
rich 0.026	warm 0.043	feel 0.023	comfortabl e 0.035	washing machine 0.017
location 0.022	great 0.024	lovers 0.018	convenient 0.028	facility 0.014
train station 0.016	service attitude 0.02	facility 0.015	transportati on 0.028	toilet 0.014
transportati on 0.015	good 0.018	robot 0.011	equipment 0.027	equipmen t 0.010

West Lake 0.015	room 0.014	theme 0.010	cozy 0.022	curtain 0.009
child 0.014	attitude 0.013	price 0.008	easy 0.021	convenient 0.009
equipment 0.012	staff 0.013	elf 0.007	tidy 0.018	cozy 0.009

## 4 Conclusions

This experiment uses Python related algorithms to analyse the UGC text big data of smart hotels, and dig out the elements and dimensions that customers pay attention to smart hotels, so as to fill the gap in the research on customer experience of smart hotels and enrich the research perspectives and forms of smart hotels.

The results show that customers' accommodation experience of smart hotel is not only related to but also different from that of traditional hotel. Whether it is a traditional hotel or a smart hotel, the breakfast, traffic conditions, service level and the software and hardware conditions of the room have always been the important factors concerned by customers, which further confirms the existing research [1-2]. Therefore, if the hotel wants to improve customer satisfaction, it is necessary to further optimize the hotel's catering service level, staff's professionalism and service level, and fully provide convenience for customers in the hotel location or transportation connection.

What's more, the experimental study found that the intelligent service of smart hotel and the intelligent hardware facilities of guest room are also the focus of customers' attention to smart hotel. In other words, the better the hotel's intelligent service level, the more it can attract young people, who are the main consumers of the hotel now or in the future, so more attention should be paid. Therefore, the upgrading of traditional hotels and the further optimization of smart hotels should give full consideration to improving the hardware facilities of hotel rooms and other customer activity places, and constantly optimize the customer's accommodation experience, which is conducive to obtaining better customer reputation and improving the hotel brand image.

## Acknowledgments

This paper is the phased achievement of Guangdong Province Joint Training Graduate Demonstration Base Project (Project No. 503170060260), Undergraduate Teaching Quality and Teaching Reform Engineering Construction Project of Guangdong Province (Project No. GDJX2020006), Wuyi University Graduate Education Innovation Project (Project No. YJS-SFJD-20-05), and Wuyi University Undergraduate Teaching Quality and Teaching Reform Project Bidding Project (Project No. JX2020046).

## References

1. Xu, F., La, L., Zhen, F., Lobsang, T., & Huang, C. (2019). A data-driven approach to guest experiences and satisfaction in sharing. *Journal of Travel & Tourism Marketing*, 36: 484-496.

2. Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, 44: 120-130.
3. Zhang, L., Qiao, X., Huang, X. (2017) *Smart Tourism Theory and Practice*. Nankai University Publishing, Tianjin.
4. Chiang, A. H., & Trimi, S. (2020). Impacts of service robots on service quality. *Service Business*, 14: 439-459.
5. Leung, R., & Loo, P. T. (2020). Co-creating interactive dining experiences via interconnected and interoperable smart technology. *Asian Journal of Technology Innovation*, 1-23.
6. Kim, J. J., Lee, M. J., & Han, H. (2020). Smart hotels and sustainable consumer behavior: Testing the effect of perceived performance, attitude, and technology readiness on word-of-mouth. *International Journal of Environmental Research and Public Health*, 17: 7455.
7. Neuhofer, B., Buhalis, D., & Ladkin, A. (2015). Smart technologies for personalized experiences: a case study in the hospitality domain. *Electronic Markets*, 25: 243-254.
8. Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer experience creation: Determinants, dynamics and management strategies. *Journal of Retailing*, 85: 31-41.
9. Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In: *The 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh. pp. 262-272.