

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/334050773>

# How do tourists evaluate Chinese hotels at different cities? Mining online tourist reviewers for new insights

Conference Paper · May 2016

CITATIONS

3

READS

51

4 authors, including:



Yong Liu

Aalto University

69 PUBLICATIONS 2,233 CITATIONS

SEE PROFILE



Feng Hu

Zhejiang Normal University

19 PUBLICATIONS 443 CITATIONS

SEE PROFILE



Hongxiu Li

Tampere University

84 PUBLICATIONS 2,263 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Big Cities meet Big Data [View project](#)



Understanding Internet gambling and its relation to problems [View project](#)

Summer 5-27-2016

# How do tourists evaluate Chinese hotels at different cities? Mining online tourist reviewers for new insights

Yong Liu

*Department of Information and Service Economy, Aalto University, Finland, Yong.Liu@aalto.fi*

Thorsten Teichert

*Department of Information and Service Economy, Aalto University, Finland*

Feng Hu

*Chair of Marketing and Innovation, University of Hamburg, Germany*

Hongxiu Li

*Information Systems Science, Turku School of Economics, University of Turku, Finland*

Follow this and additional works at: <http://aisel.aisnet.org/whiceb2016>

---

## Recommended Citation

Liu, Yong; Teichert, Thorsten; Hu, Feng; and Li, Hongxiu, "How do tourists evaluate Chinese hotels at different cities? Mining online tourist reviewers for new insights" (2016). *WHICEB 2016 Proceedings*. 67.

<http://aisel.aisnet.org/whiceb2016/67>

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in WHICEB 2016 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

## How do tourists evaluate Chinese hotels at different cities?

### Mining online tourist reviewers for new insights

*Yong Liu<sup>1\*</sup>, Thorsten Teichert<sup>2</sup>, Feng Hu<sup>2</sup>, Hongxiu Li<sup>3</sup>*

<sup>1</sup>Department of Information and Service Economy, Aalto University, Finland

<sup>2</sup>Chair of Marketing and Innovation, University of Hamburg, Germany

<sup>3</sup>Information Systems Science, Turku School of Economics, University of Turku, Finland

**Abstract:** Affected by the factors like population, economic and geographic conditions, accommodation offerings are inherently different at different cities and characterized with their specific features. This heterogeneity is not limited to the supply-side but covers the demand-side as well. For instance, business and leisure tourists may favor different travel destinations. Therefore, the development of accommodation industry needs to match the evolving demands of tourists. In this study, we utilize text mining techniques to understand English-speaking tourists' likes and dislikes with regard to hotels at different Chinese cities. Based on the studying the titles of 96,089 English reviews collected from TripAdvisor, the study seeks to explore the hotel attributes that tourists discussed in their reviews with regard to their lodging experience at a specific city. In particular, tourists' complaints are studied through the use of low rating reviews. In addition, hotel features favored by tourists are identified via associating hotel attributes with sentiment-featured words like 'great' and 'good'. Finally, an overall perspective on customer reviews is visualized in co-occurrence maps of jointly used sentiments and key-words. The research findings offer city-level strategic insights for hotel management.

Keywords: TripAdvisor, word-of-mouth, content analysis, satisfaction, hotel

### 1. INTRODUCTION

Along with the development of Web 2.0 as well as the increasing accessibility of Internet, consumers are increasingly utilizing online platform to share their reviews and comments with regard to a particular product or service, which are termed as online word-of-mouth or reviews. Many specialized online platforms have emerged which are designed to facilitate the communication of online word-of-mouth on a specific product or service, such as lodging experience at hotels. These online platforms include the websites like yelp.com, booking.com and TripAdvisor. TripAdvisor for instance has about "350 million unique monthly visitors and more than 290 million reviews and opinions covering more than 5.3 million accommodations, restaurants, and attractions" [1].

Even though the importance of online consumer reviews has been broadly acknowledged [2], [3], it is still a challenge on how to reliably extract useful insights from the online user-generated contents, such as online reviews. An increasingly large volume of online reviewers makes it impossible for an individual to even read through these reviews. Note that it might be possible for an individual to recall and synthesize tens of or hundreds of online reviews to get a consolidated understanding. The task, however, becomes impossible when the quantity of reviews is substantially increased (e.g. tens of thousands of reviews), such as when synthesizing all hotel reviews at a specific city to identify the key features of the local accommodation sector. Similar challenge may also apply to large or chain hotels, which may host thousands of customer reviews.

Based on studying online reviews of hotels at five Chinese cities, the study offers a quantitative measurement approach to identify the hotel attributes that English-speaking tourists consider when spreading their word-of-mouth online. Foreign tourists constitute an important part of customers for both accommodation and tourism industry in China. In 2014, approximately 26.36 million foreigners visited China, and 20.81 million

---

\* Corresponding author. Email: Yong.liu@aalto.fi (Yong Liu)

of them stayed at least one night [4]. Therefore, it is important to a clear understanding on what are the main concerns of foreign tourists with regard to Chinese accommodation industry, in particular when it comes to a specific city. Note that an unhappy lodging experience may deteriorate the whole travel experience of tourists, which may consequently lead to their negative perception and word-of-mouth toward the tourist destination that they visited.

To this end, we applied a text mining approach to extract the words with the highest frequency in the reviews for the purpose of identifying the most-mentioned hotel attributes. Furthermore, by associating the hotel attributes with the most-used sentiment words, like ‘great’ and ‘terrible’, it is possible to estimate at what degree a specific hotel attribute is liked or disliked by tourists. Based on a use of a large volume of online user generated reviews, strategic understanding and actionable insights can be obtained with a high external validity. The rest of the paper is organized as follows. The next section reviews relevant literature, following with a introduction of research methodology. Results are discussed in Section 4. We conclude the paper and discuss the limitations of the research in Section 5.

## **2. LITERATURE REVIEW**

The increasing popularity of online user-generated content, like online reviews, has motivated a great amount of researchers to explore it for new customer insights. By coding and analyzing 100 complaints from TripAdvisor, Reference [5] showed that complaints tends to be accompanied by advice and recommendations. Reference [6] studied a sample of 80 hotel replies to online consumer complaints that were posted on TripAdvisor. They found that hotel replies “tend to be highly formulaic and conventionalized, with thanking and apologizing among the most common moves identified” [6, p. 54]. Reference [7] examined the relationship between hotel ratings and the destinations where the hotels are located. The study indicated that destination characteristics affect ratings of hotels. Applying content analysis technique to 373 reviews extracted from TripAdvisor, reference [8] identified 26 attributes that affect ecotourists’ satisfaction with their lodging experience. Reference [9] found that TripAdvisor ratings and the number of reviews positively relate to the average size of each online booking transaction. Apparently, many studies on online tourist-generated content are based on coding a relatively small amount of online reviews.

Furthermore, a number of studies investigated how the review characteristics (e.g. text readability) and reviewer characteristics affect the perceived value of reviews (e.g. helpfulness vote). For instance, reference [10] found that more readable reviews and the review expressing extreme sentiment are more likely to be voted as helpful. This finding is partly consistent with the work of [11], who show that extreme ratings (positive or negative) are perceived to be more useful and enjoyable in comparison to moderate ratings. By studying online customer reviews from Yelp.com, the study of [12] reported that messenger and message characteristics have positive influence on the perceived usefulness of reviews.

Apart from above-mentioned studies, we noticed that a number of methodological studies are focused on developing algorithm or the mathematical basis to extract topics from a large amount of online reviews [for a summary see. 13]. This study focuses on applying text mining technique to explore the key hotel attributes valued by foreign tourists instead of the technique itself, thus, these methodology-oriented studies are not discussed in the present research.

## **3. RESEARCH METHODOLOGY**

### **3.1 Data source and data collection**

The research data were gathered from TripAdvisor. Tourists’ reviews with regard to the hotels at five Chinese cities, which were posted before September 20, 2014, are collected. The five cities are Beijing,

Shanghai, Guangzhou, Hangzhou and Sanya. We detected the language of online reviews through the use of textcat-package in R [14]. Only English reviews are considered in the study, because English is an international language and is widely used by tourists. Online reviews written in other foreign languages, such as Japanese and Russian, are not included in this study to reduce the complexity of the data analysis. As a result, 96,089 reviews written in English are retained for data analysis.

**Table 1. Average rating on hotels at different Chinese cities by English-speaking tourists**

City	Sample Size	Mean rating	Standard deviation
Beijing	40458	4.05049	1.059491
Guangzhou	9974	3.86535	1.102769
Hangzhou	4224	3.97419	1.085302
Sanya	2713	3.96940	1.145799
Shanghai	38720	4.08463	1.037976

Among the five selected cities, Sanya is the southernmost city on Hainan Island of China. Unlike Beijing and Shanghai, which are both important tourism destination and economic center, Sanya is a famous holiday resort. As shown in Table 1, Beijing and Shanghai have a large amount of reviews from International travelers while Sanya has a relatively small amount of reviews.

In the data analysis, we only selected the title of reviews to do the text mining due to the following reasons. Detailed review texts tend to be very complicated, and may include many different sentences and use both positive and negative emotional words in relation to several different hotel attributes. In this regard, a use of the title of review for extracting important hotel attributes has several advantages which are analogously to utilizing key-words [15]. First, as the title of review tends to be the summarization of the whole review or the most important information that the reviewers want to share with others, the title of review tends to be short and more concise than the detailed review. Second, the title tends to contain the most important hotel attribute that reviewers considered, other than several possible attributes. Third, the title of review tends to have a relatively simple structure. This feature facilitates us to more precisely associate hotel attribute to sentiment words.

### 3.2 Data treatment

Before extracting the keywords from the text, the raw data of the titles of reviews were handled through the following steps. First, all the uppercase letters were transformed to be lowercase letters. Second, numbers were removed while punctuation was replaced by a space. Second, the influence of a negation word ‘not’ is controlled. Specifically, the word ‘not’ is frequently used to indicate a negation of a related statement, such as ‘not good’, ‘not bad’. Therefore, without properly handled, the presence of word ‘not’ in title will violate the interpretation of the sentiment words, such as ‘good’ and ‘bad’. To avoid this problem, we removed the space between the word ‘not’ and the word followed to create a new word. For instance, ‘not good’ will be transferred to be a single term of ‘notgood’. Stop words like ‘too’ and ‘very’ are removed through the use of tm-package of R.

### 3.3 ANOVA test

Based on the above data treatment, the distribution of keyword frequency in title is examined, which follows a power law distribution ( $Y = 368167X^{-1.408}$ ;  $R^2 = 0.9704$ ). This indicates that a few keywords have high frequencies, dominating the connectivity of the semantic network that represents review titles [16], [17].

The results of an ANOVA test indicate that international tourists rated the hotels at different Chinese cities in a different manner. Tukey’s HSD test is utilized to perform a pairwise comparison on hotels rating between every two Chinese cities. The results indicate that the mean ratings differ significantly between each pair of cities, except for the comparison between Sanya and Hangzhou (probably due to smaller sample size of those

two towns). Specifically, tourists are most satisfied with hotels in Shanghai, which is followed by Beijing. Guangzhou has a significantly lower rating on its hotels than the other four cities, even though Guangzhou has much more international tourists than Hangzhou and Sanya. This implies a need for improvement for the accommodation sector in Guangzhou.

#### 4. RESULT AND DISCUSSION

Appendix A presents a list of top 30 keywords, implying the most concerned hotel attributes for international tourists. Geographic names like Beijing, Shanghai are not meaningful to be examined in this study. Thus, they are not discussed in the following. Words which represent positive emotions like ‘great’, ‘good’, ‘excellent’ are included as top keywords in all five cities, implying a general positive evaluation of hotels.

A separate evaluation of negative reviews is executed in order to isolate drivers of dissatisfaction. For this purpose, we explored the top *noun* words that appear in the low-rating reviews (overall rating = 1 or 2) to detect the hotel attributes that perform poorly. Table 2 provides an overview on the attributes that lead to dissatisfaction. In the table only specific hotel-attribute-related noun words are listed. General items like hotel, experience and place, and adjective words like terrible and bad are excluded. Across the five cities, service is always listed on top for poor evaluations. This result stays in stark contrast to the list of top keywords of all reviews (Appendix A), where location is at a higher rank in four out of five cities (with the exception of Sanya). Thus, location seems to be a driving force in generating satisfaction, whereas service seems to be a driver of dissatisfaction in the investigated Chinese cities. These patterns are investigated in the following in more detail at the level of each single city.

**Table 2. The hotel attributes that perform poorly in tourists’ evaluation**

Shanghai (n = 3298)	Sanya (n = 347)	Hangzhou (n = 451)	Guangzhou (n = 1172)	Beijing (n = 3752)
Service (12.5 %)	Service (13.5 %)	Service (9.0 %)	Service (9.3 %)	Service (9.9 %)
Location (6.6 %)	Resort (4.3 %)	Location (5.3 %)	Location (4.6 %)	Room (6.5 %)
Room (5.4 %)	Staff (2.5 %)	Room (5.0 %)	Room (4 %)	Location (6.0 %)
Staff (4.0 %)	Food (2.0 %)	Staff (3.7 %)	Dirty (2.7 %)	Staff (3.6 %)
Dirty (1.5 %)	Room (2.0 %)	Air (1.7 %)	Staff (2.6 %)	Dirty (2.5 %)

##### 4.1 Shanghai

Specifically, the words ‘better’, ‘bund’ and ‘view’ are keywords that only appear for Shanghai. The word ‘better’ is positively correlated with the word ‘much’ ( $r = 0.2$ ), ‘expected’ ( $r = 0.17$ ), implying hotels in Shanghai generally meet the expectation of tourists. This is in line with the fact that hotels in Shanghai achieve the highest rating from tourists among five cities. In addition, the popularity of the word ‘bund’ implies the bund as an important landmark to attract tourists. As another evidence, the word ‘view’ is positively correlated with the word ‘river’ ( $r = 0.16$ ) and ‘bund’ ( $r = 0.09$ ). Furthermore, we examined the words which have the highest correlation with the positive sentiment word of ‘great’ and ‘good’: The word ‘great’ is positively correlated with the word ‘location’ ( $r = 0.28$ ) while ‘good’ is positively correlated with the word ‘location’ ( $r = 0.20$ ), ‘value’ ( $r = 0.18$ ). Overall, this suggests that accommodation industry in Shanghai should be more focused on exploring the view related to river and bund.

For the 3298 reviews in which Shanghai hotels are rated as poor (rating = 2) or terrible (rating = 1), the words ‘service’, ‘location’, ‘room’, ‘staff’, and ‘dirty’ appear 413, 218, 180, 132 and 51 times respectively. This indicates that poor service is the major issue causing tourists’ dissatisfaction. Bad location of hotels also

deteriorate tourists' rating of their lodging experience.

#### 4.2 Sanya

A number of unique keywords are reported for Sanya, including 'bay', 'beach', 'family', 'food', 'holiday', 'paradise', 'relaxing' and 'resort'. This suggests that Sanya, as a tourism destination, is in particular attractive for family and holiday travelers who seek for delicious food and relaxing time at beach. We also notice that 'business', 'clean', 'comfortable', 'money' and 'price' are key words appearing at other four cities, but not for Sanya. Thus, business-traveler-oriented accessory appears to be less important for the Sanya's accommodation sector.

It is worth noting that 'price' and 'money' are not included in the top keyword list only for Sanya. Also the word 'value' (proportion = 0.008) is much less popular for hotel reviews in Sanya when compared to other cities. This in general suggests that tourists in Sanya are less sensitive to room costs, indicating the possibility to raise the room rates without affecting tourists' satisfaction. The word 'great' is positively correlated with the word 'location' ( $r = 0.15$ ), 'food' ( $r = 0.12$ ), 'rooms' ( $r = 0.11$ ), 'walk' ( $r = 0.11$ ) and 'sandy' ( $r = 0.10$ ) while 'good' is positively correlated with the word 'value' ( $r = 0.20$ ), 'location' ( $r = 0.13$ ), 'money' ( $r = 0.12$ ). This indicates that the accommodation sector in Sanya should focus on exploiting good locations, e.g. beach, and complement its hotel services with nice food. A strong correlation between 'value' and 'good', implies the possibility for generating more revenue.

347 reviews rated their lodging experience in Sanya as terrible or poor. In these reviews, 'service', 'resort', 'staff', 'food' and 'room' appear 47, 15, 9, 7 and 7 times respectively. Astonishingly, the word 'room' is not found as often within tourist complaints as in other cities. Also the word 'dirty' is not in the top list as well, differing from other cities. Holiday tourists travelling to Sanya might be less sensitive to cleanliness of the hotels, since they may spend most of their time on beach. In addition, the feature of Sanya as a city located at an island makes it unaffected from air pollution in mainland, which may probably make cleanliness a less important issue.

#### 4.3 Hangzhou

Hangzhou has three unique words of 'west', 'lake' and 'one'. This indicates that the west lake, as a famous tourist attraction of Hangzhou, attracts substantial interest from tourists. 'Beautiful' is a key word that appears in the top list of keywords for Hangzhou, Shanghai and Sanya. The word 'great' is positively correlated with the word 'location' ( $r = 0.25$ ), 'service' ( $r = 0.12$ ) while 'good' is positively correlated with the word 'location' ( $r = 0.22$ ), 'value' ( $r = 0.22$ ).

451 reviews labeled their lodging experience in Hangzhou as 'poor' or 'terrible'. They name 'service', 'location', 'room', 'staff' and 'air' which appear 41, 24, 23, 17 and 8 times in these reviews respectively. The word 'air' is unique for Hangzhou. As the air pollution becomes more serious and starts to affect Hangzhou, it seems that the demand for improved air quality starts to influence the assessment of the accommodation sector in Hangzhou.

#### 4.4 Guangzhou

Guangzhou has the lowest average hotel rating among the five cities. The list of top key-words includes the words 'bad', 'fair' and 'new', implying a large proportion of unhappy hotel guests. Also, positive sentiment words like 'fantastic' and 'wonderful', that are featured in the other four cities, are not included in the top keyword list of Guangzhou, implying a lack of 'more than expected' lodging experience for tourists. The word 'great' is positively correlated with the word 'location' ( $r = 0.20$ ), 'service' ( $r = 0.09$ ) while 'good' is positively correlated with the word 'location' ( $r = 0.22$ ), 'value' ( $r = 0.15$ ) and 'notso' ( $r = 0.09$ ) and 'price' ( $r = 0.09$ ). This suggests that location is also an important hotel attribute considered by tourists to Guangzhou, and the tourists are in general satisfied with room rate.

There are 1172 reviews giving a ‘poor’ ‘terrible’ rating on their lodging experience in Guangzhou. In the reviews, ‘service’, ‘location’, ‘room’, ‘dirty’ and ‘staff’ appear 109, 54, 47, 32 and 31 times in these reviews respectively. The word ‘dirty’ has the highest proportion among five cities.

#### 4.5 Beijing

The list of top key-words of Beijing does not include any keyword unique to Beijing. This indicates that tourists traveling to Beijing are less affected by external environmental features, such as a lake or a beach. Also the word ‘beautiful’ is not attached to the top keyword list of Beijing, which hints towards smaller enthusiasm of Beijing (business) travelers. ‘Airport’ is listed only for Beijing and Guangzhou, indicating the distance or the convenience to travel to airport is important for tourists. Therefore, hotels in Beijing and Guangzhou that have a reasonable transportation distance to airport should highlight this advantage.

For 3752 reviews in which hotels were rated as 1 or 2, ‘service’, ‘room’, ‘location’, ‘staff’ and ‘dirty’ appear 373, 246, 227, 138 and 97 times respectively. The sequence of features leading to dissatisfaction reveals another pattern unique to Beijing travelers: Room is mentioned more often here than location, indicating that Beijing travelers care more about room features than about the hotel location. Hotel managers should adjust their offerings accordingly in order to satisfy this specific need for travelers to Beijing.

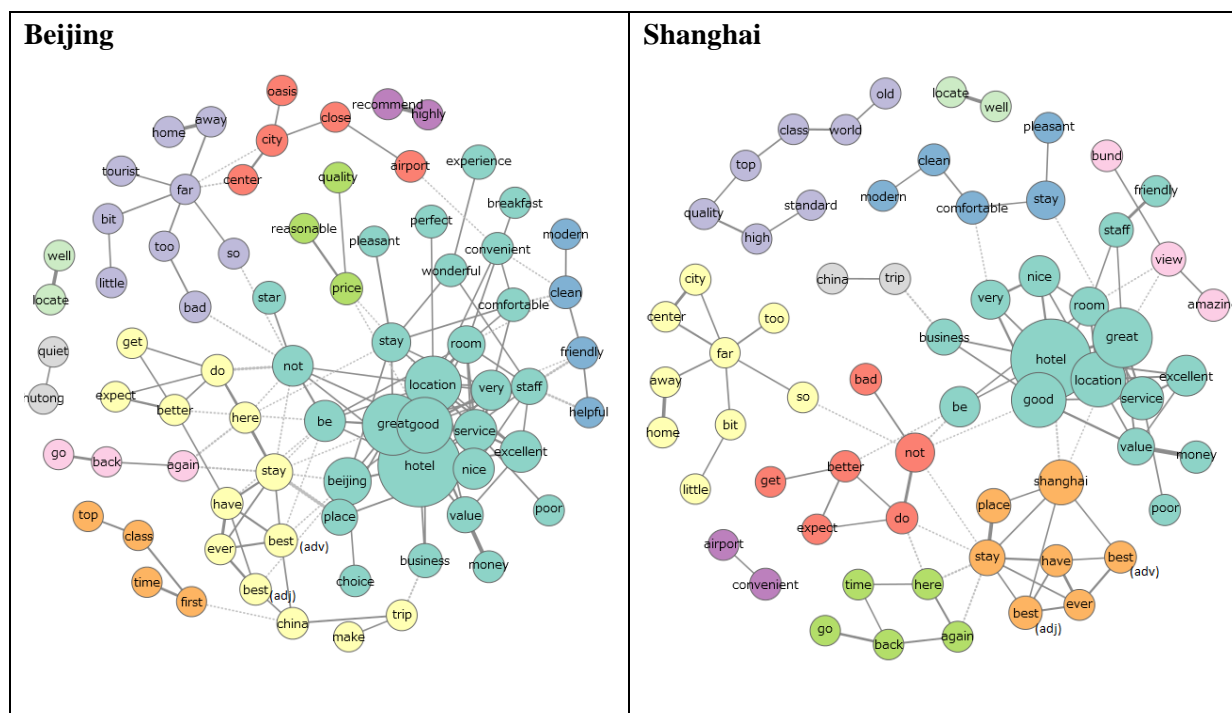
#### 4.6 In-Depth Comparison of Beijing and Shanghai

The analyses in the previous sections reveal large differences in the frequency of top keywords used by travelers between the five cities. They also indicate that keywords are applied in different contexts. For example, it has been shown that “view” is related to “bund” in Shanghai, whereas “airport” was mentioned as a special positive feature in Beijing. This demonstrates the need to further investigate the interrelations between specific words in order to get a richer picture on the underlying patterns of hotel assessments. For this reason, we compare the online reviews for the two cities of Shanghai and Beijing.

We executed a co-occurrence analysis of title words with the use of another program (KH-Coder). The analysis includes nouns, adjectives, adverbs as well as verbs in order to mirror complete sentence structures. Only words with a large term frequency were included in the analysis, whereby for display purposes only a subset of 75 (Beijing) and 64 (Shanghai) words is presented in Figure 1. This figure displays the selected words as nodes and their interrelations as lines. The size of the nodes and the strength of the lines indicate the relative frequency of (co-)occurrence. Words are located next to each other by multidimensional scaling and color-clustered by means of betweenness communities.

Figure 1 shows that both networks of word associations possess similar structures. The density of the associative network and the minimum Jaccard coefficients are around the same with 0.054 (0.032) for Beijing and 0.05 (0.043) for Shanghai. A large cluster focuses on core features of hotel offerings, including e.g. location, staff, service, value and room. However, larger differences between the two cities are to be observed at the other areas of the figure: Repurchase intention (“go back again”) is connected with an overall assessment of the accommodation experience for the Beijing travelers (better, expect, get, stay, have) whereas the intention to visit again is related to an over-fulfillment of expectations (better, expect, do) as well as to an unique hotel experience (best, stay, ever, place) in case of Shanghai. Overall, it can be stated that tourists’ evaluations on hotels in Shanghai are much more emotionally laden compared to the functional evaluations of hotels in Beijing, where reviews emphasize location in either a positive (city, center, close) or negative (bit, too, far) way.





## 5. CONCLUSION

By applying text mining techniques to study online hotel reviews with regard to different cities, a number of interesting findings are reported. First, city known as a holiday resort tend to attract different tourists, who have different demand over hotel attributes. For instance, tourists in Sanya are found to less sensitive to cleanliness issue as well as price. Also cleanliness is not among the top reasons for tourist complaints. The availability of entertainment facilities such as beach appears to be particularly important.

Shanghai is without any doubt an important economic center in China. However, the availability of nice ‘view’ related to bund and river appears to be an important hotel attribute that travelers emphasize. Astonishingly, airport is not among the top keywords, which may be influenced by the convenient public transportation to the airport in Shanghai. In contrast, hotels in Beijing and Guangzhou close to the airport have great advantages over others. Compared to Beijing, Hangzhou tends to have a relatively good air quality. However, the deteriorated air pollution seems to make air quality a concern for tourists. There are rich potentials for the development for high-quality hotels in Guangzhou. The words ‘bad’, ‘fair’ appear in those tourist reviews frequently, suggesting a need for hotels to provide ‘more than expected’ lodging experience.

When comparing the list of words that are related to tourists' complaints, we found that poor service tends to be the most important attribute leading to negative reviews. In addition, bad location of hotel is another important hotel attribute. This indicates that the selection of an adequate hotel location is important to reduce future guests' complaints. In addition, given that Chinese cities are crowded with buildings, it may be important to reduce the difficulty for guests to get access to the hotel. For those hotels located in a busy center of cities, reducing the influence of external environment, such as the noise, is important as well. Furthermore, the words 'value' and 'price' did not appear among the top words in complaints across five cities. This indicates that room rate is not a key issue that leads to tourists' complaints.

The study has a number of limitations. First, the study investigated the frequency of words in the title of reviews. An investigation on the whole review may offer more complete information. Second, the study only explored the reviews written in English regarding hotels in China, the results may not be generalized to other

tourist group or other Chinese cities. Third, the study only utilized a few sentiment words, like ‘great’ and ‘good’. An inclusion of a more complete list of sentiment words may bring more precise results in the research field.

## ACKNOWLEDGEMENT

This research was partly supported by grants from the National Natural Science Foundation of China (No. 71301089; No. 71362027).

## REFERENCES

- [1] TripAdvisor, “Fact Sheet,” 2015. [Online]. Available: [http://www.tripadvisor.com/PressCenter-c4-Fact\\_Sheet.html](http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html). [Accessed: 24-Nov-2015].
- [2] T. Hennig-Thurau, K. P. Gwinner, G. Walsh, and D. D. Gremler, “Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?,” *J. Interact. Mark.*, vol. 18, no. 1, pp. 38–52, 2004.
- [3] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, “Twitter Power: Tweets as Electronic Word of Mouth,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2169–2188, 2009.
- [4] China National Tourism Administration, “China Inbound Tourism in 2014,” 2015. [Online]. Available: <http://www.travelchinaguide.com/tourism/2014statistics/inbound.htm>. [Accessed: 05-Nov-2015].
- [5] C. Vázquez, “Complaints online: The case of TripAdvisor,” *J. Pragmat.*, vol. 43, no. 6, pp. 1707–1717, 2011.
- [6] Y. Zhang and C. Vázquez, “Hotels’ responses to online reviews: Managing consumer dissatisfaction,” *Discourse, Context Media*, vol. 6, pp. 54–64, 2014.
- [7] J. Bulchand-Gidumal, S. Melián-González, and B. González Lopez-Valcarcel, “A social media analysis of the contribution of destinations to client satisfaction with hotels,” *Int. J. Hosp. Manag.*, vol. 35, pp. 44–47, 2013.
- [8] W. Lu and S. Stepchenkova, “Ecotourism experiences reported online: Classification of satisfaction attributes,” *Tour. Manag.*, vol. 33, no. 3, pp. 702–712, 2012.
- [9] E. N. Torres, D. Singh, and A. Robertson-Ring, “Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry,” *Int. J. Hosp. Manag.*, vol. 50, pp. 77–83, 2015.
- [10] B. Fang, Q. Ye, D. Kucukusta, and R. Law, “Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics,” *Tour. Manag.*, vol. 52, pp. 498–506, 2016.
- [11] S. Park and J. L. Nicolau, “Asymmetric effects of online consumer reviews,” *Ann. Tour. Res.*, vol. 50, pp. 67–83, 2015.
- [12] Z. Liu and S. Park, “What makes a useful online review? Implication for travel product websites,” *Tour. Manag.*, vol. 47, pp. 140–151, 2015.
- [13] B. Ma, D. Zhang, Z. Yan, and T. Kim, “An Lda and Synonym Lexicon Based Approach To Product Feature Extraction From Online Consumer Product Reviews,” *J. Electron. Commer. Res.*, vol. 14, no. 4, pp. 304–314, 2013.
- [14] Kurt Hornik, P. Mair, J. Rauch, W. Geiger, C. Buchta, and I. Feinerer, “The textcat Package for n-Gram Based Text Categorization in R,” *J. Stat. Softw.*, vol. 52, no. 6, pp. 1–17, 2013.
- [15] C.-P. Hu, J.-M. Hu, S.-L. Deng, and Y. Liu, “A co-word analysis of library and information science in China,” *Scientometrics*, vol. 97, no. 2, pp. 369–382, 2013.
- [16] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science.*, vol. 286, no. 5439, pp. 509–512, 1999.
- [17] M. Steyvers and J. B. Tenenbaum, “The large-scale structure of semantic networks: statistical analyses and a model of semantic growth,” *Cogn. Sci.*, vol. 29, no. 1, pp. 41–78, 2005.

Appendix A. List of top keywords for cities.

	Beijing		Shanghai		Guangzhou		Hangzhou		Sanya	
	keyword	Frequency	keyword	Frequency	keyword	Frequency	keyword	Frequency	keyword	Frequency
1	hotel	12025 (0.086)	hotel	11597 (0.087)	hotel	3150 (0.095)	hotel	1241 (0.084)	hotel	504 (0.054)
2	great	7245 (0.052)	great	6955 (0.052)	good	1623 (0.049)	great	629 (0.043)	great	369 (0.040)
3	location	6208 (0.044)	location	5888 (0.044)	great	1303 (0.039)	good	620 (0.042)	good	284 (0.030)
4	good	6009 (0.043)	good	5821 (0.043)	location	1102 (0.033)	location	592 (0.040)	resort	267 (0.029)
5	Beijing	4113 (0.029)	Shanghai	3581 (0.027)	stay	842 (0.025)	service	353 (0.024)	service	223 (0.024)
6	stay	3337 (0.024)	stay	3040 (0.022)	nice	779 (0.023)	nice	346 (0.023)	nice	205 (0.022)
7	service	2943 (0.021)	service	3025 (0.022)	Guangzhou	702 (0.021)	stay	323 (0.022)	Sanya	180 (0.019)
8	excellent	2790 (0.020)	excellent	2630 (0.019)	service	636 (0.019)	Hangzhou	289 (0.020)	stay	171 (0.018)
9	nice	2683 (0.019)	nice	2629 (0.019)	excellent	564 (0.017)	excellent	274 (0.019)	excellent	114 (0.012)
10	room	1729 (0.012)	room	1992 (0.015)	room	463 (0.014)	lake	213 (0.014)	beautiful	113 (0.012)
11	value	1550 (0.011)	value	1723 (0.013)	value	421 (0.012)	room	209 (0.014)	place	111 (0.012)
12	staff	1532 (0.011)	business	1548 (0.011)	business	366 (0.011)	value	191 (0.013)	location	110 (0.012)
13	place	1488 (0.010)	best	1490 (0.011)	best	351 (0.010)	best	165 (0.011)	beach	107 (0.011)
14	best	1406 (0.010)	place	1120 (0.008)	place	312 (0.009)	staff	153 (0.010)	best	100 (0.011)
15	experience	1135 (0.008)	staff	1054 (0.008)	staff	253 (0.007)	west	152 (0.010)	staff	88 (0.009)
16	business	1122 (0.008)	experience	924 (0.007)	fair	216 (0.006)	place	129 (0.009)	experience	83 (0.009)
17	comfortable	954 (0.006)	comfortable	789 (0.006)	star	213 (0.006)	beautiful	118 (0.008)	room	80 (0.009)
18	perfect	791 (0.005)	perfect	773 (0.005)	money	212 (0.006)	business	112 (0.008)	china	73 (0.008)
19	wonderful	778 (0.005)	view	724 (0.005)	canton	210 (0.006)	experience	108 (0.007)	value	71 (0.008)
20	star	776 (0.005)	star	713 (0.005)	experience	203 (0.006)	comfortable	104 (0.007)	fantastic	70 (0.008)
21	clean	763 (0.005)	fantastic	687 (0.005)	comfortable	202 (0.006)	star	94 (0.006)	paradise	65 (0.007)
22	fantastic	703 (0.005)	price	673 (0.005)	convenient	189 (0.005)	price	79 (0.005)	wonderful	65 (0.007)
23	price	669 (0.004)	money	648 (0.004)	price	181 (0.005)	clean	77 (0.005)	star	63 (0.007)
24	friendly	611 (0.004)	amazing	623 (0.004)	clean	180 (0.005)	money	76 (0.005)	amazing	61 (0.007)
25	convenient	600 (0.004)	beautiful	611 (0.004)	airport	175 (0.005)	china	73 (0.005)	holiday	60 (0.006)
26	money	564 (0.004)	wonderful	576 (0.004)	new	167 (0.005)	wonderful	70 (0.005)	family	59 (0.006)
27	china	554 (0.004)	clean	574 (0.004)	china	164 (0.005)	average	69 (0.005)	relaxing	53 (0.006)
28	city	534 (0.003)	convenient	537 (0.004)	perfect	161 (0.004)	fantastic	57 (0.004)	perfect	51 (0.005)
29	airport	531 (0.003)	better	516 (0.003)	average	158 (0.004)	friendly	56 (0.004)	bay	50 (0.005)
30	average	525 (0.003)	bund	508 (0.003)	bad	143 (0.004)	one	55 (0.004)	food	49 (0.005)