Online Information Review

Application of social media analytics: a case of analyzing online hotel reviews

Wu He, Xin Tian, Ran Tao, Weidong Zhang, Gongjun Yan, Vasudeva Akula,

Application of social media analytics: a case of analyzing online hotel reviews

1. Introduction

Numerous businesses are involved in online activities by using different social media platforms such as Twitter and Facebook. Comments on social media are a source of information increasingly taken into account by the end customer and businesses. Businesses are increasingly considering social media data as a valuable and timely information source of input for making business decisions. Some businesses even use social media to include users into the product development process to improve and refine their products (Rathore, Ilavarasan & Dwivedi, 2016). Furthermore, more and more businesses are developing business analytics to extract deep meaning insights from social media data every day (Coursaris, van Osch, & Balogh, 2016; He, Zha & Li, 2013; Hoeber et al., 2016). The use of business analytics in analyzing social media data can help businesses obtain novel product ideas, suggestions, and solutions from their customers in order to achieve competitive advantages.

There is a prevalent adoption of social media in the hotel industry. As hotels operate in a competitive and dynamic environment, it is important for hotels to utilize online customer review information effectively in order to better understand their customers, improve hotel performance and compete with other hotels (Berezina et al., 2016). Studies show that online reviews have a substantial effect on consumers' purchase decision (Zhang, Cheung, & Lee, 2014; Brown, Broderick & Lee, 2007). Ye et al. (2011) indicate that a large percentage of customers rely on the online user-generated reviews to make online purchase decisions for hotels, higher than any other product category. Furthermore, hotel managers could act upon online customer reviews to change their marketing strategies and improve their services and brands (Ye et al, 2009).

As electronic Word-of-mouth (eWOM) communication contains vast amounts of consumer information and opinions, analyzing eWOM becomes one of the most efficient and powerful methods to understand customers' feeling about certain service, vendors and products (Cheung, Luo, Sia, & Chen, 2009; Tang et al., 2016). The purpose of this paper is to analyze the English written online reviews of some three to five-star hotels in four big cities in China: Beijing, Shanghai, Guangzhou, and Shenzhen. 58 three to five-star hotels were selected through TripAdvisor including 34 domestic hotels and 24 global chain hotels. In China, only three-star or above level hotels are permitted to accept foreign travelers. Currently, there are few studies examining the online reviews from English-speaking hotel customers in China although millions of foreigners visit China and stay at hotels each year (Hsu, 2015). Foreign tourists constitute an important part of customers for both accommodation and tourism industry in China (Liu et al., 2016). Therefore, it is important to understand the preferred hotel attributes, main concerns or demands of foreign tourists when they stay at hotels in China. Our study will contribute to the social media analytics literature by discovering new rich findings and providing actionable insights with implications to hotel managers in China.

Due to the availability of a large amount of user-generated data on social media, there is a growing interest in using automated computational methods such as text mining and sentiment analysis to process large amounts of user-generated data and extract meaningful knowledge and insights. Traditional content analysis methods are no longer able to meet organizations' needs to analyze the large amount of new content on a daily basis. To leverage the huge opportunity offered by user-generated data, enterprises need to develop their data analytics skills to handle large amounts of data. Studies indicate considerable performance differences in those companies that embrace the opportunities around analytics (Zikopoulos et al., 2012). Therefore, this paper seeks to show businesses how they can apply and integrate emerging technologies to analyze unstructured textual data available on social media through a case study.

The main contributions of this paper include: 1) we analyzed the online reviews from English-speaking hotel customers in China to understand their preferred hotel attributes, main concerns or demands. There are currently limited studies in this specific niche area although millions of foreigners visit China and stay at hotels each year. 2) we illustrated an approach that uses natural language preprocessing, text mining and sentiment analysis techniques to analyze online textual content such as online reviews through a case study with visualized figures. The variety of extracted knowledge shown in the case study reveals the value of the proposed approach. 3) The case study about online hotel reviews generates some meaningful finings and insights. This case study used a variety of visualization techniques to show how hotels could exploit, visualize and capitalize upon gathered data of their own and competitors to facilitate decision making about strategic planning and operation. The generated graphs such as sentiment by ratings and category correlation network contribute to the diversity of visualization techniques for conducting online content analysis and provide additional insights with regard to leveraging visualized sentiment analysis to realize benefits from business analytics. Furthermore, the study reveals that the overall review star rating correlates well with the sentiment scores for both the title and the full content of the online customer review. Compared with the title, the full review content has more insightful results with text analytics, since title is very short in many cases. The case study also revealed that both extremely satisfied and extremely dissatisfied hotel customers share a common interest in the five categories: food, location, rooms, service, and staff.

The rest of the paper is organized as follows. Section 2 provides a review of text mining and sentiment analysis. Section 3 discusses a case study that uses an approach that combines natural language preprocessing, text mining and sentiment analysis techniques to analyze online hotel reviews. Section 4 discusses the implications and insights from this case study. Conclusions and future research are given in section 5.

2. Literature Review about Text Mining and Sentiment Analysis

As users continue to post a large amount of textual information on various social media sites, there is a growing interest in using automatic methods such as text mining and sentiment analysis to process large amounts of user-generated data and extract meaningful knowledge and insights.
As an emerging technology, text mining aims to extract meaningful information from a large number of textual documents quickly (Liu, Cao, & He, 2011; He, Zha & Li, 2013). Text mining is focused on finding useful models, trends, patterns, or rules from unstructured textual data (Romero, Ventura & Garcia, 2008; He et al., 2015).

Text mining techniques have often been used to analyze large amounts of textual data to automatically extract knowledge, insights, useful patterns or trends (Zhong, Li, & Wu, 2012). For example, He, Zha and Li (2013) applied text mining to analyze unstructured text content on Facebook and Twitter sites of the three largest pizza chains: Pizza Hut, Domino's Pizza and Papa

John's Pizza. Their results reveal the value of text mining as an effective technique to extract business value from the vast amount of available social media data. Berezina et al. (2015) examined the underpinnings of satisfied and unsatisfied hotel customers by using a text-mining approach to analyze and comparing 2,510 online reviews by satisfied and dissatisfied customers. They found that satisfied customers refer to intangible aspects of their hotel stay, such as staff members, more often than unsatisfied customers. In contrast, dissatisfied customers mention more frequently the tangible aspects of the hotel stay, such as furnishing and finances. Xiang et al. (2015) used a text analytical approach to analyze a large quantity of consumer reviews extracted from Expedia.com in order to deconstruct hotel guest experience and examine its association with satisfaction ratings. Their findings reveal several dimensions of guest experience that carried varying weights and have novel, meaningful semantic compositions. They also found a strong association between guest experience and satisfaction. Dirsehan (2015) analyzed 500 positive and 500 negative comments collected from booking.com and identified controllable and uncontrollable factors. Controllable factors include staff, food, bed, service, etc. Hotel managers can train the staff to be more friendly or helpful, provide better quality food, improve the service, etc. Uncontrollable Factors are mainly location-related such as place, pool and beach, which are determined when the hotel is built and are difficult to control and change.
Tang and Guo (2015) investigated the validities and usefulness of text mining in terms of analyzing text-based electronic word-of-mouth (eWOM) communication. Their results provide initial evidence for the validity and utility of text mining and demonstrate that the linguistic indicators generated by text analysis are predictive of eWOM communicators' attitudes toward a product or service. Li et al. (2015) adopted mining technique to identify emergent hotel features of interest to international travelers. Guo et al. (2017) extracted 19 dimensions of visitor satisfaction from online hotel reviews and found heterogeneity amongst different visitor demographic segments.

More and more managers are interested in deciphering consumer's sentiment expressions from online reviews and social media comments in order to monitor product and evaluate services. For example, He et al. (2016) proposed a social media analysis framework to enhance the business value and market or business intelligence. As a special application of text mining, sentiment analysis has become very popular since it is concerned with the automatic extraction of positive or negative opinions from text (Pang & Lee, 2004). Sentiment analysis is the computational detection and study of opinions, sentiments, emotions, and subjectivities in text (Pang & Lee, 2004; Li & Wu, 2010; Liu, 2010). As texts often contain a mix of positive and negative sentiment, it is often useful to identify the polarity of sentiment in text (positive, negative, or neutral) and even the strength of sentiment expressed (Thelwall, Buckley, & Paltoglou, 2012; Pang & Lee, 2004). Sentiment analysis has been used to determine the attitude of customers and online users on some specific topics such as consumer product (e.g., books, movies, electronics) reviews, hotel service reviews, etc. Bollen et al. (2011) used sentiment analysis to mine a large corpus of Twitter messages to determine the mood of the Twitter population on a given day. They found that the mood of the Twitter population was able to predict the movement of the Dow Jones Industrial Average (DJIA) on the following day with a claimed 87.6% accuracy. Duan, Cao, Yu, & Levy (2013) used the sentiment analysis technique to mine 70,103 online user reviews posted in various online venues from 1999-2011 for 86 hotels in the Washington D.C. Sentiment analysis helped them decompose user reviews into five dimensions to measure hotel service quality and the sentiment analysis results show high level of accuracy in capturing and measuring service quality dimensions compared with existing text mining studies. He, Tian, Chen & Chong (2016) used text mining and sentiment analysis techniques to analyze and compare social media content on the Facebook sites of the three largest drugstore chains in the United States: Walgreens, CVS, and Rite Aid and found similarities and differences in the social media use among the three drugstore chains. Villarroel Ordenes (2016) conducted a text mining study using more than 45,000 consumer reviews and the results demonstrated the differential impacts of activation levels (e.g., tentative language), implicit sentiment expressions (e.g., commissive language), and discourse patterns (e.g., incoherence) on overall consumer sentiment (i.e., star ratings). Their study provides a fine-grained analysis of the implicit and explicit language used by consumers to express sentiment in text. Xiang et al. (2017) applied text analytics to compare three major online review platforms (TripAdvisor, Expedia, and Yelp) in terms of information quality related to online reviews about the entire hotel population in Manhattan, New York City. Their study found huge discrepancies in the representation of the hotel industry on these platforms. Particularly, online reviews vary considerably in terms of their linguistic characteristics, semantic features, sentiment, rating, usefulness as well as the relationships between these features. Ullrich & Brunner (2015) found that a positive customer review counteracts a negative consumer review more effectively than a positive brand response. Krawczyk & Xiang (2016) used a text analysis approach to create perceptual maps from the most frequent terms used in a data set collected hotel brands and revealed the value of online consumer reviews in representing the level of differentiation between hotel brands for understanding the market structure of the hotel industry. In addition, Yu, Li and Jai (2017) examined the top 10 hotels who participate in the TripAdvisor Green Leaders Program (eco-friendly) using content analysis from TripAdvisor and compared the customers' reviews then found that the basic practices and advanced green practices have the positive influences on customer satisfaction. They also found the terms, such as "guest training", "energy", "water", "purchasing" and "education", are significantly correlated with the overall customer satisfaction.

The above studies not only indicate the need to conduct more research in this area but also provide insights and guidance for our study. However, those studies mainly examined the online reviews for hotels in the Western countries. According to Li, Wang and Yu (2015)'s study in United States, the current website exploitation is still limited with focus on information dissemination. Little research is done to examine online reviews for English-speaking hotel customers in China. As heterogeneity exists amongst different visitor demographic segments (Guo et al., 2017) and millions of foreigners visit China each year, it would be worthwhile to explore how foreign visitors perceive hotels in China and see if we can find any new insights or knowledge from this group of people. Thus, we decided to use the text mining and sentiment analysis techniques to analyze our own data through a case study.

3. A Case Study

We conducted a case study to analyze a unique data set of user-generated online hotel reviews from Tripadvisor.com. We wrote a web crawler program to automatically extract the online user reviews from the Tripadvisor.com in December 2015. The data set contains 11,043 English online user reviews for 58 three to five-star hotels in the four largest cities in China: Beijing, Shanghai, Guangzhou, and Shenzhen. We choose three to five star hotels because only three-star or above hotels are allowed to accommodate foreign travelers in China. For each review, we collected both structured information (overall rating score and specific rating scores) and unstructured information (titles and textual review content). We used natural language preprocessing, text mining and sentiment analysis techniques to analyze these reviews.

Approach

Figure 1 lists an approach we used for integrating natural language preprocessing, text mining and sentiment analysis techniques to analyze online textual content. This approach is adapted from a text mining framework proposed by He, Zha and Li (2013).

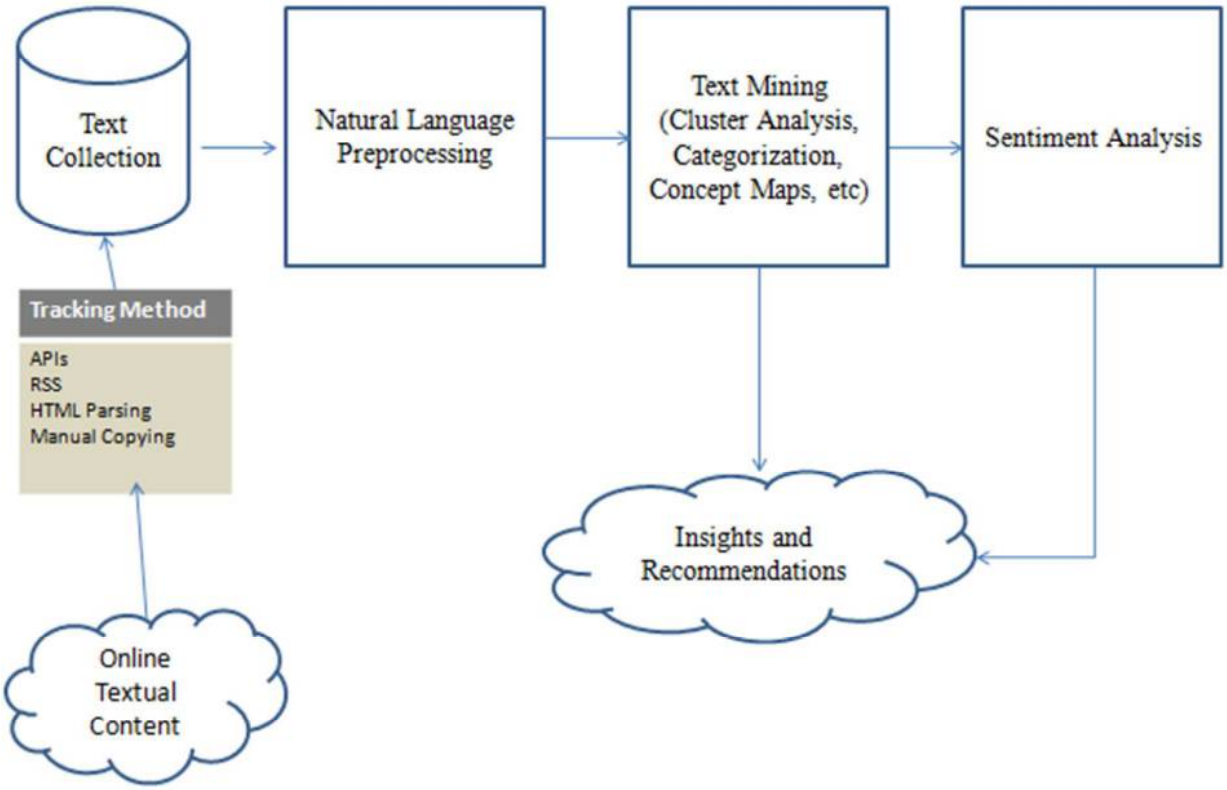


Figure 1. An Approach of Using Natural Language Preprocessing, Text Mining and Sentiment Analysis Techniques for Understanding Online Textual Content

Below is a description of the steps.
1. We wrote a program to extract all consumer reviews related to these hotels in China from tripadvisor.com and then saved the reviews into an excel spreadsheet. The program used the web scraping and HTML parsing techniques to retrieve data that exists on the tripadvisor.com website, and converted it into a format that is usable for analysis.
2. Next, we conducted natural language preprocessing to clean the text comments such as removing stop words. Data pre-processing step is needed to ensure that raw data are transformed

into a usable format, mainly by data cleaning. We used the well-known R package for text mining to analyze the texts for word frequency analysis and term co-occurrence analysis.
3. Next, we manually examined a small part of the data set and developed a list of possible categories. Traditional content analysis method can be used to identify the emerging categories from the small data sample. NVivo qualitative software was also used to facilitate the content analysis and coding. This process may be iterative and can be expanded to include more data if needed until no more new categories could be identified.
4. Afterwards, we wrote a program to go through the entire data set and put relevant words (related to the identified categories) into multiple categories. For example, the food category can include relevant words such as "foods", "breakfast", "lunch", "dinner", "snacks", "cuisine", "cuisines", "brunch", "starters", "starter", "soup", "soups", "main course", "desset", and "desserts".
5. For each category, we applied sentiment analysis techniques to further put comments into groups: positive, neutral and negative.
6. We examined the results carefully to identify emerging trends, hot topics and theme of hotels for decision making and insights. We conducted various query searches using the search feature of NVivo. The query searches are mainly used to find interesting patterns, connections, and unusual information from the qualitative textual content.

Results of the Case Study

We obtained totally 11,043 English comments captured from tripadvisor.com. Before we perform text mining techniques, we first clean all the space and special symbols in the dataset, such as “&”, “@”, “!” and so on. The most important step is processing the dataset before analyzing it. We covert all the texts to lower cases and remove common stop words in each sentence, such as “a”, “an”, “the”, and so on. Because those common stop words do not have any analytic value, we use function “stopwords (English)” in R packages to eliminate those words. After the processing, we create a document term matrix and then start to explore the data.

The famous open source statistical software, R (https://www.r-project.org/)’s text mining package is used for analyzing word frequency (Maceli, 2016). The word frequency analysis identified terms that appear frequently in these comments. Table 1 shows some of the top words that are frequently used among all comments in our dataset.

Table 1. Top 12 words with high frequencies

1. hotel	2. room	3. good	4. staff	5. great	6. stay
22182	16674	8333	7223	6049	5455
7. service	8. breakfast	9. nice	10. stayed	11. location	12. food
5392	5224	4273	4231	4096	3615

After we got those high frequency terms, we did word co-occurrence search and analysis to identify the terms with strong correlation with those high frequency terms in R. The correlation is defined between 0 and 1. For instance, if the correlation is 1, the search results will return only those words that always co-occur with the search term. The lower bound we set is 0.8 for the strength of correlation between the searched high frequency terms and result terms. That means that the program will return terms that have a co-occurrence of at least 80% and above for the searched high frequency terms. We have searched five high frequency terms and the results are listed below in Table 2. For example, term "hotel" always occurs with location, shopping and the degree of convenience. That means that when customers post their comments online they tend to mention the location and convenience of this hotel. For the term service, pleasure and luxurious are often mentioned (at least 89% of the time). Term "room" often co-occurs with the hotel's toiletries including bathtub, bathroom and toilet.

Table 2. Co-occurrence Terms

Search Term: hotel									
located	street	adjacent	probably	streets	rest	details	area	dont	outlets
0.90	0.88	0.87	0.87	0.87	0.86	0.85	0.82	0.82	0.82
shops	step	careful	installed	forced					
0.82	0.81	0.81	0.81	0.80					

Search Term: stay			
sausages	hesitation	ready	easier
0.86	0.84	0.84	0.81

Search Term: staff			
earlier	ambience	standalone	stopped
0.86	0.82	0.82	0.80

Search Term: service							
pleasure	luxurious	vast	account	Jacuzzi	excellent	concierge	day
0.92	0.89	0.88	0.87	0.84	0.83	0.83	0.81

Search Term: room									
toilet	sets	sleek	ultra	action	aware	since	river	tub	windows
0.94	0.91	0.87	0.87	0.87	0.86	0.86	0.86	0.86	0.85
actual	bathtub	disappointing	think	better	change	elevators	heart	selection	bathroom
0.83	0.83	0.82	0.82	0.81	0.81	0.81	0.81	0.81	0.80

Out of these comments, we conducted a sentiment analysis by using Google Prediction API (https://cloud.google.com/prediction/) for each comment in dataset. Google Prediction API provides pattern matching and machine learning capabilities. It can not only help solve classification type problem but also use natural processing languages (NPL) to generate the sentiment score for each comment. The result of our data set shows that about 78% comments are positive, 5% comments are negative and 17% comments are neutral, as shown in Figure 2. This gives us a basic idea of the hotel industry.

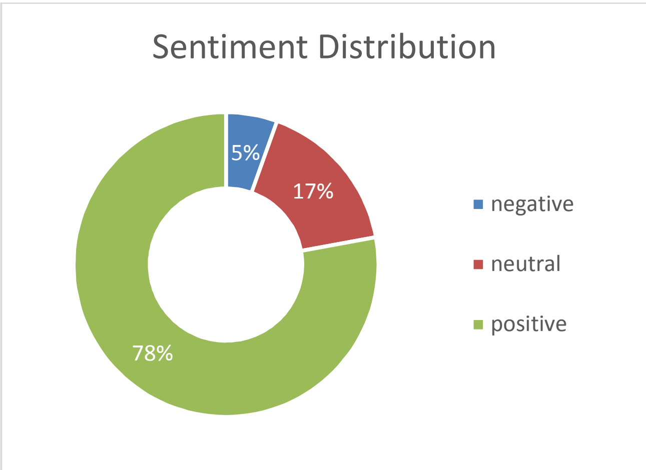


Figure 2. Overview of Sentiment Distribution

We also obtained the categories. As shown in Figure 3, there are totally 30 categories.

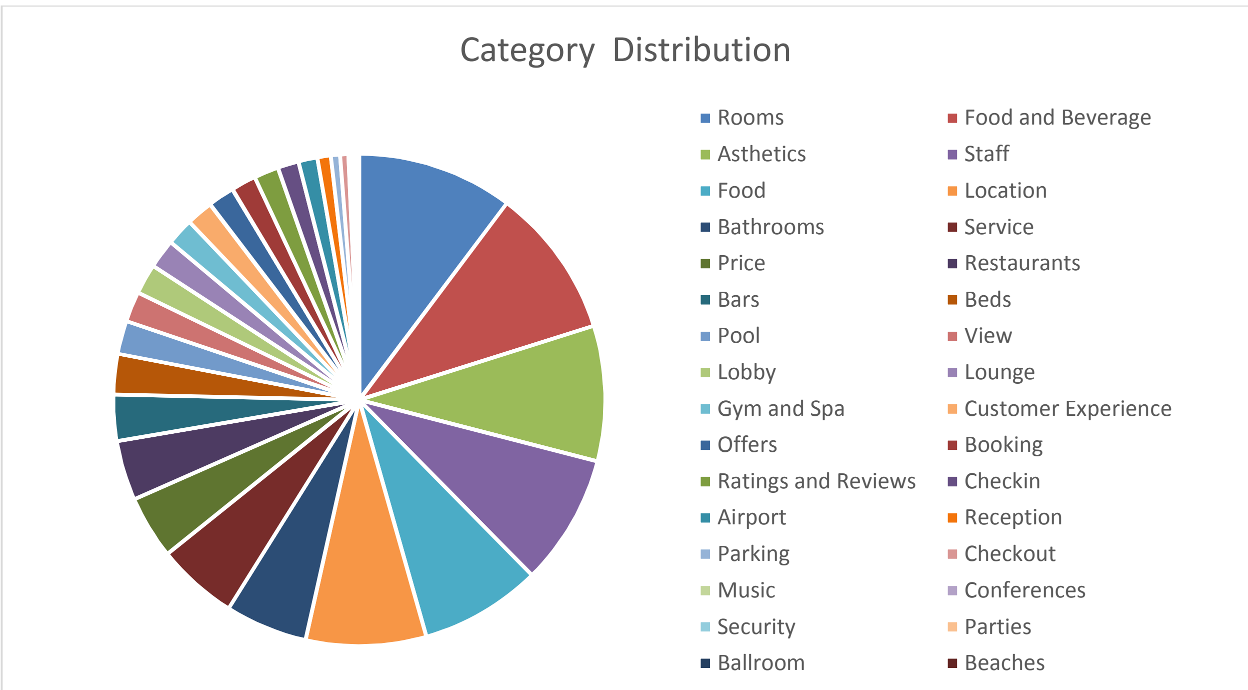


Figure 3. Category volume distribution

For each category, we compared the percentage values of positive comments. The comparison showed us a clear picture about customers' interests, as shown in Figure 4. The total volume in each category and its overall sentiments are presented in Figure 4. For example, rooms, food and beverage, aesthetics, and staff are the top four things customers care most because they have most comments volume. For every category, we can see the positive comments vs. grant total comments. For example, location has received about 43% positive comments out of all comments. The highest positive comments come from the view as well as the food and beverage category which has about 49% positive comments.

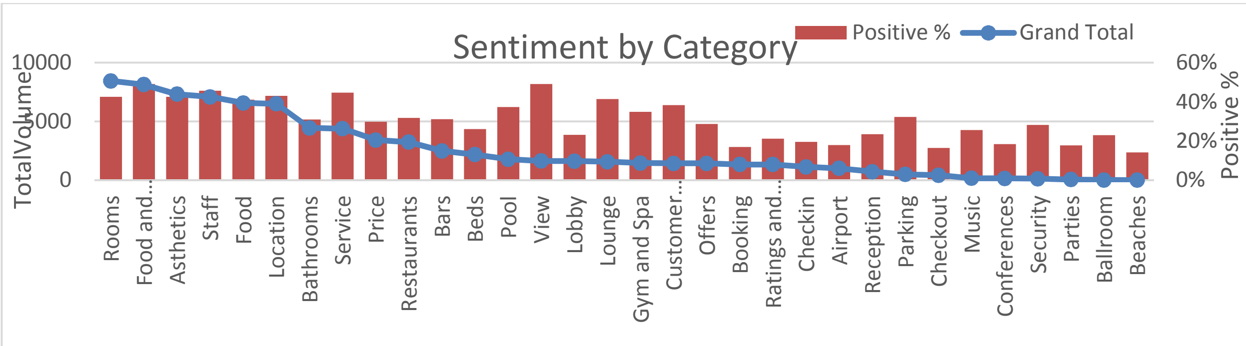


Figure 4. Sentiment of each category.

We also presented the correlation among the category. For each category, we computed the value of correlation i.e. the frequency of a comment which exists in both categories. Figure 5 shows the category correlation network we developed. We used R program to transform data into an adjacency matrix, and then we built a term-term adjacency matrix, where the rows and columns represents terms, and every entry is the number of co-occurrences of two terms. Next, we used graph.adjacency() function with R using package igraph to build a graph. In the figure 5, color of the bubble represents the sentiment for a category, green being positive, grey neutral and red negative. Thickness of the edge connecting two categories represents the frequency with which two categories occur together. Size of the bubble represents the volume for a category. From this figure, we learned that rooms, food and beverage, and aesthetics are the categories with top three highest volumes of comments. There were some complaints about parking, bed, check-in and check-out process, etc. We also found the categories that were closely related with strong relations. For instance, the "staff" category had strong correlation with the "food and beverage" category. Most of the comments were positive comments as the colors of most categories are green or alike.

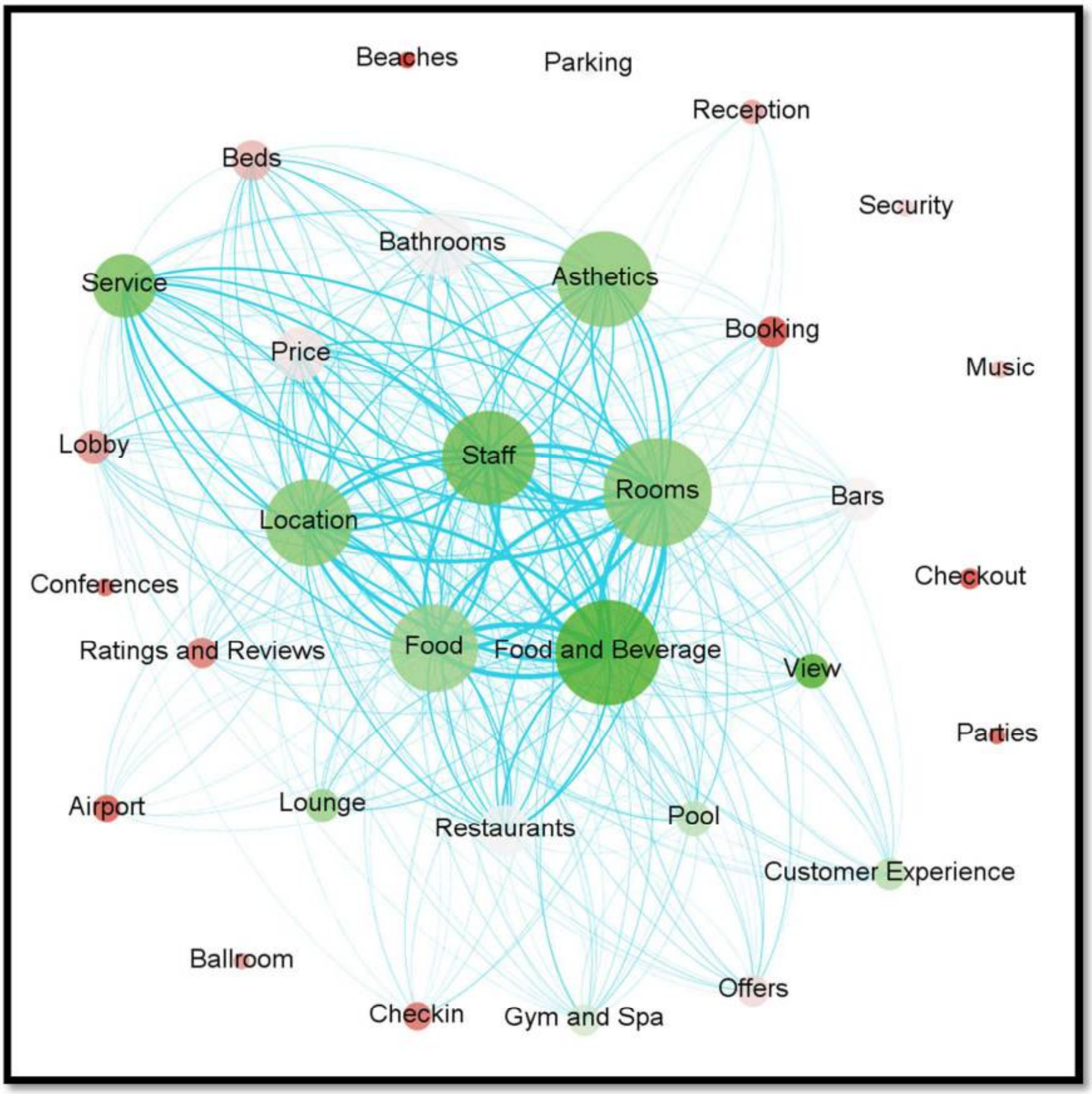


Figure 5. Category correlation network

Overall, the sentiment and category analysis give us a fundamental idea of hotel industries. We can run the same analysis for each hotel to transform comments into opinion knowledge. A full picture of knowledge of a hotel can be created by concatenating opinion knowledge with other factual knowledge such as the location, price, vacancy, and rooms, etc.

Furthermore, we examined the relationship between the overall review rating of each review and sentiment scores of the review including its title and full content. Figure 6 and 7 show the review content and title's sentiment by rating. We found that the overall review star rating correlates well with the sentiment scores for both the title and the full content of the online customer review. Compared with the title, the full review content has more insightful results with text analytics, since title is very short in many cases.

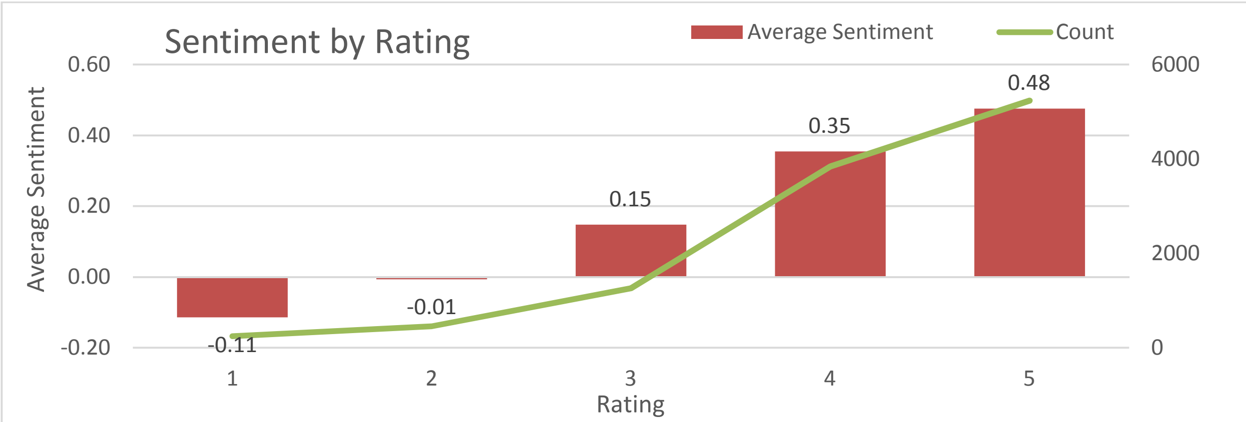


Figure 6. Review content's sentiment by rating

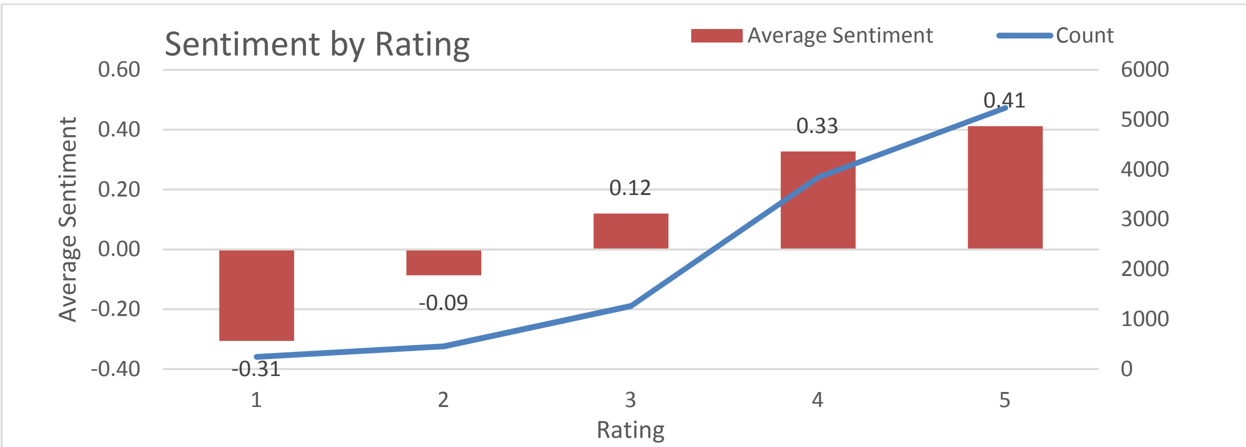
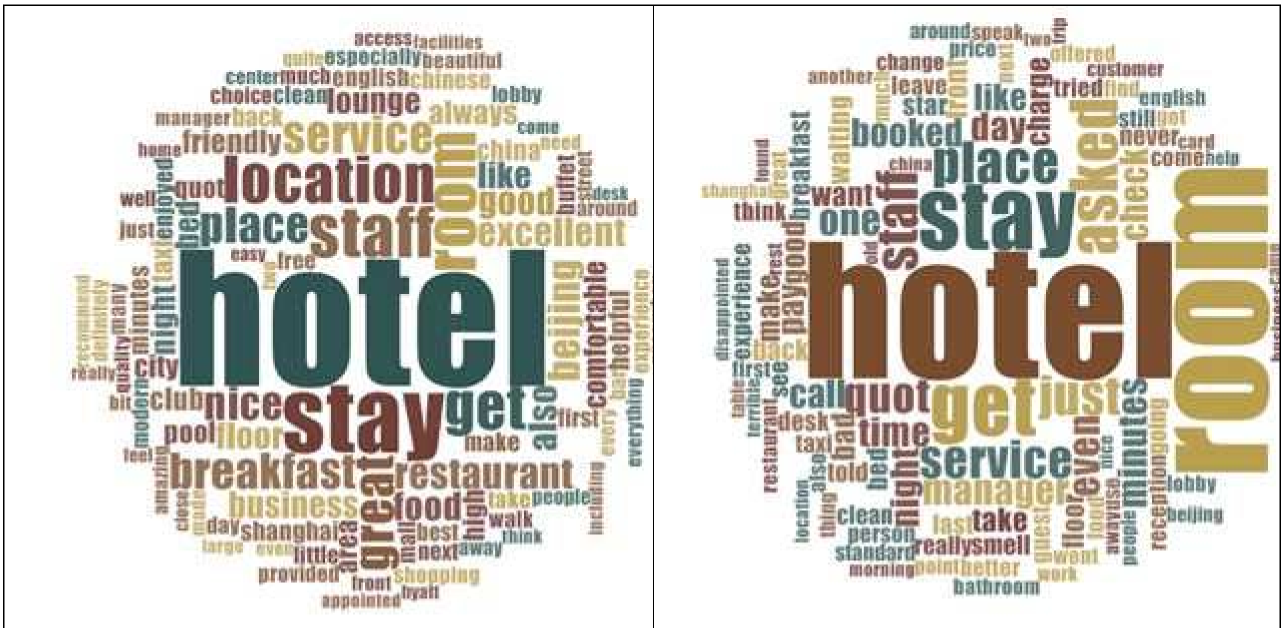


Figure 7. Review title's sentiment by rating

It is important for hotels to make customers satisfied so that they will keep coming back to the hotels later. Positive contents in online reviews can help improve hotel revenues (Anderson, 2012). In addition, Fang et al. (2016) found that the reviews expressing extreme sentiment are more likely to be voted as helpful. Park and Nicolau (2015) found that extreme ratings (positive or negative) are perceived to be more useful and enjoyable in comparison to moderate ratings. Thus, we used two data and text analytics tools including NVivo and R to analyze and compare the full content of the reviews from extremely dissatisfied customers (giving 1-star rating) and extremely satisfied customers (giving 5-star rating). The generated word clouds for extremely satisfied customers and extremely dissatisfied customers can be found in Figure 8. The word clouds reveal some common categories that are frequently mentioned in both positive and negative reviews, including food, location, rooms, service, and staff (see Table 3).



extremely satisfied customers | extremely dissatisfied customers

Figure 8. Word clouds of extremely satisfied customers and extremely dissatisfied customers

Table 3: the most frequently mentioned categories in positive and negative reviews

	Categories
Extremely Satisfied Customers	food, location, rooms, service, staff, lounge,
Extremely Dissatisfied Customers	Rooms, food, check-in, service, staff

4. Discussion and Implications

Recent years have seen an explosion in social media data. Many Internet users are sharing their experience and opinions about products and services they received on social media such as online forums. Businesses need to adapt new analytics methods and tools to develop better business intelligence and insights. This paper presents a feasible approach and a case study to show businesses how to analyze online customer reviews to discover in-depth insights and achieve deep understanding of the customer reviews on social media. Gaining insights from online customer reviews could provide valuable managerial information to hotel managers and help them identify the strengths and weaknesses of their hotels. For example, hotel managers could use our approach to monitor categories that have low sentiment scores and take actions to respond to the negative reviews immediately to reduce the possible consequence. They can also explore the hotel attributes that customers discussed in their reviews in addition to their lodging experience. Liu, Kim and Pennington-Gray (2015) suggest offering prompt responses with open, transparent, and customized information to concerned hotel customers. Hotel managers can also compare the review results over time to see if their actions have any actual impact on customer satisfaction and experience. The word co-occurrence analysis can help hotel managers identify top concerns or mentions so that they know which aspect of their hotel needs more attention.

From a managerial perspective, Berezina et al. (2016) suggest examining and monitoring the categories that have emerged from the online customer reviews to not only understand the voice of every guest, but also to see a larger picture that all of these voices would form collectively. At the final stage of text mining, an evaluation of the results is needed to see if knowledge was discovered and to determine its importance with the support of visualization techniques and representation tools such as graphs and association rules. Figure 3 uncovers 30 dimensions or categories for hotels to manage their interactions with guests. Figure 4 clearly shows that some categories receive low sentiment scores and need special attention for hotel managers to take care of. In addition, Figure 5 shows the category correlation networks and this figure can help hotel managers to see category that hotel guests are satisfied or dissatisfied with and identify inherent relations among different categories. Sometimes a problem in a category cannot be effectively solved until the hotel manager look at all related categories and address other related problems. Figure 5 can help the hotel manager see a bigger picture for potential opportunities or a series of related problems they need to resolve. As some categories, such as location may be considered uncontrollable by the hotel managers, hotel managers can choose to focus their efforts on those controllable categories to make an actionable impact. For example, they may train the staff to make them friendly or helpful, improve the food quality and variety offered by the hotel, improve the check-in service, etc. The graphs we generated from this case study such as sentiment by ratings and category correlation network show not only the ability for social media analytics to produce business insights, but also their potential for understanding how different properties of hotels are related and perceived in the minds of consumers (Krawczyk & Xiang, 2016). The case study found that the overall review star rating correlates well with the sentiment scores for both the title and the full content of the online customer review. The results support the findings of Kim, Lim, & Brymer (2015) that overall ratings are the most salient predictor of hotel performance.

This case study also provides empirical findings to help hotel managers understand the thoughts or opinions from the extremely satisfied and extremely dissatisfied hotel customers. We found that both extremely satisfied and extremely dissatisfied hotel customers share a common interest in the five categories: food, location, rooms, service, and staff. This finding offers clear managerial implications for managers to consider these five categories and suggest that those categories that make customers happy may also make them unhappy if they are not provided or there are issues with these categories (Berezina et al., 2016). Efforts should be made to enhance customers' positive perception of the five categories. For hotels with strong performance in any of the five categories, managers are recommended to highlight them as their competitive advantage for marketing and advertising.

Studies indicate that organizations focusing on analytics significantly outperform their peers on the key business metrics of growth, earnings and performance (Zikopoulos et al., 2012). Enterprises need to develop capability in collecting, storing and analyzing social media data for harvesting information and actionable knowledge for decision making and forecasting. Many enterprises are struggling in analyzing the social media data they obtained. As user-generated content (UGC) become increasingly ubiquitous, companies need to develop advanced business analytics capacity to differentiate themselves from their competitors.

As social media analytics is a young and emerging interdisciplinary research topic, the development of effective and efficient analytics techniques for social media analytics becomes essential. The results of the case study offer clear managerial implications for hotel managers using natural language preprocessing, text mining and sentiment analysis techniques. This paper contributes to this emerging area by not only presenting a feasible approach but also providing a visual analytics example for enterprises to follow. The graphs we generated such as sentiment by ratings and category correlation network contribute to the visualization techniques people can use in conducting online content analysis and provide additional insights with regard to leveraging visualized sentiment analysis to realize benefits from business analytics. Mastering these techniques will help organizations build stronger business analytics expertise to improve their existing business analytics operational practice and processes. Finally, the approach used in this paper can be used to a wide range of application domains to enhance organizations' information acquisition, transformation and exploitation capability.

..

5. Conclusion

More and more people are publicly expressing their personal thoughts and feelings using social media platforms such as online forums and Twitter on a scale we have never seen before. It is critical that businesses are provided with a feasible framework or approaches that not only help them make sense of the large amount of accumulated text, but also help them do that in an effective manner.

The goal of this study was to explore the usefulness of online reviews to understand how hotels are perceived by consumers. This paper seeks to show businesses how social media analytics can identify insights from online reviews through a case study. As a main contribution, this study presents a tested approach of using natural language preprocessing, text mining and sentiment analysis techniques to analyze online textual content. Early identification of emotions expressed by customers, and the possibility of identifying them with different categories (food, staff, service, etc.) can enable hotel managers to train the staff to be more friendly or helpful, improve the check-in or check-out service, etc. The generated results are particularly useful for rapid service monitoring, troubleshooting and evaluation. For instance, hotel managers can use the results to design customer strategies and evaluate staff performance including determining the strengths and weaknesses of existing services. The value of the proposed approach was demonstrated through a case study using online hotel reviews. The case study found that the overall review star rating correlates well with the sentiment scores for both the title and the full content of the online customer review. Compared with the title, the full review content has more insightful results with text analytics, since title is very short in many cases. The case study also revealed that both extremely satisfied and extremely dissatisfied hotel customers share a common interest in the five categories: food, location, rooms, service, and staff. In summary, the results demonstrate the value of using natural language preprocessing, text mining and sentiment analysis to categorize textual content, discover new knowledge and gain insights from a large amount of textual data. Businesses can follow our approach to guide their efforts to track, collect, and analyze various user-generated textual contents on the Internet and to enhance hotel guest experiences and reputation management (Wan & Law, 2017). Our approach also provides a good foundation for developing useful analytics tools for the hotel industry in the era of big data. For future research, we will compare how online hotel customer reviews written in different languages such as Chinese and English and see to what extent they differ and what factors lead to such a difference. We can also compare the responses of hotel management in China and western countries to online reviews and identify their respective online reputation management styles. We will also compare the online hotel reviews in different cities in China to identify which cities provide better lodging experience to foreign customers or tourists.

REFERENCES

Anderson, C.K. (2012). The impact of social media on lodging performance. *Cornell Hospitality Reports*, 12(15), 6-11.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1-24.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualising the online social network. *Journal of Interactive Marketing*, 21(3), 2-20.

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce*, 13(4), 9-38.

Coursaris, C. K., van Osch, W., & Balogh, B. A. (2016). Informing brand messaging strategies via social media analytics. *Online Information Review*, 40(1), 6-24.

Dirsehan, T. (2015). Gaining Competitive Advantage in Tourism Marketing: A Text Mining Approach to Hotel Visitors' Comments in Durrës. In *Proceedings of 4th UBT Annual International Conference on Business, Technology and Innovation* (p. 41).

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. Proceedings of the 46th Hawaii International Conference on System Sciences, 2013.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics. *Tourism Management*, 52, 498-506.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467-483.

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, *115*(9), 1622-1636.

He, W., Tian, X., Chen, Y., & Chong, D. (2016). Actionable Social Media Competitive Analytics for Understanding Customer Experiences. *Journal of Computer Information Systems*, *56*(2), 145-155.

He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, 52(7), 801-812.

He, W., Zha, S.H., & Li, L. (2013). Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry. *International Journal of Information Management*, (33:3), 464–472.

Hoeber, O., Hoeber, L., El Meseery, M., Odoh, K., & Gopi, R. (2016). Visual Twitter Analytics (Vista) Temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Information Review*, 40(1), 25-41.

Hsu, C. (2015). Brand performance of Chinese domestic vs. international hotels: Perceptions of operators as well as domestic and foreign guests. *Research in Hospitality Management*, 5(2), 123-133

Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, *44*, 165-171.

Krawczyk, M., & Xiang, Z. (2016). Perceptual mapping of hotel brands using online reviews: a text analytics approach. *Information Technology & Tourism*, 16(1), 23-43.

Li, N., & Wu, D. D.(2010). Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast. *Decision Support Systems*, (48:2), 354–368.

Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. R. (2015). Identifying emerging hotel preferences using Emerging Pattern Mining technique. *Tourism management*, *46*, 311-321.

Liu, B.(2010). Sentiment analysis and subjectivity. In: *Handbook of Natural Language Processing*, N. Indurkhya and F. J. Damerau (eds.).

Li, X., Wang, Y., & Yu, Y. (2015). Present and future hotel website marketing activities: Change propensity analysis. *International Journal of Hospitality Management*, *47*, 131-139.

Liu, B., Cao, S. G., & He, W. (2011). Distributed Data Mining for E-Business. *Information Technology & Management*, 12(1), 1–13.

Liu, B., Kim, H., & Pennington-Gray, L. (2015). Responding to the bed bug crisis in social media. International Journal of Hospitality Management, 47, 76-84.

Liu, Y., Teichert, T., Hu, F., & Li, H. (2016). How do tourists evaluate Chinese hotels at different cities? Mining online tourist reviewers for new insights.
WHICEB 2016 Proceedings. Paper 67. http://aisel.aisnet.org/whiceb2016/67

Maceli, M. (2016). Introduction to Text Mining with R for Information Professionals. *dim (dtm),* 1(51), 7891.

VILLARROEL ORDENES, F., Ludwig, S., GREWAL, D., DE RUYTER, K., & WETZELS, M. (2016). Analyzing Online Reviews Through the Lens of Speech Act Theory: Implications for Consumer Sentiment Analysis. *Journal of Consumer Research.*

Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Barcelona, Spain, pp 271-278.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67-83.

Rathore, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2016). Social media content and product co-creation: an emerging paradigm. *Journal of Enterprise Information Management*, *29*(1), 7-18.

Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.

Tang, C., & Guo, L. (2013). Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters*, *26*(1), 67-80.

Thelwall, M., Buckley,K., & Paltoglou, G. (2012). Sentiment Strength Detection for the SocialWeb. *Journal of the American Society for Information Science and Technology*, (63:1), 163–173, 2012.

Ullrich, S., & Brunner, C. B. (2015). Negative online consumer reviews: effects of different responses. *Journal of Product & Brand Management*, 24(1), 66-77.

Wan, S., & Law, R. (2017). *Leveraging Online Reviews in the Hotel Industry*. In Analytics in Smart Tourism Design (pp. 235-252). Springer International Publishing.

Xiang, Z., Schwartz, Z., Gerdes, 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.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51-65.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28, 180-182.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). 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*, *27*(2), 634-639.

Yu, Y., Li, X., & Jai, T. M. (2017). The impact of green experience on customer satisfaction: evidence from TripAdvisor. *International Journal of Contemporary Hospitality Management*, *29*(5), 1340-1361.

Zhang, K. Z., Cheung, C. M., & Lee, M. K. (2014). Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision. *International Journal of Information Management*, 34(2), 89-98.

Zhong, N., Li, Y. & Wu, S. (2012). Effective Pattern Discovery for Text Mining. *IEEE Transactions on Knowledge and Data Engineering,* 24(1), 30-44.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., & Corrigan, D. (2012). *Harness the Power of Big Data The IBM Big Data Platform.* McGraw Hill Professional.