International Journal of Hospitality Management xxx (2018) xxx-xxx

Contents lists available at ScienceDirect

International Journal of Hospitality Management

journal homepage: www.elsevier.com

Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews

Yabing Zhao[a], Xun Xu[*, b], Mingshu Wang[c]

[a] Department of Decision Sciences, College of Business, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, United States
[b] Department of Management, Operations, and Marketing, College of Business Administration, California State University, Stanislaus, One University Circle, Turlock, CA, 95382, United States
[c] Department of Geography, University of Georgia, 210 Field St, Athens, GA, 30602, United States

ARTICLE INFO

Keywords:
Online textual reviews
Technical attributes
Overall customer satisfaction
Hotel industry
Big data

ABSTRACT

Customer online reviews of hotels have significant business value in the e-commerce and big data era. Online textual reviews have an open-structured form, and the technical side, namely the linguistic attributes of online textual reviews, is still largely under-explored. Using a sample of 127,629 reviews from tripadvisor.com, this study predicts overall customer satisfaction using the technical attributes of online textual reviews and customers' involvement in the review community. We find that a higher level of subjectivity and readability and a longer length of textual review lead to lower overall customer satisfaction, and a higher level of diversity and sentiment polarity of textual review leads to higher overall customer satisfaction. We also find that customers' review involvement positively influences their overall satisfaction. We provide implications for hoteliers to better understand customer online review behavior and implement efficient online review management actions to use electronic word of mouth and enhance hotels' performance.

1. Introduction

In the e-tourism era, many customers book hotels online and post reviews after their stay. These online reviews, in the format of both textual reviews (comments) and ratings, generate an electronic-word-of-mouth (eWOM) effect, which influences future customer demand and hotels' financial performance and thus have significant business value (Xie et al., 2014).

Customers' ratings indicate their satisfaction, whose antecedents and influence have been extensively studied in the literature (e.g., Banerjee and Chua, 2016; Schuckert et al., 2015). One of the biggest strengths of researching customer ratings is that ratings can show overall customer satisfaction in a direct way. Recently, many studies have focused on textual reviews (Xiang et al., 2015; Berezina et al., 2016). The strengths of researching customer textual reviews are that they can show customer consumption experiences, highlight the product and service attributes customers care about, and provide customers' perceptions in a detailed way through the open-structure form. Researchers and hoteliers want to know both (a) the details about hotel guests' experiences to improve the corresponding product and service attributes and (b) customers' overall evaluation of the hotel stay experience to obtain a snapshot of the hotel's operational performance and overall customer satisfaction or to develop marketing strategies to better promote the hotel (Cantallops and Salvi, 2014).

However, two challenges exist when hoteliers try to understand both sides of the coin. The first challenge is the information overload of individual-level reviews or comments. Numerous comments in the open structure of online textual reviews or face-to-face conversations as feedback from hotel guests are available online and offline. The written comments often contain a substantial number of words and are time consuming to read one by one in detail. The second challenge is the lack of availability of a holistic satisfaction measure. In the face-to-face conversation environment, it is often hard to capture customers' overall evaluation of their hotel experience directly. Customers may not reveal their true evaluation, especially when they have a negative perception, because of worries about breaking the customer–seller relationship or concerns about the hotel "losing face" (Au et al., 2010). In some cases, it may be infeasible to develop a specific scale by which customers can give a single rating to evaluate the whole product or ser-

* Corresponding author.
Email addresses: Zhaoybzhao@sfsu.edu (Y. Zhao); xxu@csustan.edu (X. Xu); mswang@uga.edu (M. Wang)

https://doi.org/10.1016/j.ijhm.2018.03.017
Received 12 December 2017; Received in revised form 23 March 2018; Accepted 31 March 2018
Available online xxx
0278-4319/

vice. In the comment card and online review environment, customer comments as verbal protocols in terms of customers' online textual reviews, as opposed to direct measures, can avoid eliciting customers' perceptions (Smith and Bolton, 2002). The direct measurement of customer ratings in terms of closed-ended survey questions can confound the data of customers' true evaluation because of variations in survey design from different review platforms (Weber, 1985; Xiang et al., 2017).

A technical approach to link the relationship between customers' overall satisfaction and their textual comments is needed to address these major challenges. Technical attributes of textual reviews can explain significant variations in customer ratings, and technical attributes of online textual reviews can have a significant effect on customer ratings (Geetha et al., 2017). To link the two sides of the coin, this study uses customers' online review behavior to predict their overall satisfaction with hotels. Many previous studies focus on the indications and contents of customer online reviews (e.g., Xiang et al., 2015; Xu and Li, 2016), but few studies discuss the linguistic style, namely the technical attributes of the online reviews themselves (e.g., Geetha et al., 2017). The main reasons lie in the fact that examining technical attributes of online textual reviews is an extremely costly task with unstable and difficult-to-interpret measurements (Chevalier and Mayzlin, 2006; Godes and Mayzlin, 2004).

Previous studies have found inconsistency in customers' opinions mined from their textual reviews and their ratings, and the sentimental interplay between customer textual reviews and customer ratings can be influenced by their satisfaction level (Zhang et al., 2016b). Previous studies have found that the sentiments of online textual reviews and customer ratings are highly correlated (e.g., Geetha et al., 2017; He et al., 2017); however, the relationship among other technical attributes of online textual reviews, such as subjectivity, diversity, readability, length, and customer ratings is still largely under-explored (Geetha et al., 2017). To fill this research gap, this study aims to provide a full picture of the role of technical attributes of online textual reviews and bridge the technical aspects of customer reviews with their indications of overall satisfaction with hotels. We aim to understand how customers behave in writing online reviews in terms of what types of words they use and how long they write to reflect their overall evaluation of their hotel stay experience. This leads to the first question of this study: What is the effect of the linguistic attributes of online textual reviews, including subjectivity, diversity, readability, polarity, and length of individual review, on overall customer satisfaction? Subsequently, the second research question is as follows: Given those technical attributes of textual reviews, what are the most important technical attributes showing customer opinions about hotels, as measured by the highest influential level on customers' overall satisfaction?

In addition, different customers can exhibit different online review behaviors and perceptions of hotels depending on their demographic background, such as language group (Schuckert et al., 2015), and trip information, such as travel purpose (Xu et al., 2017). Different levels of review involvement and engagement in the online community (i.e., active or non-active) reveal customers' personalities and aspects of their hotel stay and review experience, which influence their perception of hotels (Zhang et al., 2010). However, the role of the reviewer's involvement in the online review community in influencing overall customer satisfaction is still unknown. To fill this research gap, we pose our third research question: What is the effect of review involvement on customers' overall satisfaction?

The main contribution of this study is that it bridges the technical side, namely the linguistic style of online reviews, with overall customer satisfaction. This is one of the first studies to investigate the role of technical variables of online customer reviews, including subjectivity, diversity, readability, sentiment polarity, and length of review, in predicting customers' overall satisfaction along with the role of customers' review involvement in influencing their overall satisfaction. In addition, the importance of the role of these technical variables of online customer reviews in influencing customers' overall satisfaction is examined.

Examining the relationship between the technical attributes of online textual reviews and customers' overall satisfaction can help hotels and online hotel booking agents to obtain richly structured descriptions of customers' sentiments and other technical information from the unstructured online textual reviews. It can also help them better design feedback systems to raise the quality of information received and thus to enhance their products and services based on customers' online textual reviews and ratings (Zhang et al., 2016b). The relationship between the technical attributes of online textual reviews and customers' ratings also influences future customers' demands because customers tend to read both textual reviews and ratings to justify their consistency (Chevalier and Mayzlin, 2006; Ludwig et al., 2013). Customer ratings supported by lengthy textual reviews containing rich information are favored by customers, and thus hotels should identify and promote the most influential reviews and provide instructions to motivate customers to write powerful reviews (Ludwig et al., 2013).

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 proposes the hypotheses. Section 4 introduces the methodology. Section 5 presents the results. Section 6 discusses the results. Section 7 provides theoretical and managerial implications, and Section 8 concludes the study.

2. Literature review

2.1. Motivation and impact of hotel online reviews

Customers are generally motivated by four incentives to write online reviews. The first is altruism and reciprocity. Customers posting online reviews based on this motive seek to help future hotel guests make better decisions about hotel stay choices and help hotels improve their service operations (Yoo and Gretzel, 2011). The second is fulfilling customers' psychosocial needs. Customers posting online reviews for this reason can show their satisfaction and admiration or their dissatisfaction and complaints toward a hotel (Cantallops and Salvi, 2014). The third is customers' social needs. They want to obtain a positive reputation in an online community such as by being voted "helpful" (Kwok and Xie, 2016), gain social identification in the travel community (Cheung and Thadani, 2012), or anticipate hotel managers' online responses (Gu and Ye, 2014). The fourth incentive is economic by which they earn rewards from an online review platform when they post reviews (Hennig-Thurau et al., 2004). Customers' linguistic style is influenced by their motivation for writing online reviews (Ludwig et al., 2013).

The impacts of online reviews are mainly disseminated through the generated eWOM and include influencing future customers' purchase intentions, trust, customer demand, and hotels' financial performance (Sparks and Browning, 2011; Vermeulen and Seegers, 2009). Hotels use online customer reviews to understand customers' expectations and needs and improve the corresponding products and services (Gu and Ye, 2014).

2.2. Examining hotel online textual reviews

Compared with ratings, which are structured, online textual reviews are unstructured user-generated contents (Zhang et al., 2016b). Thus, online textual reviews can reflect customers' consumption experience and perceptions in more detail compared with customer ratings (Xu and Li, 2016). Previous studies focusing on hotel online reviews can be categorized into two types. The first category focuses on the contents of textual reviews to find the attributes mentioned by hotel guests and their perceptions of their hotel stay experiences. These attributes include room quality, staff attitude and behavior, location, access, value, food, and so on (Xu and Li, 2016). The perceptions include customer satisfaction and dissatisfaction (Berezina et al., 2016) based on the assumptions that positive reviews indicate satisfaction and negative reviews indicate dissatisfaction. Many studies use text mining techniques including content analysis (Li et al., 2013), frequency analysis (Xiang et al., 2015), text-link analysis (Berezina et al., 2016), and latent semantic analysis (Xu and Li, 2016) to examine attributes of the hotel products and services that customers care about. Zhang and Mao (2012) used content analysis of online reviews to predict customers' revisiting intention and referrals of the hotel.

Recently, a second category of hotel online reviews has emerged and drawn increasing attention in online review studies: the technical side of hotel textual reviews. The technical analytics of online reviews can help hospitality companies make forecasts such as review helpfulness (Ma et al., 2018), customer conversion rates (Ludwig et al., 2013), and hotel performance (Blal and Sturman, 2014). Gao et al. (2018) claimed that online textual reviews are rich opinion resources and used sentiment analysis to extract comparative relations from online textual reviews of restaurants to help restaurants identify their competitors to gain competitiveness.

Regarding the relationship between the technical side of hotel textual reviews and online customer ratings, Geetha et al. (2017) focused on the sentiment polarity of online customer reviews and found that it influences customer ratings. He et al. (2017) used natural language preprocessing, text mining, and sentiment analysis techniques to analyze online hotel textual reviews, and they found that the sentiment scores of the title and contents of online customer reviews had a high correlation with overall customer ratings of hotels. Their results confirmed Qu et al.'s (2008) study indicating that most attribute sentiments derived from the textual reviews were significantly correlated with customers' overall rating. This also supports the findings from Kim et al.'s (2015) study showing that overall ratings are the most critical predictor of hotel performance.

Because online reviews can be written by customers who have different cultural backgrounds and different languages, examining the technical attributes of textual reviews written in different languages and with different cultural backgrounds can be meaningful (Tian et al., 2016). Regarding language, Tian et al. (2016) claimed that examining the sentiment of textual reviews written in different languages can help hotel managers better understand their customers and improve the corresponding products and services. Wu et al. (2017) examined the language style of online reviews and uncovered that the persuasive power is different between figurative and literal language styles. The authors employed a text mining approach to find the influence of the linguistic style of online reviews on customers' conversion rates among product websites (Ludwig et al., 2013). Regarding culture, Zhang et al. (2016b) found that because Chinese consumers have been identified as typical collectivists, they behave differently from people in Western countries, and thus they exhibit a relatively looser sentimental interplay between the textual review and ratings. They also found that satisfied or neutral consumers are more likely to show confounding sentiment signals in relation to the textual review and ratings.

The technical attributes of textual reviews can also be influenced by many other factors. Xiang et al. (2017) compared different online review platforms between booking websites and social media and concluded that the information quality between these platforms is different. Zhang et al. (2010) examined the demographic information of the writers of the online reviews and compared the persuasive power of online reviews written by customers and editors.

Our study contributes to this research stream: the technical side study of hotel textual reviews by examining customers' evaluation of hotels through the linguistic style of their online reviews. We extend Geetha et al.'s (2017) study by including more technical variables of customer reviews: subjectivity, diversity, readability, and length. In addition, the role of the reviewers' identity: review involvement in influencing their ratings, is also discussed. We use big data analytics to predict overall customer satisfaction through various technical variables of customer reviews, which can meet the hotel owners' need to predict the future performance of hotels, benchmark properties, forecast occupancy rates, and improve the corresponding operations in the fierce competition among hotels (Pan and Yang, 2017).

3. Hypotheses development

3.1. Theoretical background

Signal theory provides a theoretical foundation for this study. Signal theory describes the signal behavior between two parties where information asymmetry exists (Connelly et al., 2011). Many products and services offered by hotels are intangible, which leads to information asymmetry between hotels and customers about the quality of products and services. Customers write online reviews after their stay, and the contents and linguistic characteristics of customer online reviews serve as signals about their perceptions of their hotel stay to hotel managers and future customers (Casaló et al., 2015; Geetha et al., 2017). What customers write (i.e., the contents) and how customers write (i.e., the linguistic style) signal their satisfaction or dissatisfaction with hotel product and service attributes. As an indirect communication approach to hoteliers and future customers, customer online reviews efficiently alleviate the effects of information asymmetry and strongly influence future customers' hotel booking intentions and behavior (Cantallops and Salvi, 2014).

Customer online ratings show their satisfaction with hotels. According to expectation-confirmation theory, the generation mechanism of customer satisfaction is the comparison between pre-purchase expectation and perceived quality of products and services after consumption. If customers' perceived quality is higher than their expectation, customers are satisfied. If not, they are dissatisfied (Oliver, 1980). In online reviews, customers mention the perceived quality of products and services, their pre-expectation, or both to show why they are satisfied or dissatisfied.

3.2. Subjectivity

Objective information describes hotel products and services; any other information, such as expressing emotion in online reviews, is considered "subjective." Objective information reflects cognitive behavior, and subjective information reflects affective behavior (Anand et al., 1988). Customers with cognitive behavior often compare current experiences with past experience and thus are more rational (Rose et al., 2011). Customers with affective behavior are more likely to complain, showing their affective dissatisfaction (Heung and Lam, 2003). Customers writing subjective reviews are more emotional and thus tend to generate more extreme, negative evaluations toward hotels when they perceive that the product and service offerings are unfair (Schoefer and Ennew, 2005). We propose the following hypothesis:

H1

The subjectivity of online reviews has a negative effect on customer ratings.

3.3. Diversity

Customers usually use diverse words to generally describe several positive attributes of hotel products and services (Xiang et al., 2015). Diversity for the purpose of this study refers to the redundancy of words in online reviews. Higher diversity indicates that customers use fewer redundant words in their online reviews. The diversity of words in positive reviews comes from both the multiple attributes of the hotel products and services the customers described and the descriptive words they used (Xiang et al., 2015). Negative reviews reflect customer complaint behavior as a way to release negative emotion, warn future customers, and seek hotels' responses and compensation (Gu and Ye, 2014). Negative reviews often focus on specific aspects of the product and service attributes that the customers are dissatisfied with, and they tend to describe these attributes in detail (Xu and Li, 2016). While complaining, customers tend to use similar words to describe certain products and services and show future customers and hoteliers why they are dissatisfied with the hotel (Berezina et al., 2016). Extremely negative words are often repetitively used to express customers' criticism and complaints (Bradley et al., 2015). We propose the following hypothesis:

H2
The diversity of online reviews has a positive effect on customer ratings.

3.4. Readability

Readability refers to the difficulty of understanding the meaning of online reviews. Higher readability indicates that readers require a higher level of education and maturity to understand the meaning of the texts. The linguistic style of a review with a higher readability score usually implies that the writer is more educated (Hu et al., 2012). People with higher-level education are more likely to be critical, which arouses negative emotion to generate customer dissatisfaction (Westbrook and Oliver, 1991). When detailing the cons of hotel products and services, customers tend to use more words with higher complexity (Xu and Li, 2016). Customers also tend to use more advanced words to describe their experiences in detail when they are unsatisfied with hotels' product and service attributes and wish to persuade hoteliers and future customers (Xu and Li, 2016). The following hypothesis is proposed:

H3
The readability of online reviews has a negative effect on customer ratings.

3.5. Sentiment polarity

Sentiment implies customer emotion, including negative extreme emotions such as frustration and anger and positive extreme emotions such as delight or excitement (Geetha et al., 2017). Sentiment polarity is the degree of positive or negative sentiment that customers express when writing online reviews. Higher polarity shows more positive sentiment. Positive emotions can enhance the perceived quality of products and services, which is an antecedent for customer satisfaction, while negative emotions are an antecedent for customer dissatisfaction (Dai et al., 2015). Customers tend to evaluate their consumption experience more positively when they are in a positive emotional state compared with when they are in a negative emotional state (Isen, 1987). Negative emotions trigger customers' criticism and induces them to provide a biased evaluation of their experience and rate it more negatively (McColl-Kennedy and Sparks, 2003). We thus propose the following hypothesis:

H4
The sentiment polarity of online reviews has a positive effect on customer ratings.

3.6. Review length

Customers tend to post more words and sentences with more detailed descriptions of the negative aspects of hotels' products and services compared with their description of the positive aspects (Xu and Li, 2016). Longer reviews indicate that customers put more review effort into commenting on products and services (Chevalier and Mayzlin, 2006), which often happens when they experience a negative consumption emotion (Verhagen et al., 2013). Customers use more words to express their frustration, anger, and depression when they encounter the cons of products and services (Berezina et al., 2016). Customer complaints are frequently mixed with neutral or even positive reviews, which makes online reviews longer (Bradley et al., 2015). Customers with negative perceptions tend to write more detailed reviews to seek identification and support from the travel community and make their reviews more persuasive (Salehan and Kim, 2016). The following hypothesis is presented:

H5
The length of online reviews has a negative effect on customer ratings.

3.7. Review involvement

Customers' involvement in the online review community as reviewers influences their ratings. Customers with high review involvement are frequent hotel customers who have more experience and thus provide more professional reviews. Customers' higher involvement in online reviews indicate their higher expertise in evaluating hotel products and services because they have a higher a degree of competence and knowledge about hotel operations (Liu and Park, 2015). Future customers often seek reviews written by review experts (Zhang et al., 2010). In turn, review experts often seek more helpfulness votes from future potential customers and more readership in recognition of their contribution and impact in the review community (Salehan and Kim, 2016). This motivates frequent review writers to post more objective reviews, which will help future customers to better choose hotel options, rather than subjective and emotional reviews (Bronner and De Hoog, 2011). Customers with higher review involvement are also more confident about making judgements and have higher self-esteem (Zhou and Guo, 2017). Thus, they are less influenced by other reviewers and can often provide a more objective evaluation of hotel products and services (Clark and Goldsmith, 2005).

Frequent hotel customers are also more experienced and have more chances to compare products and services between hotels, which enhances their capability to provide more objective reviews. In addition, frequent hotel customers can better understand the operations of hotels, and thus usually have a higher tolerance than non-frequent hotel customers when service failures happen (Lewis and McCann, 2004). Moreover, frequent hotel customers writing online reviews are often motivated by altruism and reciprocity instead of vengeance and the need to vent, so they tend to write fewer extremely negative evaluations of hotels (Yoo and Gretzel, 2011). The following hypothesis is thus proposed:

H6
Customers' involvement in the online review community has a positive effect on customer ratings.

4. Data analysis

4.1. Data collection

We collected data from tripadvisor.com. There are two reasons for this choice. First, tripadvisor.com is one of the largest world social media platforms dedicated to travel. It has more than 300 million members and 500 million reviews of hotels, restaurants, and other travel-related businesses around the world, making it easy to collect big data of online reviews. Second, tripadvisor.com has implemented many methods to check the quality of each online customer review to ensure a relatively high quality of review contents. It scrutinizes the IP and email addresses of the online review writer and tries to detect suspicious patterns and obscene or abusive language before a review is posted to the website. It also allows users to report suspicious contents, and these reports are followed up with an assessment by a team of quality assurance specialists. This ensures the validity of the online customer reviews.

We developed a customized Python program, which automatically collected online reviews of hotels available on TripAdvisor's website in June 2017. The coding and data analytics were developed based on the following five steps. First, among some candidate Python libraries (e.g., Beautiful Soup, Scrapy, Selenium), we selected Selenium (version 3.9.0) because it is a web browser automation tool that is capable of handling dynamically generated web pages to collect all the data visible on the websites in a real time. Second, we followed the framework of the Selenium package to develop several Python scripts to extract the contents of interest to this study (i.e., a list of all hotels in San Francisco, overall customer ratings, customer textual reviews, hotel ranking, user profile). Because most hotels' reviews are posted across multiple pages, we also had to employ pagination in Selenium. Third, we ran the customized scripts to automatically extract all reviews for each hotel and repeated this process for all hotels in San Francisco. The program automatically opened each hotel's web page and searched for each review comment for that hotel. Once a review comment block was identified based on its html pattern, the program would extract relevant information as previously coded (review text, user profile, etc.) and save the parsed text (e.g., removing non-ASCII characters) to a local database (i.e., MySQL). The fourth step was post-processing. After obtaining the complete review data in MySQL, we used the Testimonial toolkit and the methods introduced in Section 4.2 to compute all technical attribute variables of online reviews needed for this study. Finally, we converted the data format and imported the data into SAS for empirical analysis.

We took San Francisco as a sample city to collect data because of the high popularity of hotels and their online reviews. We excluded records with incomplete information required in this study (e.g., textual review, overall rating). We also excluded hotels with fewer than 10 reviews to reduce review bias. The final sample yielded 127,629 individual-level reviews for 155 out of 217 hotels in San Francisco from April 2001 to June 2017. For each review sample, customer textual reviews, overall customer ratings, and customer involvement (indicated by contributor level endorsed by TripAdvisor) information was collected directly from the website. Fig. 1 shows an example of an eligible review in our sample.

4.2. Variables and measurements

The dependent variable in this study was a customer's self-reported overall rating of the hotel on a scale of one to five, which has been widely used in literature and is extracted directly from various online platforms (e.g., Ganu et al., 2013, Geetha et al., 2017; Liu and Park, 2015). The independent and control variables are described in detail below.

We calculated both *subjectivity* and *polarity* measurements based on the Stanford Natural Language Toolkit (NLTK) with a naïve Bayes classifier (Manning et al., 2014), which has drawn tremendous attention in academia (e.g., Giatsoglou et al., 2017; Gu and Kim, 2015; Krishna et al., 2017). Specifically, we implemented the Testimonial toolkit formula in the TextBlob Python library (Giatsoglou et al., 2017; Loria et al., 2014; Micu et al., 2017) and used a sentiment analysis tool in the library that uses deep learning techniques to calculate subjectivity and polarity measurements. The subjectivity score ranges from 0 to 1, where a higher value indicates a more subjective text. A smaller value of subjectivity indicates that more objective words are used to describe products and services instead of revealing emotions or evaluating (Giatsoglou et al., 2017; Saif et al., 2016). Similarly, the polarity score is a continuous variable from -1 to 1. A greater value for the polarity score indicates a more positive sentiment (emotion) of the text, with 1 showing extremely positive sentiment, such as excitement and delight, and -1 showing extremely negative sentiment, such as frustration and anger. A value of 0 shows neutral sentiment (Cho et al., 2014; Deng et al., 2017; Geetha et al., 2017). Both subjectivity and polarity measure-

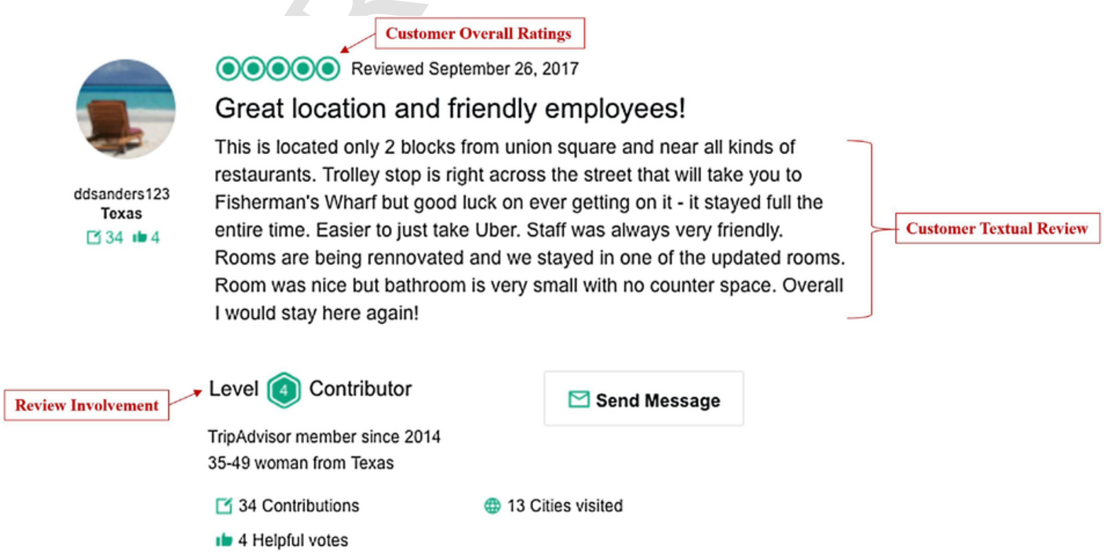

Fig. 1. Screenshot of one review sample on Tripadvior.com.

ments have been introduced and discussed in previous studies as part of sentiment analysis (e.g., Cho et al., 2014; Deng et al., 2017; Giatsoglou et al., 2017; Saif et al., 2016).

Diversity refers to the lexical diversity of a review and ranges from 0 to 1 based on a linguistic metrics measurement calculated by the ratio of unique words to total words in the text (Lahuerta-Otero and Cordero-Gutiérrez, 2016; Zhang et al., 2016a). A higher value suggests fewer redundant words and more lexical diversity in the review.

Readability refers to how easily a reader can understand a text and includes two aspects of the text, namely content and presentation. While the presentation of reviews (e.g., typography and web page design) is homogeneously defined by TripAdvisor, the contents of reviews have various levels of vocabulary and syntactical complexity. We used the Gunning Fog Index (Gunning, 1969) as our readability measure because it has been employed in extant studies (Fang et al., 2016, Li et al., 2017). The Gunning Fog Index is one of the best-known readability measures and has been widely applied to the level of reading difficulty for diverse types of writing. The Gunning Fog Index estimates the number of years of formal education in the U.S. school system that a person needs to understand a text on the first read. We used Python library to compute readability.

Length is the review length measured by the number of words in each online review (Li et al., 2017; Liu and Park, 2015; Zhang et al., 2016a). Because the number of words is widely distributed (ranging from fewer than 10 words to thousands of words), we took the natural logarithm transformation to deflate and normalize this measurement in the data analysis.

Involvement indicates a reviewer's involvement in the online review community. It is measured by a user's contribution level endorsed by TripAdvisor and ranges from 0 to 6, where a Level 6 contributor indicates that the reviewer is engaged in the TripAdvisor community to the greatest extent in terms of the number of reviews posted (Filieri et al., 2015; Liu et al., 2018). Our measurement is consistent with those of previous studies (e.g., Liu et al., 2018) that used the badge level of a member to measure a reviewer's involvement. Our measurement is also consistent with previous studies (e.g., Liu and Park, 2015; Zhou and Guo, 2017) that used the number of previous reviews written by a reviewer to measure review expertise.

Hotel ranking is a unique ranking for each hotel given by TripAdvisor directly based on the overall reviews the hotel received. Referring to previous studies (e.g., Fang et al., 2016), we used this as a control variable because individual customers may be able to give higher ratings for and have more positive perceptions of highly ranked hotels (Casaló et al., 2015; Ye et al., 2009). Hotel or attraction ranking is relatively stable over a long period because of their relatively stable popularity (Fang et al., 2016). In our sample, the hotels were ranked from 1 (highest) to 217 (lowest) without continuum because some hotels with few reviews were excluded from the sample. The description, value range, and method to construct all the variables are summarized in Table 1. A flow chart of the analysis of this study can be found in Fig. 2.

5. Empirical results

5.1. Main results

The descriptive statistics of all variables are provided in Table 2. We conducted a multivariate linear regression analysis following the model specified in Eq. (1). The regression results are presented in Table 3. To address the potential violations of OLS assumptions, we used robust (heteroscedasticity-consistent) standard errors to estimate *t*-statistics in our regression analysis. Variance inflation factors (VIF) were re-

Table 1
Description of the variables.

Variable	Description	Measurement	Method	Reference
Customer Ratings	Overall customer evaluation of the hotel from one to five ratings	Interval	Collected directly from TripAdvisor	Ganu et al. (2013), Geetha et al. (2017), and Liu and Park (2015)
Subjectivity	Review sentiment subjectivity; a higher value indicates a more subjective review	Ratio	Calculated by Testimonial toolkit in TextBlob Python library	Deng et al. (2017), Giatsoglou et al. (2017), Loria et al. (2014), and Saif et al. (2016)
Diversity	Review lexical diversity with a linguistic matrix, where a higher value suggests fewer redundant words and higher lexical diversity	Ratio	The number of unique words/ The number of total words	Lahuerta-Otero and Cordero-Gutiérrez (2016) and Zhang et al. (2016a)
Readability	Measured by Gunning Fog Index, a readability test in linguistics that estimates the years of formal education in the U.S. school system that a person needs to understand a text on the first reading	Interval	Calculated by *readability* in TextBlob Python library	Fang et al. (2016), Gunning (1969), and Li et al. (2017)
Polarity	Review sentiment polarity; a higher value indicates more positive emotion	Ratio	Calculated by Testimonial toolkit in TextBlob Python library	Cho et al. (2014), Deng et al. (2017), Geetha et al. (2017), Loria et al. (2014) and Micu et al. (2017)
Length	Natural logarithm of the number of words in the textual reviews	Ratio	Natural logarithm of the word count in the review	Li et al. (2017), Liu and Park (2015), and Zhang et al. (2016a)
Involvement	Measured by user contribution level endorsed by TripAdvisor; higher involvement level indicates that the reviewer is engaged in more activities in the TripAdvisor review community	Interval	Collected directly from TripAdvisor	Filieri et al. (2015), Liu et al. (2018), Liu and Park (2015) and Zhou and Guo (2017)
Hotel Ranking	Endorsed by TripAdvisor; a smaller value indicates a more favorable ranking (1 is the best)	Ordinal	Collected directly from TripAdvisor	Casaló et al. (2015), Fang et al. (2016) and Ye et al. (2009)

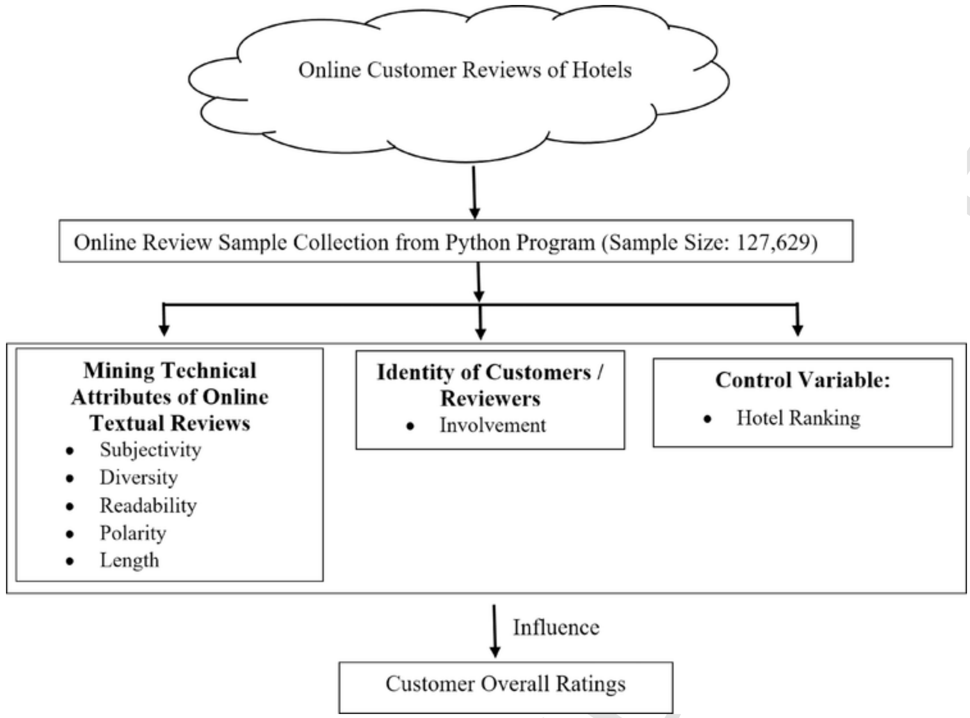

Fig. 2. Research design framework.

Table 2
Descriptive statistics of variables.

Variable	Mean	Std. Dev.	Min	Max
Customer Ratings	4.01	1.08	1	5
Subjectivity	0.54	0.16	0	1
Diversity	0.76	0.11	0.25	1
Readability	7.68	2.44	2.40	22
Polarity	0.25	0.18	−1	1
Length	4.38	1.14	0	7.82
Involvement	3.01	2.04	0	6
Hotel Ranking	68.28	50.08	1	217

Remark: Number of observations = 127,629; *Length* is log-transformed value.

Table 3
Regression results.

Variable	Coefficient Estimation	Standard Error	*t*-stat Value	VIF
Intercept	4.02***	0.05	88.17	
Subjectivity	−0.83***	0.03	−32.56	2.10
Diversity	0.27***	0.04	6.42	3.78
Readability	−0.01***	<0.01	−4.86	1.45
Polarity	3.19***	0.02	155.79	1.55
Length	−0.03***	<0.01	−5.44	5.07
Involvement	0.01***	<0.01	4.23	1.04
Hotel Ranking	−0.01***	<0.01	−112.97	1.08
Adjusted R^2	38.51%			
F Value	11,421.9***			
Number of Observations *n*	127,629			

Remark: *$p < 0.1$, **$p < 0.01$, *** $p < 0.001$.

ported to provide evidence that multicollinearity issues are not a concern in our data because all VIFs are well below the typical benchmark value of 10 (Neto et al., 2016; Wooldridge, 2015; Zhou and Li, 2012). The Durbin–Watson statistical test score is 1.893, suggesting no presence of autocorrelation. The results in Table 3 indicate that all hypotheses are supported.

$$\begin{aligned}
Customer\ Ratings_{ij} = \beta_0 &+ \beta_1 \text{Subjectivity}_{ij} \\
&+ \beta_2 \text{Diversity}_{ij} \\
&+ \beta_3 \text{Readability}_{ij} + \beta_4 \text{Polarity}_{ij} \\
&+ \beta_5 \text{Length}_{ij} + \beta_6 \text{Involvement}_i \\
&+ \beta_7 \text{Hotel Ranking}_j + \varepsilon_{ij},
\end{aligned} \tag{1}$$

where the subscript *i* and *j* indicate reviewer *i* and hotel *j*, respectively.

5.2. Robustness check

Following previous studies (e.g., Zhou and Guo, 2017), we conducted an additional analysis to ensure the robustness of our empirical results with alternative measurement of control variables and alternative model specifications.

5.2.1. Alternative measurement of control variable

Because hotel rankings might change over time, we used hotel star ratings (ranging from 0 to 5) as an alternative measurement of the control variable, in which the star level of a hotel is relatively stable during the period (Xiang et al., 2015). From the results of Model 1 in Table 4, we found consistent results and came to the same conclusions regarding the hypotheses of technical attributes of online reviews as found in the main results. We then tested the model without any control variable included (i.e., Model 2 in Table 4) and the model with only technical attributes included (i.e., Model 3 in Table 4). The main results are still consistent, and conclusions regarding hypotheses are the same as those reached with the main results.

5.2.2. Alternative model specifications

We used a fixed effect model to test whether our results are consistent and robust. We aimed to determine whether there exists a hotel- or reviewer-specific effect across online reviews. We treated the variables of user involvement (shown in Table 5a), hotel ranking (shown in Table 5b), and hotel star level (shown in Table 5c) as a fixed effect to

Table 4
Robustness test results through alternative measurement of control variable.

Variable	Coefficient Estimation (Standard Error)		
	Model 1	Model 2	Model 3
Intercept	2.91*** (0.05)	3.66*** (0.05)	3.73*** (0.05)
Subjectivity	−0.92*** (0.02)	−0.87*** (0.02)	−0.90*** (0.02)
Diversity	0.27*** (0.04)	0.12*** (0.04)	0.11*** (0.04)
Readability	−0.01*** (0.00)	−0.01*** (0.00)	−0.01*** (0.00)
Polarity	3.52*** (0.02)	3.65*** (0.02)	3.66***(0.02)
Length	−0.03***(0.01)	−0.05***(0.01)	−0.04***(0.01)
Involvement	0.01***(0.01)	0.02***(0.01)	N/A
Hotel Star Level	0.18*** (0.01)	N/A	N/A
Adjusted R^2	33.15%	31.09%	30.93%
F Value	9044.30***	9599.26***	11432.8***
Number of Observations n	127,629	127,629	127,629

Remark: *$p < 0.1$, **$p < 0.01$, *** $p < 0.001$.

Table 5
Robustness test results through alternative model specifications.

Variable	Coefficient Estimation (Standard Error)		
	Model 1	Model 2	Model 3
Intercept	4.04*** (0.04)	2.34*** (0.22)	3.87*** (0.05)
Subjectivity	−0.83*** (0.02)	−0.79*** (0.02)	−0.88*** (0.02)
Diversity	0.27*** (0.04)	0.27*** (0.04)	0.31*** (0.04)
Readability	−0.01*** (0.00)	−0.01*** (0.00)	−0.01*** (0.00)
Polarity	3.19*** (0.02)	3.13*** (0.02)	3.41*** (0.02)
Length	−0.03*** (0.00)	−0.03*** (0.00)	−0.03*** (0.00)
Involvement	Fixed Effect	0.01*** (0.00)	0.01*** (0.00)
Hotel Ranking	−0.01*** (0.00)	Fixed Effect	N/A
Hotel Star Level	N/A	N/A	Fixed Effect
Adjusted R^2	38.53%	39.62%	34.87%
F Value	6664.65***	522.76***	4880.82***
Number of Observations n	127,629	127,629	127,629

rerun the analysis, and the results in Table 5 suggest that the respective conclusions of the influence of technical attributes of online reviews on overall customer satisfaction are the same as for the main results (as seen in Table 3).

6. Discussion

6.1. Technical attributes of customer textual reviews

Our results support H1: higher subjectivity of an online review leads to lower customer ratings. Customers often consider the online review platform a place to complain about their consumption experience. The emotions of frustration and anger are revealed when they complain (Sparks et al., 2016). This drives them to post more emotional words and details, showing their negative perceptions in the online reviews, which results in higher subjectivity of the review text. These subjective complaints caused by customer dissatisfaction with hotel products and services expressed through online reviews are reflected in low ratings.

H2 is also supported: higher diversity of online reviews leads to higher customer ratings. Customers tend to use more varied words to describe multiple product and service attributes to praise hotels and more similar words to describe certain product and service attributes negatively and give detailed reasons that they are dissatisfied with the hotels. The negative reviews contain more similar extremely negative words to express customers' complaints and criticism.

The analysis supports H3: higher readability has a negative effect on customer rating. The online reviews with higher readability use more advanced words and are more likely to be written by customers with higher education levels. People with higher education tend to engage more in critical thinking, so their online reviews are more critical, and, in turn, the higher readability of reviews is associated with lower overall ratings. Customers also tend to describe the cons of the hotel products and services in more detail and use more advanced words compared with describing pros to elaborate the detailed reasons for their dissatisfaction (Xu and Li, 2016).

Our results support H4: higher sentiment polarity leads to higher customer ratings. Sentiment polarity reflects customer emotions when writing online reviews. Higher sentiment polarity indicates that more positive words than negative words are used in their online reviews (Geetha et al., 2017). Customers with higher sentiment polarity in their online reviews express positive emotions, such as excitement and delight, which leads to higher ratings.

H5 is also supported by the results: online reviews of longer length lead to lower customer ratings. Negative reviews are usually longer compared to positive reviews with one or more complaints incorporated (Bradley et al., 2015). Customers who provide negative descriptions of the hotel's products and services often use more words to seek public revenge, engage with other customers, and inform future customers about their awful experience (Sparks and Browning, 2010), so they post longer reviews and low ratings online.

6.2. Review involvement

Our results support H6: higher review involvement of customers makes them rate the hotels higher. Customers with more involvement in the online review community have stayed in more hotels, which makes it easier for them to compare hotels. Their reviews are more like expert reviews, which are more objective. In addition, frequent online reviewers are more motivated to show altruism and reciprocity to help future customers to choose hotels, which discourages them from complaining and posting extremely negative evaluations compared with less frequent online reviewers (Yoo and Gretzel, 2011; Bradley et al., 2015). Furthermore, frequent hotel guests tend to be more tolerant because they have stayed at many hotels and know better how to resolve unpleasant experiences during their stay rather than just posting very negative ratings online to express their dissatisfaction. Their extensive hotel stay experience also makes their evaluation of hotel products and services less biased.

6.3. The relative importance of the variables

Based on the data analysis using multiple regression, we found that all independent variables contribute significantly to overall customer satisfaction. To compare the relative importance of the variables, we examined the standardized coefficient $\hat{\beta}$of each independent variable. We found that sentiment polarity has the highest influence on overall customer satisfaction ($\hat{\beta}_4 = 0.53$), with subjectivity the second highest influence ($\hat{\beta}_1 = -0.12$), and review length ($\hat{\beta}_5 = -0.03$), diversity (

$\hat{\beta}_2 = 0.03$), readability ($\hat{\beta}_3 = -0.01$), and review involvement ($\hat{\beta}_6 = 0.01$) follow.

The reason that sentiment polarity and subjectivity have a higher influence on customer satisfaction in comparison with other independent variables in our study is that sentiment polarity describes the consumption emotion of customers and the emotions expressed through their textual reviews (Geetha et al., 2017), which highly influence customer satisfaction (Mano and Oliver, 1993). Customers tend to evaluate products more positively when they are in a positive emotional state than when they are in a negative emotional state (Isen, 1987). When customers experience positive emotions facing services, they tend to adopt acceptance behavior and generate more satisfaction (Yalch and Spangenberg, 2000). When customers experience more negative extreme emotions, they are more likely to exhibit detailed, systematic, and complex judgmental processes (Forgas, 1994). Customers become more critical in their thinking when they are in a negative emotional state (McColl-Kennedy and Sparks, 2003). The negative emotions trigger customers to engage in counterfactual thinking and evaluate products and services more negatively (McColl-Kennedy and Sparks, 2003).

Subjectivity shows customers' cognitive and affective level (Anand et al., 1988). Higher subjectivity shows that a customer is more affective and thus more likely to complain and express their dissatisfaction (Heung and Lam, 2003); lower subjectivity shows that a customer is more cognitive, which makes him or her compare experience with expectation and past experience and thus judge products and services more rationally (Andreassen and Lindestad, 1998). Although the review length, diversity, readability, and review involvement have relatively less influence on customer satisfaction compared with sentiment polarity and subjectivity, they still cannot be ignored because of their significance of effect on customer satisfaction, as Table 3 shows.

7. Theoretical and managerial implications

7.1. Theoretical implications

Many hospitality studies have focused on online customer reviews with the higher popularity of online users and the availability of online customer review data. Compared with most previous studies, which focus on the contents of online reviews of hotels, this study focuses on the technical attributes of online reviews to examine the relationship between the customers' writing style of online reviews and their overall satisfaction. The study provides four main theoretical implications and contributions.

First, this study shows the relationship between customer online textual reviews and ratings. Compared with ratings, textual reviews can more fully reflect the customer's consumption experience and perception in detail because of their open structure. Our study focused on this open structure and found that customers' linguistic style in writing online reviews serves as a signal to predict their overall satisfaction. This supports and extends signal theory by revealing that the linguistic characteristics of an online review signal overall customer evaluation of hotels to future customers and hoteliers.

Second, the findings of this study provide a comprehensive view of the roles of technical attributes of the online reviews in predicting customer ratings. Our paper is one of the first to introduce several new attributes of online reviews in hospitality, such as subjectivity, diversity, readability, and so forth. Also, the relative importance of these technical attributes is compared. The results show the added business value of the technical attributes of online reviews.

Third, this study examines the role of customer identity in terms of customers' involvement in the online review community in influencing their satisfaction. Aspects of customers' identities lead them to have different needs for hotel products and services, different perceptions, and different online review behaviors among various segments of customers. This reveals the role of customer identity in the expectation-confirmation theory about the generation of customer satisfaction.

Last, one of the complexities of researching online customer reviews is the substantial amount and open structure of its information. This paper uses a sample of 127,629 reviews to show how to use big data to analyze the business value of online textual reviews in the hospitality industry. We were able to deal with the huge amount of information by mining the technical attributes of online textual reviews. It provides a roadmap to examine the patterns for writing online reviews and the generated eWOM effect from the technical attributes of online customer textual reviews.

7.2. Managerial implications

Customer ratings are direct measurements of customers' perceptions. Textual reviews measure customer perception with verbal protocols, which are indirect measurements of customers' perception and satisfaction, with the advantage of avoiding eliciting customers' perception that otherwise might not have appeared in the evaluations (Smith and Bolton, 2002). In this way, customer textual reviews reflect the customer's perception and consumption experience more fully.

The findings of this study can motivate hoteliers to mine more attributes from customer textual reviews and to investigate customer online review behavior and its relationship with overall customer ratings in depth. These online reviews generate a high eWOM effect that influences future customers' booking decisions. However, because of the open structure of online reviews and their substantial information, online review analysis remains challenging.

Our study uses data mining methodologies that offer a practical approach for hoteliers to understand the linguistic style of online customer reviews and how these textual reviews are related to customers' overall ratings. Hoteliers not only need to be aware of the contents of the online reviews and improve the products and services customers give feedback on but also need to emphasize the technical attributes implied by the online reviews.

Providing prompt and efficient responses to online customers' negative reviews is an effective approach to implement service recovery actions and retain customers (Gu and Ye, 2014). Hotels may try to learn about unsatisfactory experiences from customer reviews for future improvement. However, faced with limited resources and priority rules, hoteliers should focus on online textual reviews that have more subjective words showing personal emotion (higher subjectivity), fewer diverse words (lower diversity), more advanced words (higher readability), more negative emotion (lower sentiment polarity), and longer length (more words). Although this may take more time compared with dealing with short, easy online textual reviews (e.g., shorter reviews with lower readability), these review attributes reflect higher levels of customer dissatisfaction and should be targeted for response first. By addressing these reviews and taking the appropriate actions, hotels can improve their service and reputation and benefit from the more positive eWOM effect.

Hoteliers should focus on online reviews written by nonfrequent travelers with less review involvement in the online review community because their perceptions of hotels tend to be more negative compared with those of frequent travelers. Providing efficient service guidance and communication, especially when service failures happen, can alleviate negative perceptions and enhance tolerance of service quality (Anderson et al., 2009). Prompt online response with a commitment to service improvement and compensation can be helpful to reduce the dissatisfaction of nonfrequent travelers and maintain the loyalty of future customers (Gu and Ye, 2014). Hotels can also develop promotion programs to motivate customers to be more actively engaged in the online review community and thereby benefit from more positive eWOM effect to attract future customers.

Open face-to-face discussion with customers and providing customer comment cards are other ways to obtain indirect and open perceptions of customers. Hoteliers can translate a face-to-face conversation to text data and use the methodologies in this study to generate different technical attributes of the text so they can also predict overall customer satisfaction from those conversations and comment cards. For some customers, such as those from Asian countries, there is a culture of longer power distance and face threat when they show their dissatisfaction directly by ratings (Zourrig et al., 2009). Customers may feel uncomfortable evaluating the hotel product and service directly when they have an extremely negative perception. Thus, indirect measurements of their perception from conversation and comments avoid eliciting their perception directly and can help hoteliers to obtain the customers' actual perception and predict their overall satisfaction.

8. Conclusions, limitations, and future research directions

8.1. Conclusions

This study uses a sample of 127,629 online reviews to predict customers' overall satisfaction through the technical attributes of online textual reviews and reviewers' identity. We find that certain technical attributes—subjectivity, readability, and length–significantly negatively influence customer ratings, and diversity and sentiment polarity significantly positively influence customer ratings. Customers' review engagement positively influences ratings. The findings of this study illustrate the relationships among the linguistic style of online customer reviews, customers' identity, and overall customer perception and satisfaction.

8.2. Limitations and future research directions

The limitations of this study primarily lie in the following facts. First, the sample only contains reviews for hotels in one city and from a single online review platform. Future researchers can extend this study by collecting more samples for multiple cities from various sources. Second, the technical attributes of online textual reviews can be influenced by the languages the customers use and the customers' cultural background. Examining and comparing the technical attributes of online textual reviews written in different languages and in different cultures can be another extension. Third, the variables of review involvement and hotel ranking can change over time. Future studies should examine these variables dynamically.

In addition, future studies can explore the technical attributes of titles of online reviews or predict customer ratings of other hospitality industries such as restaurants and airlines through online customer reviews. Furthermore, the overall customer ratings examined in this study show overall customer satisfaction. Future studies can also examine the relationship between the technical attributes of textual reviews and customer ratings for specific aspects of hotel products and services, such as room quality, staff performance, and location, and explore different online review behaviors that may exist with respect to each aspect.

References

Anand, P., Holbrook, M.B., Stephens, D., 1988. The formation of affective judgments: the cognitive-affective model versus the independence hypothesis. J. Consum. Res. 15 (3), 386–391.

Anderson, S.W., Baggett, L.S., Widener, S.K., 2009. The impact of service operations failures on customer satisfaction: evidence on how failures and their source affect what matters to customers. Manuf. Service Oper. Manag. 11 (1), 52–69.

Andreassen, T.W., Lindestad, B., 1998. Customer loyalty and complex services: the impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise. Int. J. Service Ind. Manag. 9 (1), 7–23.

Au, N., Law, R., Buhalis, D., 2010. The impact of culture on eComplaints: evidence from Chinese consumers in hospitality organisations. Inf. Commun. Technol. Tourism 2010, 285–296.

Banerjee, S., Chua, A.Y., 2016. In search of patterns among travellers' hotel ratings in TripAdvisor. Tourism Manag. 53, 125–131.

Berezina, K., Bilgihan, A., Cobanoglu, C., Okumus, F., 2016. Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. J. Hospitality Market. Manag. 25 (1), 1–24.

Blal, I., Sturman, M.C., 2014. The differential effects of the quality and quantity of online reviews on hotel room sales. Cornell Hospitality Q. 55 (4), 365–375.

Bradley, G.L., Sparks, B.A., Weber, K., 2015. The stress of anonymous online reviews: a conceptual model and research agenda. Int. J. Contemp. Hospitality Manag. 27 (5), 739–755.

Bronner, F., De Hoog, R., 2011. Vacationers and eWOM: Who posts, and why, where, and what?. J. Travel Res. 50 (1), 15–26.

Cantallops, A.S., Salvi, F., 2014. New consumer behavior: a review of research on eWOM and hotels. Int. J. Hospitality Manag. 30, 41–51.

Casaló, L.V., Flavian, C., Guinaliu, M., Ekinci, Y., 2015. Do online hotel rating schemes influence booking behaviors?. Int. J. Hospitality Manag. 49, 28–36.

Cheung, C.M.K., Thadani, D.R., 2012. The impact of electronic word-of-mouth communication: a literature analysis and integrative model. Decis. Support Syst. Vol. 54 (No.1), 461–470.

Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Marketing Res. 43 (3), 345–354.

Cho, H., Kim, S., Lee, J., Lee, J.S., 2014. Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. Knowl.-Based Syst. 71, 61–71.

Clark, R.A., Goldsmith, R.E., 2005. Market mavens: psychological influences. Psychol. Marketing 22 (4), 289–312.

Connelly, B.L., Certo, S.T., Ireland, R.D., Reutzel, C.R., 2011. Signaling theory: a review and assessment. J. Manag. 37 (1), 39–67.

Dai, H., Luo, X.R., Liao, Q., Cao, M., 2015. Explaining consumer satisfaction of services: the role of innovativeness and emotion in an electronic mediated environment. Decis. Support Syst. 70, 97–106.

Deng, S., Sinha, A.P., Zhao, H., 2017. Adapting sentiment lexicons to domain-specific social media texts. Decis. Support Syst. 94, 65–76.

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 Manag. 52, 498–506.

Filieri, R., Alguezaui, S., McLeay, F., 2015. Why do travelers trust TripAdvisor?: Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism Manag. 51, 174–185.

Forgas, J.P., 1994. The role of emotion in social judgments: an introductory review and an Affect Infusion Model (AIM). Eur. J. Soc. Psychol. 24 (1), 1–24.

Ganu, G., Kakodkar, Y., Marian, A., 2013. Improving the quality of predictions using textual information in online user reviews. Inf. Syst. 38 (1), 1–15.

Gao, S., Tang, O., Wang, H., Yin, P., 2018. Identifying competitors through comparative relation mining of online reviews in the restaurant industry. Int. J. Hospitality Manag. 71, 19–32.

Geetha, M., Singha, P., Sinha, S., 2017. Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Manag. 61, 43–54.

Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., Chatzisavvas, K.C., 2017. Sentiment analysis leveraging emotions and word embeddings. Expert Syst. Appl. 69, 214–224.

Godes, D., Mayzlin, D., 2004. Using online conversations to study word-of-mouth communication. Marketing Sci. 23 (4), 545–560.

Gu, X., Kim, S., 2015. What parts of your apps are loved by users? (T). In: 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE. pp. 760–770.

Gu, B., Ye, Q., 2014. First step in social media: measuring the influence of online management responses on customer satisfaction. Prod. Oper. Manag. 23 (4), 570–582.

Gunning, R., 1969. The fog index after twenty years. J. Bus. Commun. 6 (2), 3–13.

He, W., Tian, X., Tao, R., Zhang, W., Yan, G., Akula, V., 2017. Application of social media analytics: a case of analyzing online hotel reviews. Online Inf. Rev. 41 (7), 921–935.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., Gremler, D.D., 2004. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. J. Interact. Market. 18 (No. 1), 38–52.

Heung, V.C., Lam, T., 2003. Customer complaint behaviour towards hotel restaurant services. Int. J. Contemp. Hospitality Manag. 15 (5), 283–289.

Hu, N., Bose, I., Koh, N.S., Liu, L., 2012. Manipulation of online reviews: an analysis of ratings, readability, and sentiments. Decis. Support Syst. 52 (3), 674–684.

Isen, A.M., 1987. Positive affect, cognitive processes: and social behavior. Adv. Exp. Soc. Psychol. 20, 203–253.

Kim, W.G., Lim, H., Brymer, R.A., 2015. The effectiveness of managing social media on hotel performance. Int. J. Hospitality Manag. 44, 165–171.

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., et al., 2017. Visual genome: connecting language and vision using crowdsourced dense image annotations. Int. J. Comput. Vision 123 (1), 32–73.

Kwok, L., Xie, K.L., 2016. Factors contributing to the helpfulness of online hotel reviews: does manager response play a role?. Int. J. Contemp. Hospitality Manag. 28 (No. 10), 2156–2177.
Lahuerta-Otero, E., Cordero-Gutiérrez, R., 2016. Looking for the perfect tweet: the use of data mining techniques to find influencers on Twitter. Comput. Human Behav. 64, 575–583.
Lewis, B.R., McCann, P., 2004. Service failure and recovery: evidence from the hotel industry. Int. J. Contemp. Hospitality Manag. 16 (1), 6–17.
Li, H., Ye, Q., Law, R., 2013. Determinants of customer satisfaction in the hotel industry: an application of online review analysis. Asia Pac. J. Tourism Res. 18 (7), 784–802.
Li, H., Zhang, Z., Meng, F., Janakiraman, R., 2017. Is peer evaluation of consumer online reviews socially embedded? – An examination combining reviewer's social network and social identity. Int. J. Hospitality Manag. 67, 143–153.
Liu, Z., Park, S., 2015. What makes a useful online review? Implication for travel product websites. Tourism Manag. 47, 140–151.
Liu, X., Schuckert, M., Law, R., 2018. Utilitarianism and knowledge growth during status seeking: evidence from text mining of online reviews. Tourism Manag. 66, 38–46.
Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., 2014. Textblob: simplified text processing. Secondary TextBlob: Simplified Text Processing.
Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E.C., Wetzels, M., Pfann, G., 2013. More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. J. Marketing 77 (1), 87–103.
Ma, Y., Xiang, Z., Du, Q., Fan, W., 2018. Effects of user-provided photos on hotel review helpfulness: an analytical approach with deep leaning. Int. J. Hospitality Manag. 71, 120–131.
Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D., 2014. The Stanford Corenlp Natural Language Processing Toolkit Paper Presented at the ACL (System Demonstrations).
Mano, H., Oliver, R.L., 1993. Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction. J. Consum. Res. 20 (3), 451–466.
McColl-Kennedy, J.R., Sparks, B.A., 2003. Application of fairness theory to service failures and service recovery. J. Service Res. 5 (3), 251–266.
Micu, A., Micu, A.E., Geru, M., Lixandroiu, R.C., 2017. Analyzing user sentiment in social media: implications for online marketing strategy. Psychol. Market. 34 (12), 1094–1100.
Neto, J.Q.F., Bloemhof, J., Corbett, C., 2016. Market prices of remanufactured: used and new items: evidence from eBay. Int. J. Prod. Econ. 171, 371–380.
Oliver, R.L., 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. J. Market. Res. 460–469.
Pan, B., Yang, Y., 2017. Forecasting destination weekly hotel occupancy with big data. J. Travel Res. 56 (7), 957–970.
Qu, Z., Zhang, H., Li, H., 2008. Determinants of online merchant rating: content analysis of consumer comments about Yahoo merchants. Decis. Support Syst. 46 (1), 440–449.
Rose, S., Hair, N., Clark, M., 2011. Online customer experience: a review of the business-to-consumer online purchase context. Int. J. Manag. Rev. 13 (1), 24–39.
Saif, H., He, Y., Fernandez, M., Alani, H., 2016. Contextual semantics for sentiment analysis of Twitter. Inf. Process. Manag. 52 (1), 5–19.
Salehan, M., Kim, D.J., 2016. Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics. Decis. Support Syst. 81, 30–40.
Schoefer, K., Ennew, C., 2005. The impact of perceived justice on consumers' emotional responses to service complaint experiences. J. Services Market. 19 (5), 261–270.
Schuckert, M., Liu, X., Law, R., 2015. A segmentation of online reviews by language groups: how english and non-english speakers rate hotels differently. Int. J. Hospitality Manag. 48, 143–149.
Smith, A.K., Bolton, R.N., 2002. The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments. J. Acad. Market. Sci. 30 (1), 5–23.
Sparks, B.A., Browning, V., 2010. Complaining in cyberspace: the motives and forms of hotel guests' complaints online. J. Hospitality Market. Manag. 19 (7), 797–818.
Sparks, B.A., Browning, V., 2011. The impact of online reviews on hotel booking intentions and perception of trust. Tourism Manag. 32 (6), 1310–1323.
Sparks, B.A., So, K.K.F., Bradley, G.L., 2016. Responding to negative online reviews: the effects of hotel responses on customer inferences of trust and concern. Tourism Manag. 53, 74–85.
Tian, X., He, W., Tao, R., Akula, V., 2016. Mining online hotel reviews: a case study from hotels in China. In: Proceeding of Twenty-second Americas Conference on Information Systems. San Diego, CA, USA.
Verhagen, T., Nauta, A., Feldberg, F., 2013. Negative online word-of-mouth: behavioral indicator or emotional release?. Comput. Human Behav. 29 (4), 1430–1440.
Vermeulen, I.E., Seegers, D., 2009. Tried and tested: the impact of online hotel reviews on consumer consideration. Tourism Manag. 30 (1), 123–127.
Weber, R.P., 1985. Basic Content Analysis. Sage, Beverly Hills, CA.
Westbrook, R.A., Oliver, R.L., 1991. The dimensionality of consumption emotion patterns and consumer satisfaction. J. Consum. Res. 18 (1), 84–91.
Wooldridge, J.M., 2015. Introductory Econometrics: A Modern Approach. Nelson Education.
Wu, L., Shen, H., Fan, A., Mattila, A.S., 2017. The impact of language style on consumers′ reactions to online reviews. Tourism Manag. 59, 590–596.
Xiang, Z., Schwartz, Z., Gerdes, J.H.J., Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction?. Int. J. Hospitality Manag. 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 Manag. 58, 51–65.
Xie, K.L., Zhang, Z., Zhang, Z., 2014. The business value of online consumer reviews and management response to hotel performance. Int. J. Hospitality Manag. 43, 1–12.
Xu, X., Li, Y., 2016. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach. Int. J. Hospitality Manag. 55, 57–69.
Xu, X., Li, Y., Lu, A.C.C., 2017. A comparative study of the determinants of business and leisure travelers satisfaction and dissatisfaction. Int. J. Serv. Oper. Manag.(ahead-of-print).
Yalch, R.F., Spangenberg, E.R., 2000. The effects of music in a retail setting on real and perceived shopping times. J. Bus. Res. 49 (2), 139–147.
Ye, Q., Law, R., Gu, B., 2009. The impact of online user reviews on hotel room sales. Int. J. Hospitality Manag. 28 (1), 180–182.
Yoo, K.H., Gretzel, U., 2011. Influence of personality on travel-related consumergenerated media creation. Comput. Human Behav. 27 (2), 609–621.
Zhang, J.J., Mao, Z., 2012. Image of all hotel scales on travel blogs: its impact on customer loyalty. J. Hospitality Market. Manag. 21 (2), 113–131.
Zhang, Z., Ye, Q., Law, R., Li, Y., 2010. The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. Int. J. Hospitality Manag. 29 (4), 694–700.

D. Zhang, L. Zhou, J.L. Kehoe, I.Y. Kilic, What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews, J. Manag. Inf. Syst. 33 (2) (2016)a, 456–481.
Zhang, X., Yu, Y., Li, H., Lin, Z., 2016. Sentimental interplay between structured and unstructured user-generated contents: an empirical study on online hotel reviews. Online Inf. Rev. 40 (1), 119–145.
Zhou, S., Guo, B., 2017. The order effect on online review helpfulness: a social influence perspective. Decis. Support Syst. 93, 77–87.
Zhou, K.Z., Li, C.B., 2012. How knowledge affects radical innovation: knowledge base, market knowledge acquisition, and internal knowledge sharing. Strateg. Manage. J. 33 (9), 1090–1102.
Zourrig, H., Chebat, J.C., Toffoli, R., 2009. Consumer revenge behavior: a cross-cultural perspective. J. Bus. Res. 62 (10), 995–1001.