International Journal of Information Management 37 (2017) 673–683

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors

Xun Xu[a], Xuequn Wang[b,*], Yibai Li[c], Mohammad Haghighi[d]

[a] Department of Management, Operations, and Marketing, College of Business Administration, California State University, Stanislaus, One University Circle, Turlock, CA, 95382, United States
[b] School of Engineering and Information Technology, Murdoch University, 90 South Street, Perth, WA, 6150, Australia
[c] Operations & Information Management Department, Kania School of Management, The University of Scranton, Brennan Hall 322, Scranton, PA, 18510, United States
[d] Faculty of Management, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:
Customer satisfaction
Customer dissatisfaction
Online textual reviews
Text mining
Regression

ABSTRACT

With the rapid development of information technology, customers not only shop online—they also post reviews on social media. This user-generated content (UGC) can be useful to understand customers' shopping experiences and influence future customers' purchase intentions. Therefore, business intelligence and analytics are increasingly being advocated as a way to analyze customers' UGC in social media and support firms' marketing activities. However, because of its open structure, UGC such as customer reviews can be difficult to analyze, and firms find it challenging to harness UGC. To fill this gap, this study aims to examine customer satisfaction and dissatisfaction toward attributes of hotel products and services based on online customer textual reviews. Using a text mining approach, latent semantic analysis (LSA), we identify the key attributes driving customer satisfaction and dissatisfaction toward hotel products and service attributes. Additionally, using a regression approach, we examine the effects of travel purposes, hotel types, star level, and editor recommendations on customers' perceptions of attributes of hotel products and services. This study bridges customer online textual reviews with customers' perceptions to help business managers better understand customers' needs through UGC.

1. Introduction

With the rapid development of information technology, customers not only shop online—they also post reviews on social media. This information created based on social media platforms is often referred to as user-generated content (UGC). UGC can help reduce the perceived risk of online shopping before customers make purchasing decisions (Ladhari & Michaud, 2015). UGC also provides opportunities for firms to receive customer feedback and improve corresponding attributes of products and services, generating critical business value (Habibi, Laroche, & Richard, 2014; Suo, Sun, Hajli, & Love, 2015; Wang, Hsiao, Yang, & Hajli, 2016; Xie, Zhang, & Zhang, 2014). Therefore, business intelligence and analytics (BI & A) is increasingly advocated to analyze UGC and support firms' marketing activities.

BI & A has become increasingly important for firms' profit and operations. Using effective BI & A, firms can reduce their marketing costs by better understanding customer preferences and implementing market segmentation (Wixom & Watson, 2010). It is predicted that, in the United States alone, there will be a shortage of 140,000 to 190,000 professionals with deep analytical skills by 2018 (Manyika et al., 2011). However, UGC such as online customer textual reviews is often in open form and has no structural restrictions. Therefore, although UGC includes richer information regarding customer purchasing experiences and perceptions (Berezina, Bilgihan, Cobanoglu, & Okumus, 2016) than customer ratings, it is quite challenging for firms to apply BI & A and harness UGC. As a result, the value of UGC is insufficiently discovered and analyzed, and firms find it challenging to generate marketing insights regarding what aspects of products or services make customers feel satisfied or dissatisfied based on UGC.

The numerous customer online reviews posted on social media and online shopping websites have sped up the demand for big data analytics and corresponding techniques. Dealing with unstructured texts is among one of the biggest challenges of big data analytics (Gandomi & Haider, 2015). Sentiment analysis is often used to mine customers' sentiments from their online reviews (Schumaker, Jarmoszko, & Labedz, 2016). Sentiment analysis can detect the implicit expressions of customers' emotions in their texts (Balahur, Hermida, & Montoyo, 2012). To further examine customer online reviews, our study focuses on examining customer satisfaction and dissatisfaction.

* Corresponding author.
E-mail addresses: xxu@csustan.edu (X. Xu), xuequnwang1600@gmail.com (X. Wang), yibai.li@scranton.edu (Y. Li), mhaghighi@ut.ac.ir (M. Haghighi).

http://dx.doi.org/10.1016/j.ijinfomgt.2017.06.004
Received 21 February 2017; Received in revised form 29 April 2017; Accepted 15 June 2017
0268-4012/ © 2017 Elsevier Ltd. All rights reserved.

freepaper.me

Although customer satisfaction and dissatisfaction have been well examined in previous literature, few studies have discussed customer satisfaction and dissatisfaction using customers' online textual reviews (Xiang, Schwartz, Gerdes, & Uysal, 2015). In these few studies, descriptive methods such as frequency analysis and content analysis are most often used (e.g., Li, Ye, & Law, 2013). Therefore, more studies are needed to explore how to better utilize UGC to capture customers' insights and bridge the gap between customer satisfaction/dissatisfaction and UGC.

With the hospitality industry as its context, the objective of this study is to analyze UGC to identify key attributes driving customer satisfaction and dissatisfaction toward hotel products and service attributes and to examine the effect of travel purposes, hotels types, star level, and editor recommendations on customers' perceptions of attributes of hotel products and services offered by hotels with different types. This study has two research questions. First, *what are the key attributes driving customer satisfaction and dissatisfaction, which are reflected in customer reviews posted on online booking websites (booking.com in this study), toward hotel products and service attributes?* To answer this question, our study conducts latent semantic analysis (LSA), a text mining approach, to analyze online customer textual reviews from booking.com. We selected LSA because this technique is helpful in extracting and representing human nature words (Kulkarni, Apte, & Evangelopoulos, 2014). We consider positive online reviews to indicate customer satisfaction and negative online reviews to indicate customer dissatisfaction. Thus, the attributes driving customer satisfaction are mined from positive reviews, and the attributes driving customer dissatisfaction are mined from negative reviews. Second, *what are the effects of travel purposes, hotels types, star level, and editor recommendation on customers' perceptions of attributes of hotel products and services?* To answer this question, our study conducts regression based on textual reviews.

Our study makes important contributions that benefit both researchers and business practitioners. Academically, our study bridges customer satisfaction and dissatisfaction with online customer textual reviews by reflecting customers' perceptions of attributes of products and services. To process the textual data, our study combines text mining and regression methods to extract and represent customer consumption experiences and perceptions. Our study thus contributes to previous BI & A literature by proposing a new approach to capture consumer perceptions. Previous studies mainly used customer ratings to examine their satisfaction with preset questions reflecting certain attributes of products and services (e.g., Schuckert, Liu, & Law, 2015). By using LSA and regression based on customer textual reviews, our study can thoroughly examine customer perceptions of attributes of products and services through detailed consumption experiences. In this way, firms can practically identify relevant product and services attributes leading to customer satisfaction and dissatisfaction as well as how customers' perceptions of those attributes are influenced by firms' market positioning and strategies. Those results can thus provide these firms with a roadmap for improving service quality and firm performance and for better targeting the market by implementing appropriate market positioning and strategies.

The rest of our study is structured as follows. Section 2 reviews the relevant literature; Section 3 discusses the theoretical background and develops the hypotheses; Section 4 describes the research method; Section 5 analyzes the data and presents the results; Section 6 discusses the results; Section 7 provides implications, limitations, and opportunities for future studies; and Section 8 concludes the study.

2. Literature review

2.1. Business intelligence and analytics

BI & A is referred to as "the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand its business and market and make timely business decisions" (Chen, Chiang, & Storey, 2012). BI & A was initially used to focus on structured data stored in commercial relational database management systems. Since the 2000s, Web 2.0-based systems such as social media have generated a large amount of unstructured UGC, resulting in great opportunities and challenges for BI & A.

Recent BI & A literature has begun to understand how to better analyze and harness UGC on social media. For example, Chau and Xu (2012) developed a framework to automatically collect and analyze blog content. Park, Huh, Oh, and Han (2012) proposed a social-network–driven inference framework to determine the accuracy and reliability of customer profiles. He, Zha, and Li (2013) applied text mining to analyze text content on the Facebook and Twitter sites of the pizza chains. More recently, Wang et al. (in press) conducted a content analysis of 26 big data implementation cases in health care and identified five major big data analytics capabilities and potential benefits. Those studies made important progress regarding to how to analyze and harness data on social media.

However, conducting BI & A is still challenging because of the lack of widely adopted methods for effectively analyzing and harnessing data on social media. The value of social media data has thus been insufficiently explored to support marketing activities. Therefore, more BI & A studies are needed to examine how to analyze social media data and capture consumer perception of UGC, as well as understand how BI & A can help create business value. Our study tries to fill this gap by using text mining to analyze social media data. With LSA and regression, our method can bridge customers' online textual reviews and customer perceptions and generate important marketing insights for business practitioners. Our study particularly focuses on customer satisfaction and dissatisfaction toward particular attributes of products and services.

2.2. Customer satisfaction and dissatisfaction

Although customer satisfaction and dissatisfaction have been widely examined in previous hospitality literature (e.g., Gu & Ye, 2014; Matzler & Sauerwein, 2002; Sim, Mak, & Jones, 2006), few studies treat them separately (Zhou, Ye, Pearce, & Wu, 2014), despite the fact that customer satisfaction and dissatisfaction are different constructs (Chowdhary & Prakash, 2005). Specifically, most studies use the overall satisfaction score to measure customer satisfaction and do not differentiate low satisfaction from dissatisfaction. Indeed, according to the two-factor theory (Matzler & Sauerwein, 2002), customer satisfaction and dissatisfaction can coexist. Some factors, such as excitement factors, can generate customer satisfaction at a high level but do not result in customer dissatisfaction at a low level. Other factors, such as basic factors, can generate customer dissatisfaction at a low level but may not result in customer satisfaction at a high level. To better illustrate the different formation mechanisms of customer satisfaction and dissatisfaction, our study views customer satisfaction and dissatisfaction separately and identifies their corresponding influential factors.

In terms of methodologies, most previous studies use surveys to examine customer satisfaction and dissatisfaction. Although surveys can obtain first-hand data, they may not allow researchers to identify all of the product and service attributes that affect customer satisfaction and dissatisfaction. To avoid the limitations of the survey method, our study uses customer online textual reviews, a specific type of UGC, to examine customer satisfaction and dissatisfaction, which have received little attention in previous literature (Xiang et al., 2015; Berezina et al., 2016). Online textual reviews reflect customer satisfaction and dissatisfaction in a more inclusive and comprehensive way because of their open structure, the availability of big data samples, and the anonymity of respondents. Therefore, our study can complement previous literature by examining customer satisfaction and dissatisfaction toward attributes of products and services with online customer textual reviews.

freepaper.me

2.3. Customer online textual reviews

Although previous studies on customer reviews has mainly focused on customer ratings (e.g., Schuckert et al., 2015), recent studies have begun to pay more attention to textual reviews (Xiang et al., 2015). Analyzing textual reviews can provide deeper insights than exploring customer general experiences and overall satisfaction. The content of textual review studies includes customer complaints, experiences, and satisfaction (Lee and Hu, 2005; Xiang et al., 2015) and firms' responses toward the textual reviews (Gu & Ye, 2014). Customer online textual reviews show customer experiences in a more detailed way because of their open structure and can therefore reflect customer perceptions more accurately (Berezina et al., 2016). These reviews thus give rise to the electronic word-of-mouth (eWOM) effect, which influences future online customers' purchasing intentions (Cantallops & Salvi, 2014). Previous literature has mainly used content analysis and frequency analysis to analyze textual reviews (e.g., Li et al., 2013). To better harness customer textual reviews, our study applies LSA and examines positive and negative textual reviews separately to identify specific attributes of hotel products and services leading to customer satisfaction and dissatisfaction.

2.4. Sentiment analysis in big data analytics

With the fast development of big data, big data analytics has become another significant component of BI & A (Chen et al., 2012). According to Russom (2011), "Big data analytics is where advanced analytic techniques operate on big data" (p. 8). Researchers believe that the combination of big data and analytics is the path from insights to value (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Unstructured text, which is the focus in this study, is one of the most challenged formats of data in big data analytics (Gandomi and Haider, 2015). Six main types of big data analysis techniques have been used in previous studies: data mining, web mining, visualization methods, machine learning, optimization methods, and social network analysis (Yaqoob et al., 2016). This study can be clustered into the web mining category; we use LSA to analyze customer online reviews.

One of the most widely adopted BI & A techniques in analyzing big data of UGC is sentiment analysis (Pang & Lee, 2008). Sentiment analysis or, interchangeably, opinion mining, is "the computational study of people's opinions, attitudes and emotions toward an entity" (Medhat, Hassan, & Korashy, 2014, p. 1093). The entity can represent individuals, events, products, or services. In practice, numerous companies have used sentiment analysis to develop marketing strategies by assessing and predicting public attitudes toward their brand (Cambria, Schuller, Xia, & Havasi, 2013).

Sentiment analysis is often used to analyze big data of UGC posted on both social media and online shopping websites. Regarding social media, Schumaker et al. (2016) focused on texts on Twitter using sentiment analysis to predict wins and spread in the premier league. Twitter is also used to extract and evaluate conversational patterns using sentiment analysis (Lipizzi, Iandoli, & Marquez, 2015). Meire, Ballings, and Van den Poel (2016) used sentiment analysis to analyze Facebook texts to explore the added value of auxiliary data such as user profile information. Regarding online booking/shopping websites, Salehan and Kim (2016) collected data from Amazon.com and used sentiment analysis to predict the readership and helpfulness of customer online reviews.

The sentiments mined from online customer reviews include positive and negative sentiments (Wang, Sun, Ma, Xu, & Gu, 2014). Positive sentiments include delight, joy, and satisfaction, and negative sentiments include anger, fear, guilt, sadness, frustration, and dissatisfaction (Balahur et al., 2012). In this study, we further examines customer online reviews posted on hotel booking websites and analyze customers' satisfaction and dissatisfaction toward the various attributes of products and services. Our study contributes to previous literature by connecting customer online reviews with the various attributes of products and services offered by various types of hotels. In this way, the reasons for customers' satisfaction/dissatisfaction are revealed. The business values of customer online reviews are thus significantly reflected because business can improve the corresponding product and service attributes based on the customers' specific reviews regarding those attributes.

3. Theoretical foundation and hypotheses development

3.1. Theoretical foundation

The theoretical foundation of this study is expectation-disconfirmation theory. According to expectation-disconfirmation theory, customers compare their expectations with the perceived quality of products and services before consuming them. When the expectation is no greater than the perceived quality, customers are satisfied; otherwise, customers are dissatisfied (Oliver, 1980).

Customers' expectations before consumption are influenced by many factors, including product- and service-provider–related factors and customer-related factors. In the context of online hotel booking, product- and service-provider–related factors can include hotel stars, editor recommendation, and hotel type. Those factors are referred to by online customers (Srinivasan, Anderson, & Ponnavolu, 2002) and influence customer expectations (Boulding, Kalra, Staelin, & Zeithaml, 1993).

Customer-related factors include customers' areas of focus (Lohan, Lang, & Conboy, 2011). Customers can have different areas of focus regarding the attributes of products and services, depending on their own identities, requirements, information, and relationships with providers. In the context of online hotel booking, customers can book hotels for holidays or business purposes. Because of the different travel purposes, business and leisure travelers can have different areas of focus on the attributes of hotel products and services. Below, we develop our hypotheses based on expectation-disconfirmation theory.

3.2. Hypotheses development

3.2.1. The influence of travel purpose on customer satisfaction and dissatisfaction

Customers travel mainly for business or leisure purposes. Customers have different focuses on hotels' products and services because their identities, requirements, and travel purposes are different. Business and leisure travelers emphasize different attributes of hotel products and services when they select hotels (Yavas & Babakus, 2005), pertaining to their expectations of the performance of hotel attributes (Chu & Choi, 2000). Therefore, business and leisure travelers have different evaluations and perceptions of the perceived quality of hotel products and services (Kashyap & Bojanic, 2000). Therefore, this study proposes the following hypotheses:

Hypothesis 1a. *Travel purpose influences customer satisfaction toward the attributes of hotel products and services.*

Hypothesis 1b. *Travel purpose influences customer dissatisfaction toward the attributes of hotel products and services.*

3.2.2. The influence of hotel star level on customer satisfaction and dissatisfaction

The star level of hotel products and services indicates the quality and variability of the offered products and services. Higher star level hotels have higher room rates. The star level indicates the monetary value of the hotel's product and services and influences customers' expectations (Liu, Law, Rong, Li, & Hall, 2013). Customers expect higher obtained utility when they pay more. Therefore, higher room rates at higher star level hotels positively influence customer

فری پیپر freepaper.me

expectations of the attribute quality of hotel products and services and influence customer satisfaction and dissatisfaction (Anderson, Fornell, & Lehmann, 1994).

Lower star level hotels are typically considered budget hotels, while higher star level hotels are often thought of as luxury hotels. Customers staying in budget hotels often pay more attention to the functional value of the hotel products and services attributes, such as cleanliness, staff, and location (Ren, Qiu, Wang, & Lin, 2016). In comparison, customers staying in luxury hotels emphasize the hedonic value of hotel product and service attributes, such as luxury room amenities and additional services such as recreational services (Heo & Hyun, 2015). Therefore, different star levels of hotels have different perceived qualities among customers, which influences their perceptions. And so, this study proposes the following hypotheses:

Hypothesis 2a. *Hotels' star levels influence customer satisfaction toward the attributes of hotel products and services.*

Hypothesis 2b. *Hotels' star levels influence customer dissatisfaction toward the attributes of hotel products and services.*

3.2.3. *The influence of editor recommendation on customer satisfaction and dissatisfaction*

Many customers book hotels online. Online ratings and reviews from past customers and editors generate an eWOM effect, which influences online customers' booking intentions (Cantallops & Salvi, 2014). The source of these reviews plays an important role in influencing customer expectation of hotels during the booking process (Zeithaml, Bitner, & Gremler, 2006). Because editor recommendations are not anonymous and written by professional people (i.e., well-known online travel entities), they are valued more by online customers compared with anonymous customer online reviews (Casalo, Flavian, Guinaliu, & Ekinci, 2015). Editor recommendations usually contain credible information about hotels and therefore enhance customer expectations and hotels' attractiveness (Zhang, Ye, Law, & Li, 2010). Editor recommendations also indicate hotels' reputation and goodwill, increasing customer trust toward the hotels (Fuller, Serva, & Benamati, 2007). All of these influence customer satisfaction and dissatisfaction toward hotels. Therefore, this study proposes the following hypotheses:

Hypothesis 3a. *Editor recommendations influence customer satisfaction toward the attributes of hotel products and services.*

Hypothesis 3b. *Editor recommendations influence customer dissatisfaction toward the attributes of hotel products and services.*

3.2.4. *The influence of hotel type on customer satisfaction and dissatisfaction*

Customers have different levels of familiarity toward hotels of different types in terms of individual hotels and chain hotels (Kandampully & Suhartanto, 2003). Chain hotels are more familiar to customers because of their brands, which tend to lead to customers' more frequent visits and their loyalty (Kandampully & Suhartanto, 2003). The higher familiarity also enhances customers' predictive expectations of the perceived performance of the hotels and thus influences disconfirmation and customer perception (Tam, 2008).

Chain hotels and individual hotels also have different organizational structures, which influences their operating efficiency (Botti, Briec, & Cliquet, 2009). Compared with individual hotels, chain hotels have more standardized operating policies, which includes rules for daily operations, employee training, and technology applications (Yeung & Law, 2004). Individual hotels have more customized products and services with a higher variation in quality standards, which generate differences in customers' perceived quality (Haktanir & Harris, 2005). Therefore, this study proposes the following hypotheses:

Hypothesis 4a. *Hotel type influences customer satisfaction toward the attributes of hotel products and services.*

Hypothesis 4b. *Hotel type influences customer dissatisfaction toward the attributes of hotel products and services.*

4. Methodology

To achieve the objectives of this study, researchers need to (a) identify the most prominent hotel attributes associated with customer satisfaction and dissatisfaction and (b) to quantify the extent to which a customer is satisfied or dissatisfied with these identified attributes. The traditional methods used to achieve these objectives include gap analysis (Parasuraman, Berry, & Zeithaml, 1991; Teas, 1993), linear regression (Cronin & Taylor, 1994), conjoint analysis (Danaher, 1997), and content analysis (Torres & Kline, 2013). However, these methods rely on self-completion hotel customer satisfaction surveys (Danaher, 1997) or customer letters (Torres & Kline, 2013) to collect data. These methods require researchers to identify the service attributes subjectively.

This study, instead, uses an enormous amount of online textual reviews as its data source. The online review data are then analyzed using LSA and text regression. The following sections describe the details of the data collection and data analysis process.

4.1. *Data collection*

The online customer review data were collected from the world's largest third-party hotel booking website, booking.com (Gössling & Lane, 2015). The website provides hotel room reservations and other travel-related services, and it also allows customers to rate and review the hotels they stay in. Over the years, booking.com has accumulated an enormous amount of UGC, which serves as an ideal data source for this study. Another reason we selected booking.com as the data source for this study is that it verifies customers' input—only customers who have booked rooms on booking.com are allowed to post ratings and reviews, which ensures the authenticity of the UGC.

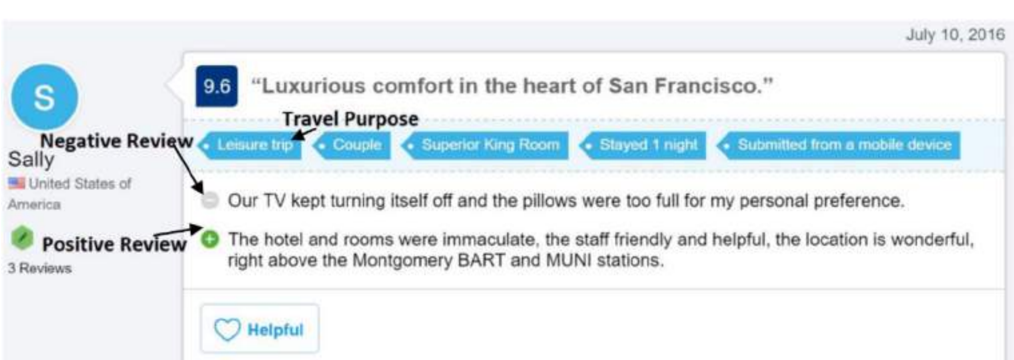

Fig. 1. Screenshot of customer online textual review webpage on Booking.com.

freepaper.me

Fig. 2. Screenshot of hotel booking webpage on Booking.com.

Moreover, we used booking.com as our data source because it requests customers to post their positive and negative reviews simultaneously, but separately, as shown in Fig. 1. This can help us to find factors of customer satisfaction from positive reviews and factors of customer dissatisfaction from negative reviews separately. We also hope that posting both positive and negative reviews can address customers' cognitive bias toward negative reviews.

Customers are asked to post their positive textual comments and negative comments separately, as shown in Fig. 1. Additionally, booking.com also collects information on customers' travel purpose (business trip vs. leisure trip), as shown in Fig. 1, and the hotel's star level, editor recommendation, and hotel type, as shown in Fig. 2.

Referring to Xiang et al. (2015)'s sampling methodology, we collected online customer reviews from hotels in the 100 largest U.S. cities based on the U.S. Census Bureau population estimate (U.S. Census Bureau, 2015). These cities include New York, Los Angeles, Chicago, Houston, and 96 other cities. For each city, we filtered the hotels according to their star levels (star level from 0 to 5). Then, for each star level, we generated two random numbers, a and b, between 1–20 and collected review sample a for hotel b according to the review position and the hotel position listed on booking.com. We repeated this process for six times for each star level of hotels in each city. Reviews with missing values were excluded. Eventually, we collected 3596 data records. The demographics of customers whose reviews were analyzed are shown in Table 1. Table 1 shows the nationality and travel purpose of customers whose reviews are analyzed. Regarding nationalities, we present countries that have 10 or more customers whose reviews were analyzed. As Table 1 shows, our samples are from 20 of these countries, with most samples from the United States. Regarding travel purpose, around four-fifths of the customers were on leisure trips while around one-fifth were on business trips.

4.2. LSA and regression method

Text mining is a significant technique in business intelligence (Gao, Chang, & Han, 2005; McKnight, 2005). Text mining refers to the process of extracting useful, meaningful, and nontrivial information from unstructured text to overcome information overload (Netzer, Feldman, Goldenberg, & Fresko, 2012). Traditional business intelligence has mainly focused on structural data; however, beginning around 2000, more and more researchers have argued that text mining is a powerful extension of business intelligence (Sullivan, 2000). The specific text mining method used in this study is a well-accepted text mining technique called LSA (Li & Joshi, 2012; Lin et al., in press). LSA is an algebraic-statistical method that can detect the underlying topical structure of a document corpus and extract the hidden semantic structures of words and sentences (Evangelopoulos, 2011). LSA, also called latent semantic indexing (LSI), was first proposed by a group of computer scientists at Bell Communication Research, University of Chicago, and University of Western, who developed the method for retrieving information from text. LSA was designed to overcome the limitations of the conventional vector retrieval method or vector space model (Lochbaum and Streeter, 1989). Different from the vector space model, in which text documents are mapped in a multidimensional vector space literally constructed from the vocabulary of text documents, LSA constructs vector space using dimension reduction techniques such as singular value decomposition (SVD) and maps the text documents in the higher-order vector space. This approach partially overcomes the problem of variability in human word choices in the vector space model. Thus, LSA is more appropriate for information retrieval (Dumais, 1992). The higher-order vector space is also a semantic structure, and so LSA allows information retrieval at the semantic level rather than syntax level (Dumais, 1992). In the previous literature, LSA has been applied to enormous areas such as search engines (Berry & Browne, 2005), recommendation systems (Resnick & Varian, 1997), image retrieval (La Cascia, Sethi, & Sclaroff, 1998), speech recognition (Bellegarda, 2000), facial image processing (Hayashi, Yasumoto, Ito, & Koshimizu, 2001), and video retrieval (Liu & Chen, 2007).

Following the well-established text mining procedures discussed in prior studies (Kulkarni et al., 2014), this study applied LSA in three steps: preprocessing, term frequency matrix transformation, and SVD. For preprocessing, we followed the procedure discussed in prior studies (Sidorova, Evangelopoulos, Valacich, & Ramakrishnan, 2008). We tokenized the customer reviews and removed stop words such as "and," "the," "is," and "are." We removed the single character tokens and the tokens that only existed in one review, then applied the stemming method to get to the root of the token. N-gram algorithm was applied to identify phrases (e.g., *comfortable room*, *friendly staff*, or *walking distance*). This step resulted in a two-term frequency matrix, one for positive reviews with more than 1000 tokens and one for negative reviews with more than 1500 tokens.

The second step was term frequency matrix transformation. In this step, we applied the term frequency–inverse document frequency (TF-IDF) weighting method (Husbands, Simon, & Ding, 2001). Term frequency ($tf_{i,j}$) is calculated as $tf_{i,j} = \frac{n_{i,j}}{n_j}$, where n_i is the number of occurrences of token i in document j, and $n_{i,j}$ is the total number of tokens in document j. The IDF is calculated as $idf_i = log\left(\frac{N}{df_i}\right)$, where N is the

Table 1
Customer demographic information.

Nationality	Number	Percentage	Nationality	Number	Percentage
United States	2808	78.09%	Brazil	14	0.39%
UK	134	3.73%	Switzerland	14	0.39%
Canada	113	3.14%	Ireland	13	0.36%
Australia	77	2.14%	Hong Kong	12	0.33%
Germany	32	0.89%	China	11	0.31%
Netherlands	23	0.64%	France	11	0.31%
Saudi Arabia	23	0.64%	UAE	10	0.28%
Italy	20	0.56%	Other	198	5.51%
New Zealand	20	0.56%			
Japan	18	0.50%	**Travel Purpose**	**Number**	**Percentage**
Mexico	15	0.42%	Leisure	2876	79.98%
Nigeria	15	0.42%	Business	720	20.02%
Spain	15	0.42%			
			Total Samples	3596	100%

freepaper.me

number of documents in the database and df_i is the frequency of documents that have the token i. Finally, the weight of each token in each review is calculated as $w_{i,j}$, where $w_{i,j} = tf_{i,j} \times idf_i$. This weighting method is widely adopted in text mining studies and LSA studies (Sidorova et al., 2008).

The third step, SVD, was the dimension reduction algorithm. SVD decomposes the TF-IDF weighted matrix, denoted as matrix A, into the production of three matrices: an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V. That is, $A_{mn} = U_{mm}S_{mn}V_{nn}^T$, where $U^TU = I$ and $V^TV = I$. The columns of V are orthonormal eigenvectors of A^TA; the columns of U are orthonormal eigenvectors of AA^T; and S is a diagonal $m \times n$ matrix containing the square roots of eigenvalues from U or V in descending order. The computation of SVD is discussed in the prior literature, such as in Golub and Reinsch (1970), Klema and Laub (1980), and Baker (2005).

LSA generates three matrices: a document-by-term matrix, a term-by-factor matrix, and a singular value matrix. The singular values in the singular value matrix indicate the significance of the factors—the significance of the identified hotel attributes. The specific hotel attributes for each factor require researchers to look into the document-by-term matrix and the term-by-factor matrix. In accordance with prior studies (Evangelopoulos, 2011), we associated each factor with its high-loading terms and documents in the document-by-term matrix and the term-by-factor matrix to assist this process.

Regression was then applied to conduct the remaining part of the data analysis. Prior studies showed that dimension reduction techniques such as the LSA in this study significantly enhance text regression performance (Ngo-Ye & Sinha, 2012). After finding the factors of customers' online positive and negative textual evaluation, we conducted regression using vector space (Ngo-Ye & Sinha, 2014). The independent variables were the four dummy variables showing travelers' purposes, hotel star, editor recommendation, and hotel type. Business travelers were coded as 1, and leisure travelers were coded as 0. Lower star level hotels (those with 1, 2, and 3 stars) were coded as 0, and higher-star level hotels (4 and 5 stars) were coded as 1. Previous studies found that hotels with four stars or above have different characteristics in pricing, hedonic function, size, distance to attractions, and availability of parking spaces compared with hotels with three stars or fewer (Espinet, Saez, Coenders, & Fluvià, 2003; Heo & Hyun, 2015). The hotels with four or more stars are often considered as luxury hotels, while the hotels with three or less stars are usually considered as economy hotels (Huang, Wang, & Wang, 2015). This is why we coded hotels with three or fewer stars in one category and hotels with four or more stars in the other category.

Editor-recommended hotels are coded as 1; otherwise, they were coded as 0. Chain hotels are coded as 1, and individual hotels are coded as 0. Each regression has one dependent variable: the coordinate of each review vector space on each positive/negative factor. A higher loading shows that the corresponding customer's online review has a higher relevance (e.g., there is more detail about one particular attribute compared with other attributes) for the corresponding attribute (shown by the factor). The methodology of text mining and regression is summarized in Fig. 3.

5. Results

5.1. Factors of customers' positive evaluations

The factors of customers' positive evaluations of hotels are identified in the LSA, as shown in Table 2. The LSA results indicate that these top factors cover over 95% of all unique terms and reviews. For each factor, we selected the top 10 terms as the high-loading terms from more than 4000 terms contained in each of the factors. A singular value reflects the extent of the factor's explanation of the variance (Baker, 2005). A higher singular value indicates that the corresponding factor is more significant when measured by more words and phrases in customer online reviews that reflect the corresponding factor. Four positive factors are identified: friendly staff, comfortable room, good location, and good value.

5.2. Factors of customers' negative evaluations

As above, the factors of customers' negative evaluations toward hotels are identified in the LSA, as shown in Table 3. Five negative factors are identified: low value, uncomfortable and dirty rooms, unfriendly and inefficient staff, amenities and facility issues, and operations issues.

5.3. Results of regressions

In the nine regressions, the independent variables are the four dummy variables showing travelers' purpose, hotel stars, editor recommendations, and hotel type; the dependent variable is the corresponding positive/negative factor. The regression model is shown in Formula (1):

$$f\ (Attributes^*) = \alpha_0 + \alpha_1(Purpose) + \alpha_2(Star) + \alpha_3(Recommendation) + \alpha_4(Type)^{**} \tag{1}$$

*Nine attributes in total, which include four positive attributes (friendly staff, comfortable room, good location, good values) and five negative attributes (low value, uncomfortable and dirty room, unfriendly and inefficient staff, amenities and facility issues, operations issues).

*Coding: leisure travelers = 0, business travelers = 1; low star level hotels = 0, high star level hotels = 1; editor nonrecommended hotels = 0, editor recommended hotels = 1; individual hotels = 0, chain hotels = 1.

Variance inflation factor (VIF) of all independent variables is smaller than 3, which indicates that multicollinearity is probably not an issue (Neto, Bloemhof, & Corbett, 2016). The regressions results are presented in Tables 4 and 5. We find that hotel star level is a significantly influential factor for customers' satisfaction and dissatisfaction toward all specific attributes. Editor recommendation is negatively related to customer dissatisfaction toward operation issues. Hotel type has negative effects on customer dissatisfaction toward unfriendly and inefficient staff. The rest of the relationships between dummy variables and customers' perceptions of specific attributes are not significant. Comparatively, more influential factors are found for customer dissatisfaction toward specific attributes than customer satisfaction.

6. Discussion

This study aims to identify customer satisfaction and dissatisfaction toward attributes of hotel products and services as well as effects of travel purposes, hotel types, star level, and editor recommendation on customers' perceptions of attributes of hotel products and services by analyzing online customer textual reviews. Using text mining and regression, our study shows that four factors (i.e., friendly staff, comfortable room, good location, and good value) significantly influence customers' satisfaction, and five factors (i.e., low value, uncomfortable and dirty room, unfriendly and inefficient staff, amenities and facility issues, and operations issues) significantly influence customers' dissatisfaction. Further, hotel star level significantly influences customer satisfaction and dissatisfaction toward all attributes. Editor recommendation is negatively related to customer dissatisfaction toward operation issues. The hotel type has a negative effect on customer dissatisfaction toward unfriendly and inefficient staff.

freepaper.me

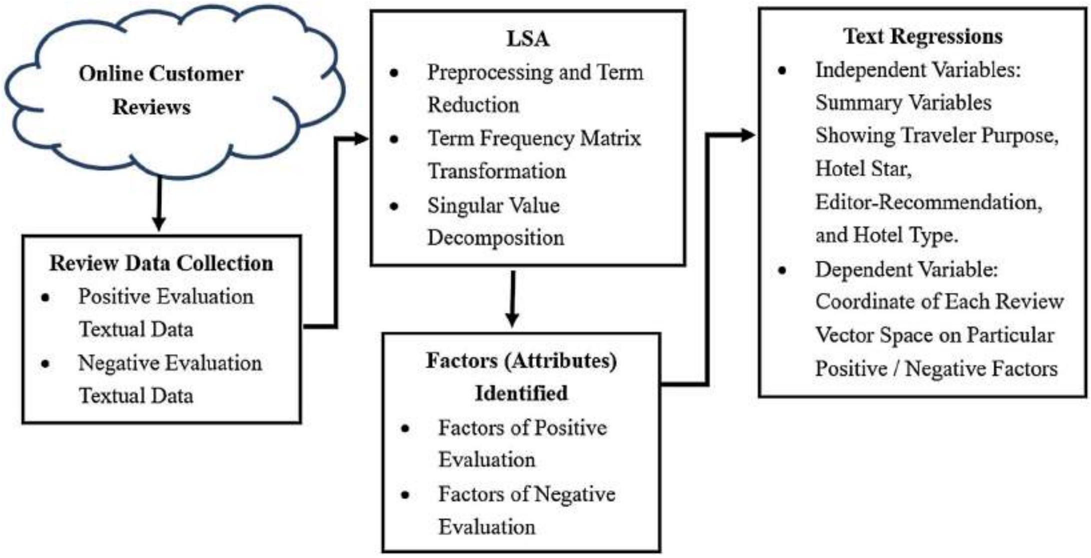

Fig. 3. Analyzing the influential factors of customer online reviews' relevance to particular attributes using text mining and regressions.

6.1. The influence of travel purposes on customer satisfaction and dissatisfaction

Our results do not support hypotheses 1a and 1b. Table 4 shows that travel purpose does not significantly influence customer satisfaction toward the attributes of staff, room, location, and value. Table 5 indicates that travel purpose does not significantly influence customer dissatisfaction toward the attributes of value, room, staff, amenities and facility, and operations issues. A possible reason for this finding is that although business and leisure travelers have different requirements (Yavas & Babakus, 2005) and focus on different attributes of hotel products and services, each attribute has a combined value that balances both types of travelers' requirements. For example, hotels often offer various types of guest rooms. Some rooms with additional functional amenities and facilities are targeted to business travelers, while other rooms with city or ocean views are targeted to leisure travelers. Thus, although business and leisure travelers have different focuses on each attribute, their satisfaction and dissatisfaction toward this particular attribute is similar.

6.2. The influence of hotel's star level on customer satisfaction and dissatisfaction

Our results fully support hypotheses 2a and 2b. A hotel's star level significantly influences customer satisfaction and dissatisfaction toward every attribute of hotel product and services. Results from Table 4 indicate that customers staying in higher star level hotels have lower satisfaction toward each attribute of hotel products and services. Therefore, although higher star level hotels typically offer higher-quality products and services, customers also have higher expectation toward their products and services because of the higher costs they incur. Such paid higher cost influences the expectation-disconfirmation relationship and reduces customer satisfaction toward higher star level hotels.

Regarding customer dissatisfaction, as shown in Table 5, customers staying in luxury hotels are more dissatisfied toward the associated value, consistent with the findings from the previous literature (Gumasta, Kumar Gupta, Benyoucef, & Tiwari, 2011). This shows that the perceived extra utility of higher star level hotels products and services may not match the higher costs incurred. Regarding the attributes of room, staff, amenities and facility, and operations issues, customers staying in lower star level hotels are more dissatisfied toward them. The reason is that lower star level hotels typically offer products and services with lower quality, less variety, and less professionalism and therefore incur customer dissatisfaction and complaints.

6.3. The influence of editor recommendation on customer satisfaction and dissatisfaction

Our results do not support hypothesis 3a and only partially support hypothesis 3b. Customers staying in editor-recommended hotels and non–editor-recommended hotels have similar satisfaction levels regarding most attributes of hotel products and services. Although editor-recommended hotels typically have higher-quality products and services, customers have higher expectations based on the additional information provided by the editor recommendation. Therefore, the comparison between perceived quality and expectation is similar to non–editor-recommended hotels. Similar logic can be applied to customer dissatisfaction toward the attributes of hotel products and services, with one exception: customers staying in editor-recommended hotels have less dissatisfaction toward operations issues. One possible explanation is that operations performance is one of the most critical criteria to be evaluated during the editor recommendation process. Operations performance is easier to improve compared with other attributes with higher costs such as physical settings of guest rooms. Therefore, editor-recommended hotels typically have fewer operations

Table 2
Factors of customer satisfaction.

Factors	Interpretations (Labels)	Singular Values	High-Loading Terms
Factor 1	Friendly Staff	3.355	help_staff, friendli_help_staff, staff_great, staff, staff_excel, friendli_staff, desk_staff, great_help, front_desk_staff, extrem_help
Factor 2	Comfortable Room	3.130	bed_comfi, room_clean_comfort, bed_pillow, nice_clean, bed_pillow_comfort, room_spaciou, bedroom, comfort_clean, clean_bed, room_great
Factor 3	Good Location	3.065	conveni_locat, river, beach, locat_airport, attract, locat_citi, like_conveni, locat_highwai, good_locat, excel_locat
Factor 4	Good Value	2.924	reason_price, good_price, good_valu, price_good, monei, rate, valu_monei, good_valu_monei, price_reason, reason_rate

freepaper.me

Table 3
Factors of customer dissatisfaction.

Factors	Interpretations (Labels)	Singular Values	High-Loading Terms
Factor 1	Low Value	3.690	expens, expens_park, breakfast_expens, bit_expens, room_expens, room_price, expens_room, resort_fee, monei, room_rate
Factor 2	Uncomfortable and Dirty Room	3.427	uncomfort, bed_uncomfort, sleep, sheet, room_dirti, pillow_soft, bed_soft, mattress, room_uncomfort, bed_old
Factor 3	Unfriendly and Inefficient Staff	3.314	rude, staff_rude, person, desk_clerk, front_desk_clerk, clerk, unfriendli, staff_unfriendli, manag_rude, rude_unhelp
Factor 4	Amenities and Facility Issues	3.207	Wi_fi, internet, lobbi, tv, updat. outdat, air_condition, facil, hotel_bit_date, renov
Factor 5	Operations Issues	3.183	smoke, smell_smoke, room_smell, reserv, shower_terribl, dark, loud, wait_park, noisi_pool, alarm_kept

and services failures compared with non-recommended hotels, resulting in less customer dissatisfaction (Lee, Sing, & Chan, 2011).

6.4. *The influence of hotel type on customer satisfaction and dissatisfaction*

Our results do not support hypothesis 4a and only partially support hypothesis 4b. Customers staying in chain hotels and individual hotels have similar satisfaction toward each attribute of hotel products and services. One possible reason is that many branded hotel chains franchise previously individual hotels to use their brands, which reduces the differences in product and service quality between individual hotels and chain hotels. Similar logic can be applied to customer dissatisfaction toward the attributes of hotel products and services with the only exception being that customers staying in chain hotels have less dissatisfaction toward unfriendly and inefficient staff. One possibility is that chain hotels have standards regarding policies and implementation of employee training (Yeung & Law, 2004). However, for individual hotels, because of higher variation services and less employee training, customers may feel more dissatisfied with certain staffs' attitude and behaviors.

7. Implications, limitations, and opportunities for future studies

7.1. *Theoretical implications*

Our study makes three important theoretical contributions. First, by analyzing online customer textual reviews with LSA, our study provides better understanding regarding customer satisfaction and dissatisfaction toward attributes of hotel products and services. Instead of only focusing on overall customer satisfaction, our study examines customer satisfaction and dissatisfaction separately. Using LSA, our study identifies specific attributes contributing to customer satisfaction and dissatisfaction. Specifically, our analyses show that friendly staff, comfortable room, good location, and good value significantly influence customers' satisfaction, while low value, uncomfortable and dirty room, unfriendly and inefficient staff, amenities and facility issues, and operations issues significantly affect customers' dissatisfaction. Thus, with LSA, our study is able to identify relevant product and service attributes affecting customer satisfaction and dissatisfaction in the context of hospitality and avoid missing important attributes. Thus, our study bridges customer online textual reviews with their perceptions. These results complement the previous literature on customer satisfaction, and future studies can further validate our results through surveys.

Second, our study applies regression to examine how various hotel and customer factors influence customer satisfaction and dissatisfaction toward attributes identified through LSA. Based on expectation-disconfirmation theory, we argue that customers probably feel satisfied when the perceived quality meets their expectations; otherwise, they may feel dissatisfied. Here customers' expectations are influenced by service-provider and customer-related factors. In the context of hospitality, we identify four relevant factors: hotel stars, editor recommendation, hotel type, and purpose. Our results show that hotel stars have significant effects on all attributes related to customer satisfaction and dissatisfaction, while editor recommendation and hotel type have effects on certain attributes related to customer dissatisfaction. These results clarify the role of service-provider and customer-related factors on customer satisfaction and dissatisfaction and provide further insights regarding the formation mechanisms of customer satisfaction and dissatisfaction.

Third, in this study, we find that star level is the most important factor influencing both customer satisfaction and dissatisfaction toward various attributes of products and services. This extends the three-factor theory (Füller & Matzler, 2008) that incorporates basic factors, excitement factors, and performance factors. Basic factors only influence customer dissatisfaction, and excitement factors only influence customer satisfaction; performance factors, however, affect both customer satisfaction and dissatisfaction. The good performance of these factors generates customer satisfaction, while poor performance stirs customer dissatisfaction. The three-factor theory in previous studies has been used to examine the influential factors of customer satisfaction and dissatisfaction from the product and service attributes perspective, but this study examines influential factors from the hotel perspective. We find that hotels' star levels are a performance factor in hotel properties, and this factor influences both customer satisfaction and dissatisfaction. Our results show that the star level is negatively related to both customer satisfaction toward comfortable room and customer dissatisfaction toward uncomfortable and dirty room. One possible reason is higher star level hotels have higher standards and implement more efforts and costs in improving various product and service attributes, which lowers customer dissatisfaction. However, the higher star level creates higher expectation, and the higher charges arouse the customers' need to receive more utilities to facilitate equal exchange. This stimulates the possibilities of lowering customer satisfaction toward various product and service attributes.

Table 4
Regressions of positive attributes.

		Dependent Variable			
		Friendly Staff	Comfortable Room	Good Location	Good Value
Independent Variables	Coefficient Parameter	Coefficient Estimates	Coefficient Estimates	Coefficient Estimates	Coefficient Estimates
(Intercept)	α_0	0.008***	0.008***	0.002***	0.004***
Purpose	α_1	−0.004	0.004	0.002	0.001
Star	α_2	−0.034**	−0.035**	−0.036**	−0.034**
Recommendation	α_3	0.003	−0.001	0.006	0.016
Type	α_4	−0.024	−0.016	−0.004	−0.005

freepaper.me

Table 5
Regressions of negative attributes.

			Dependent Variable			
		Low Value	Uncomfortable and Dirty Room	Unfriendly and Inefficient Staff	Amenities and Facility Issues	Operations Issues
Independent Variables	Coefficient Parameter	Coefficient Estimates	Coefficient Estimates	Coefficient Estimates	Coefficient Estimates	Coefficient Estimates
(Intercept)	α_0	0.004***	0.006***	0.004***	0.004***	0.003***
Purpose	α_1	−0.011	−0.018	−0.003	0.01	−0.005
Star	α_2	0.081***	−0.048***	−0.032*	−0.036**	−0.032*
Recommendation	α_3	0.003	−0.016	−0.025	0.014	−0.029*
Type	α_4	−0.016	−0.012	−0.030*	−0.012	0.007

The methodologies applied in this study—LSA and regression—provide efficient approaches for future research on online customer textual reviews. LSA and regression methodology can efficiently extract and represent customers' perceptions from their reviews and thus can be used to examine the perceived quality of products and services.

Our study contributes to BI & A literature by providing an approach for future studies to examine customer perceptions from textual reviews. Our study also enriches existing customer satisfaction studies by providing a new perspective through our data source, research methodologies, and discussion of factors influencing customer satisfaction and dissatisfaction.

7.2. Managerial implications

Our study also has important managerial implications for business practitioners. Online textual reviews have business values that are reflected in their generated eWOM and influence customer purchase intention and demand (Xie et al., 2014). Online textual reviews can provide a way for businesses to understand customer needs and improve their products and services. Compared with customer ratings, online textual reviews can show more details about customers' consumption experiences and customer perceptions because of their open structure. Thus, managers can obtain more insights regarding customers' expectations and needs and their perceived quality of product and services.

This study provides a way for business managers to use online textual reviews. The different factors of customer satisfaction and dissatisfaction identified by LSA help managers understand the different generation mechanisms of customer satisfaction and dissatisfaction and therefore to set up priorities to enhance customer overall satisfaction and implement service recovery actions to alleviate customer dissatisfaction.

When hotels try to lower customer dissatisfaction and avoid negative reviews posted by customers, they can focus on the five factors shown in Table 3. Similarly, when hotels aim to enhance customer satisfaction and increase positive reviews, they can focus on the factors shown in Table 2. The singular values shown in Tables 2 and 3 indicate the importance of each factor. Given the limited resources of businesses, hotels can use our results to set priorities for each attribute and create a roadmap to increase their performance. In addition to online textual reviews and survey e-mails, feedback cards can also be combined to identify specific areas that need further improvement.

Businesses can also utilize the eWOM generated by customer online reviews (Cantallops & Salvi, 2014). Businesses can advocate the advantages of their products and services mentioned in customer online reviews and use the favorable online textual reviews as successful cases for marketing. Managers can also examine customer online textual reviews using our method in different time periods and thus find ways to improve each product and service attribute and how customers' perceptions change over time. This can also provide feedback of businesses' performance improvements and guidelines for future improvements.

Further, results from our regression analysis can help managers better understand the key factors influencing customer satisfaction and dissatisfaction. Different hotels, depending on the star level, type, and other factors, target different customers. Hotels can implement market segmentation strategies to identify different needs from customers with different travel purposes and demographics. In this way, they can better serve customers. Our results show that star level has significant effects on all attributes from customer satisfaction and dissatisfaction. Therefore, hotels need to improve their star levels for their products and services such that customers feel more satisfied toward the attributes of these products and services. Moreover, we find that customers' dissatisfaction toward operations issues is significantly higher toward non–editor-recommended hotels compared with editor-recommended hotels, and customers' dissatisfaction toward the behaviors and attitudes of staff is significantly higher toward individual hotels compared with chain hotels. Therefore, businesses might need to enhance certain criteria for its operations and employee training and learn lessons from benchmarking companies, such as chain hotels that have stronger reputations and brand effects. Improving product and service attributes to obtain accreditation and certification (e.g., a third-party recommendation) can be an approach to alleviate customer dissatisfaction efficiently.

7.3. Limitations and opportunities for future studies

Our study only collected data from one source. Although booking.com is the world's largest third-party hotel booking website and provides sufficient customer reviews, our results may be limited and should thus be interpreted cautiously. Additionally, this study does not distinguish between environmental variables such as customer nationality or hotel region. This study incorporated customers from various countries and hotels in various cities in the same model with no differentiation. A comparative study that examines customers' nationalities and hotels in different regions would be particularly useful.

Additionally, future studies can extend our study through the following ways. First, customer overall satisfaction from online customer reviews can be studied. A comparison between customer overall satisfaction found in customer ratings and customer textual reviews can be conducted. Second, this study focuses on mining customer satisfaction and dissatisfaction toward various products and service attributes from customer online reviews. It is possible that customer online reviews cannot reflect all factors (i.e., products and service attributes) that influence customer satisfaction and dissatisfaction. Future studies can also examine the determinants of customer satisfaction and dissatisfaction from other sources such as customer comments cards, interviews with customers, surveys, and so on. Third, other types of customer perception such as customer loyalty, customer familiarity, and customer willingness to pay can also be examined through online customer reviews. More product and service attributes can also be

فری پیپر freepaper.me

gleaned from customer online reviews when focusing on different types of products and service providers. Finally, a comparison can be conducted between customer perceptions found in online reviews and offline reviews (e.g., customer comment cards). For online reviews, different review sources such as hotel booking websites or various social media websites may also influence customers' perceptions. Therefore, a study of different review sources of customer perceptions is another possibility that should be considered.

8. Conclusion

Our study identified the key attributes driving customer satisfaction and dissatisfaction toward hotel products and services and examined effects of travel purposes, hotel types, star level, and editor recommendation on customers' perceptions of attributes of hotel products and services. We found four key attributes that drive customer satisfaction (i.e., staff, room, location, and value), and five key attributes that affect customer dissatisfaction (i.e., value, room, staff, amenities and facility, and operations issues). Additionally, our results showed that customer satisfaction and dissatisfaction toward attributes of products and services are significantly different among hotels with different star levels. Our study contributed to the BI & A literature by providing a text mining and regression approach to analyze online customer textual reviews, thus providing insights for business managers to use online customer textual reviews and examine customer perceptions.

References

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66.

Baker, K. (2005). *Singular value decomposition tutorial*. [https://www.ling.ohio-state.edu/~kbaker/pubs/Singular_Value_Decomposition_Tutorial.pdf (Accessed 3 February 2016).].

Balahur, A., Hermida, J. M., & Montoyo, A. (2012). Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, *53*(4), 742–753.

Bellegarda, J. R. (2000). Large vocabulary speech recognition with multispan statistical language models. *IEEE Transactions on Speech and Audio Processing*, *8*(1), 76–84.

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 and Management*, *25*(1), 1–24.

Berry, M. W., & Browne, M. (2005). *Understanding search engines: mathematical modeling and text retrieval*. Society for industrial and applied mathematics.

Botti, L., Briec, W., & Cliquet, G. (2009). Plural forms versus franchise and company-owned systems: A DEA approach of hotel chain performance. *Omega*, *37*(3), 566–578.

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, *30*, 7–27.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15–21.

Cantallops, A. S., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, *30*, 41–51.

Casalo, L. V., Flavian, C., Guinaliu, M., & Ekinci, Y. (2015). Do online hotel rating schemes influence booking behaviors? *International Journal of Hospitality Management*, *49*, 28–36.

Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quarterly*, *36*(4), 1189–1216.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chowdhary, N., & Prakash, M. (2005). Service quality: Revisiting the two factors theory? *Journal of Services Research*, *5*(1), 61–75.

Chu, R. K. S., & Choi, T. (2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: A comparison of business and leisure travelers. *Tourism Management*, *21*, 363–377.

Cronin, J. J., Jr, & Taylor, S. A. (1994). SERVPERF versus SERVQUAL: Reconciling performance-based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing*, *58*(1), 125–131.

Danaher, P. J. (1997). Using conjoint analysis to determine the relative importance of service attributes measured in customer satisfaction surveys. *Journal of Retailing*, *73*(2), 235–260.

Dumais, S. (1992). *Enhancing performance in latent semantic indexing (LSI) retrieval*. [http://www2.denizyuret.com/ref/dumais/Enhancing_LSI__Dumais_1991.pdf. Accessed on 15 February 2017.].

Espinet, J. M., Saez, M., Coenders, G., & Fluvià, M. (2003). Effect on prices of the attributes of holiday hotels: A hedonic prices approach? *Tourism Economics*, *9*(2), 165–177.

Evangelopoulos, N. (2011). Citing Taylor: Tracing Taylorism's technical and sociotechnical duality through Latent Semantic Analysis. *Journal of Business and Management*, *17*(1), 57–74.

Füller, J., & Matzler, K. (2008). Customer delight and market segmentation: An application of the three-factor theory of customer satisfaction on life style groups. *Tourism Management*, *29*(1), 116–126.

Fuller, M. A., Serva, M. A., & Benamati, J. (2007). Seeing is believing: The transitory influence of reputation information on e-commerce trust and decision making. *Decision Sciences*, *38*(4), 675–699.

Gössling, S., & Lane, B. (2015). Rural tourism and the development of Internet-based accommodation booking platforms: A study in the advantages, dangers and implications of innovation. *Journal of Sustainable Tourism*, *23*(8–9), 1386–1403.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Gao, L., Chang, E., & Han, S. (2005). Powerful tool to expand business intelligence: Text mining. *Proceedings of world academy of science, engineering and technology*, *8*, 110–115.

Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische Mathematik*, *14*(5), 403–420.

Gu, B., & Ye, Q. (2014). First step in social media: Measuring the influence of online management responses on customer satisfaction. *Production and Operations Management*, *23*(4), 570–582.

Gumasta, K., Kumar Gupta, S., Benyoucef, L., & Tiwari, M. K. (2011). Developing a reconfigurability index using multi-attribute utility theory. *International Journal of Production Research*, *49*(6), 1669–1683.

Habibi, M. R., Laroche, M., & Richard, M. O. (2014). Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities. *International Journal of Information Management*, *34*(2), 123–132.

Haktanir, M., & Harris, P. (2005). Performance measurement practice in an independent hotel context. *International Journal of Contemporary Hospitality Management*, *17*(1), 39–50.

Hayashi, J., Yasumoto, M., Ito, H., & Koshimizu, H. (2001). Method for estimating and modeling age and gender using facial image processing. *Proceedings of seventh international conference on virtual systems and multimedia*, 439–448.

He, W., Zha, S., & 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.

Heo, C. Y., & Hyun, S. S. (2015). Do luxury room amenities affect guests' willingness to pay? *International Journal of Hospitality Management*, *46*, 161–168.

Huang, K. T., Wang, J. C., & Wang, Y. C. (2015). Analysis and benchmarking of greenhouse gas emissions of luxury hotels. *International Journal of Hospitality Management*, *51*, 56–66.

Husbands, P., Simon, H., & Ding, C. H. Q. (2001). On the use of the singular value decomposition for text retrieval. *Computational Information Retrieval*, *5*, 145–156.

Kandampully, J., & Suhartanto, D. (2003). The role of customer satisfaction and image in gaining customer loyalty in the hotel industry? *Journal of Hospitality & Leisure Marketing*, *10*(1–2), 3–25.

Kashyap, R., & Bojanic, D. C. (2000). A structural analysis of value, quality, and price perceptions of business and leisure travelers. *Journal of Travel Research*, *39*, 45–51.

Klema, V., & Laub, A. (1980). The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control*, *25*(2), 164–176.

Kulkarni, S. S., Apte, U. M., & Evangelopoulos, N. E. (2014). The use of Latent Semantic Analysis in operations management research. *Decision Sciences*, *45*(5), 971–994.

La Cascia, M., Sethi, S., & Sclaroff, S. (1998). Combining textual and visual cues for content-based image retrieval on the world wide web. *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, 24–28.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2), 21.

Ladhari, R., & Michaud, M. (2015). EWOM effects on hotel booking intentions, attitudes, trust, and website perceptions. *International Journal of Hospitality Management*, *46*, 36–45.

Lee, C. C., & Hu, C. (2005). Analyzing hotel customers' E-complaints from an internet complaint forum. *Journal of Travel & Tourism Marketing*, *17*(2–3), 167–181.

Lee, M. J., Sing, N., & Chan, E. S. W. (2011). Service failures and recovery actions in the hotel industry: A text-mining approach. *Journal of Vacation Marketing*, *17*(3), 197–207.

Li, Y., & Joshi, K. D. (2012). The state of social computing research: A literature review and synthesis using the latent semantic analysis approach. *Proceedings of Americas Conference on Information Systems*.

Li, H., Ye, Q., & Law, R. (2013). Determinants of customer satisfaction in the hotel industry: An application of online review analysis. *Asia Pacific Journal of Tourism Research*, *18*(7), 784–802.

Lin, X., Li, Y., & Wang, X. (in press). Social commerce research: Definition, research themes and the trends. International Journal of Information Management.

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. *International Journal of Information Management*, *35*(4), 490–503.

Liu, D., & Chen, T. (2007). A topic-motion model for unsupervised video object discovery. *Proceedings of 2007 IEEE conference on computer vision and pattern recognition*, 1–8.

Liu, S., Law, R., Rong, J., Li, G., & Hall, J. (2013). Analyzing changes in hotel customers' expectations by trip mode. *International Journal of Hospitality Management*, *34*, 359–371.

Lochbaum, K. E., & Streeter, L. A. (1989). Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing & Management*, *25*(6), 665–676.

freepaper.me

Lohan, G., Lang, M., & Conboy, K. (2011). *Having a customer focus in agile software development. Information systems development*, . New York: Springer, 441–453.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute [http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation; Accessed 4 February 2017].
Matzler, K., & Sauerwein, E. (2002). The factor structure of customer satisfaction: An empirical test of the importance grid and the penalty-reward-contrast analysis. *International Journal of Service Industry Management*, *13*(4), 314–332.
McKnight, W. (2005). Text data mining in business intelligence. *DM Review*, *15*(1), 80.
Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.
Meire, M., Ballings, M., & Van den Poel, D. (2016). The added value of auxiliary data in sentiment analysis of Facebook posts. *Decision Support Systems*, *89*, 98–112.
Neto, J. Q. F., Bloemhof, J., & Corbett, C. (2016). Market prices of remanufactured,used and new items: Evidence from eBay. *International Journal of Production Economics*, *171*, 371–380.
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*(3), 521–543.
Ngo-Ye, T. L., & Sinha, A. P. (2012). Analyzing online review helpfulness using a regressional reliefF-enhanced text mining method. *ACM Transactions on Management Information Systems*, *3*(2), 1–20.
Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, *61*, 47–58.
Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*, 460–469.
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.
Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing*, *67*(4), 420–450.
Park, S., Huh, S., Oh, W., & Han, S. P. (2012). A social network-based inference model for validating customer profile data. *MIS Quarterly*, *36*(4), 1217–1237.
Ren, L., Qiu, H., Wang, P., & Lin, P. M. C. (2016). Exploring customer experience with budget hotels: Dimensionality and satisfaction. *International Journal of Hospitality Management*, *52*, 13–23.
Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58.
Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, *19*, 40.
Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30–40.
Schuckert, M., Liu, X., & Law, R. (2015). A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *International Journal of Hospitality Management*, *48*, 143–149.
Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84.
Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline? *MIS Quarterly*, *32*(3), 467–482.
Sim, J., Mak, B., & Jones, D. (2006). A model of customer satisfaction and retention for hotels? *Journal of Quality Assurance in Hospitality & Tourism*, *7*(3), 1–23.
Srinivasan, S. S., Anderson, R., & Ponnavolu, K. (2002). Customer loyalty in e-commerce: An exploration of its antecedents and consequences. *Journal of Retailing*, *78*(1), 41–50.
Sullivan, D. (2000). The need for text mining in business intelligence. *DM Review*, *10*, 12–16.
Suo, Q., Sun, S., Hajli, N., & Love, P. E. (2015). User ratings analysis in social networks through a hypernetwork method. *Expert Systems with Applications*, *42*(21), 7317–7325.
Tam, J. L. M. (2008). Brand familiarity: Its effects on satisfaction evaluations. *Journal of Services Marketing*, *22*(1), 3–12.
Teas, R. K. (1993). Expectations, performance evaluation, and consumers' perceptions of quality. *Journal of Marketing*, *57*(4), 18–34.
Torres, E. N., & Kline, S. (2013). From customer satisfaction to customer delight: Creating a new standard of service for the hotel industry. *International Journal of Contemporary Hospitality Management*, *25*(5), 642–659.
US Census Bureau, Population Division. (2015). Table 1: Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More, Ranked by July 1, 2014 Population: April 1, 2010 to July 1, 2014 − United States — Places of 50,000+ Population. 2014 Population Estimates. Retrieved from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.
Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*, *57*, 77–93.
Wang, Y., Hsiao, S. H., Yang, Z., & Hajli, N. (2016). The impact of sellers' social influence on the co-creation of innovation with customers and brand awareness in online communities. *Industrial Marketing Management*, *54*, 56–70.
Wang, Y., Kung, L., & Byrd, T. A. (in press). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change.
Wixom, B., & Watson, H. (2010). The IB-based organization. *International Journal of Business Intelligence Research*, *1*(1), 13–28.
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? *International Journal of Hospitality Management*, *44*, 120–130.
Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, *43*, 1–12.
Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., et al. (2016). Big data: From beginning to future. *International Journal of Information Management*, *36*(6), 1231–1247.
Yavas, U., & Babakus, E. (2005). Dimensions of hotel choice criteria: Congruence between business and leisure travelers. *International Journal of Hospitality Management*, *24*(3), 359–367.
Yeung, T. A., & Law, R. (2004). Extending the modified heuristic usability evaluation technique to chain and independent hotel websites. *International Journal of Hospitality Management*, *23*, 307–313.
Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2006). *Service marketing*. Boston: McGraw-Hill.
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. *International Journal of Hospitality Management*, *29*(4), 694–700.
Zhou, L. Q., Ye, S., Pearce, P., & Wu, M. Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, *38*, 1–10.

فری پیپر freepaper.me