

Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality

Wenjing Duan
The George Washington
University
wduan@gwu.edu

Qing Cao
Texas Tech University
qing.cao@ttu.edu

Yang Yu
Texas Tech University
yang.yu@ttu.edu

Stuart Levy
The George Washington
University
slevy@gwu.edu

Abstract

This study aims to look beyond the quantitative summary to provide a more comprehensive view of online user-generated content. We obtain a unique and extensive dataset of online user reviews for hotels across various review sites and over a long time periods. We use the sentiment analysis technique to decompose user reviews into five dimensions to measure hotel service quality. Those dimensions are then incorporated into econometrics models to examine their effect in shaping users' overall evaluation and content generating behavior. The results suggest that different dimensions of user reviews have significantly differential impact in forming user evaluation and driving content generation.

1. Introduction

The unprecedented and increasing prevalence of the Internet has enabled the fast growth of user-generated content in various social media platforms, which has become the primary source of information for both consumers and businesses. According to Jones [25], social media sites come in a number of forms, including strictly review sites (e.g., TripAdvisor and Yelp), social networks and microblogging sites (e.g., Facebook, Twitter), social sharing (e.g., YouTube), and purchase/review sites (e.g., Amazon and Travelocity). In particular, user-generated reviews for products and services posted on the Internet have become the most powerful information sharing tool for digital customer-to-customer interactions. Online users increasingly rely on such digitized word of mouth information to evaluate products and services. Business are more and more recognizing the importance of using the information to collect feedback and evaluations of their own and their competitors' products and performance, to better understand consumer behavior and expectations, as well as identifying market opportunities.

Recent studies have shown strong support that online user reviews affect retail sales and consumer purchase decisions ([2], [8], [12], [13], [17], [18], [19],

[40]). However, the large majority of these studies have only used the quantitative summaries of user-generated content, such as overall valence and volume of user review ratings, to represent user opinions. There are only a handful recent studies that formally incorporate and test the impact of the textual content of user-generated reviews ([2], [18], [45]). Among those studies, only [2] considers multi-dimensional features in evaluating the economic impact of online user reviews, whereas economics and marketing theories (e.g., [59]) specify that products and services have multiple attributes. Individuals tend to place different important levels of importance on different attributes and features of products and services. Therefore, only considering the quantitative summarization, even with uni-dimensional classification of the textual content, is not sufficient in generating a comprehensive and accurate evaluation of user-generated content.

In this study, our objectives are to investigate user-generated content, particularly online user reviews, by considering both quantitative aspects and textual content from multi-dimensional perspectives. In particular, we study the vast amount of online user-generated reviews for the hotel industry. Online user reviews have exploded in recent years, revolutionizing the hotel industry. Travel reviews from other consumers influence half of all hotel purchase decisions, or over \$10 billion in online travel. In contrast to the majority of extant studies that focus on one-time purchase experience goods such as books, movies, and consumer electronics, we are one of the first studies to focus on the service industry, particularly the hotel industry, which can also be viewed as repeated purchase products.

We draw upon the well-established service quality literature, which is largely built on more subjective self-reporting survey methods, as the foundation of our study. In the current study, we use the more objective online user reviews to better measure hotel service quality and performance. We obtained an extensive data source of online hotel user reviews across various review sites and over a long time period, from nearly the very existence of online hotel user reviews. The richness of the data not only allows us to better

understand the trend of user reviews and market structure, but also enables us to closely examine the dynamics of the reviews. We employ the sentiment analysis technique to mine the textual content of user reviews and classify them into different dimensions according to the theoretical conceptualization of service quality and performance in extant literature. The sentiment analysis results show high level of accuracy in capturing and measuring service quality dimensions compared with existing text mining studies. We then use an econometric modeling technique to examine the potential differential effect of different service quality dimensions. Our results suggest that different dimensions of service quality embedded in user reviews account for significantly different weight in the review textual content. In addition, the effect of different dimensions also varies considerably in forming users overall evaluation of hotels and driving consumers online reviewing behavior.

The paper proceeds as follows. Related work is reviewed and discussed to provide the theoretical background and foundation for our study. We then describe the data and provide detailed discussion on sentiment analysis procedures. Next we formulate the econometrics models and present the estimation results. We conclude the paper by discussing study implications and suggesting future research directions.

2. Theoretical Background and Related Literature

As the global economy is moving towards heavy service-orientation, service quality is becoming the quintessential critical successful factor for business performance ([4], [9]). Online reviews represent one particular aspect of consumer-generated media and have exploded in recent years, revolutionizing the retail and service industries. In this study, we draw on two main streams of research from both the marketing and information systems domains. The first stream is on identifying and measuring service quality and service performance. The second stream is on examining the impact of digital WOM (e.g., online user reviews) on retail sales and consumer decision-making.

SERVQUAL vs. SERVPERF Models

Two most prevalent service quality models developed in the literature are SERVQUAL and SERVPERF. [55] propose a model measuring service quality (a.k.a. SERVQUAL) based on customers' evaluation of their service encounters. In other words, SERVQUAL can be conceptualized as the "gap" between what customers feel that a service should offer (their expectations) and their perceptions of the actual performance of the service [54]. Subsequently, five

generic service quality dimensions (TANGIBLES, RELIABILITY, RESPONSIVENESS, ASSURANCE and EMPATHY) are established and these five dimensions are measured by 22 survey items in totality for expectations and perceptions respectively [53]. Since its inception, SERVQUAL has been widely accepted as the de facto model in service quality research and it has been applied in various service industries such as the healthcare ([29], [32]), banking ([33], [70]), food and wine ([34], [49]), information systems ([24]), retailing ([53], [56]) and hospitality ([22]).

However, the SERVQUAL model also draws its share of criticisms from service quality researchers (e.g., [3], [5], [10], [29], [33]). One of the main criticisms lies in the basic assumption of how the service quality is determined in the SERVQUAL model (i.e., gap analysis). [5] argues that the gap scores between expectations and perceptions of service quality are unlikely to be distinct from their component scores (e.g., expectations and perceptions) as there is little theoretical or empirical evidence to support the expectation-perception gap as the basis for measuring service quality [9]. As such SERVQUAL may suffer various reliability and validity problems ([29], [32]). In addition, the operationalization of service quality in SERVQUAL also confounds satisfaction and attitude [10]. To overcome these shortcomings of the SERVQUAL model, [10] proposes the SERVPERF model, which are derived from the perception only scores. Employing the SERVPERF, [3] find SERVPERF is superior to SERVQUAL for measuring service quality. However, [7] do not find significant difference between the two models in terms of predictive validity. [29] further argue that while the perception based SERVPERF model seems to possess superior power in reliability and validity, the gap based SERVQUAL model appears to be more viable in finding service deficiencies. We believe that choice of service quality model is contingent upon the nature of the research and the data. The predominant usage of survey methodology to measure service quality and performance in extant study has set many constraints in terms of conceptualization and measurement of consumers' true perceptions.

The service quality evaluated by consumers for both SERVQUAL and SERVPERF models comprise of the same five dimensions, namely, TANGIBLES, RELIABILITY, RESPONSIVENESS, ASSURANCE and EMPATHY ([10], [55]). According to [55], tangibles refer to the appearance of physical facilities, equipment, and personnel; reliability represents the ability to perform the promised service dependably and accurately; responsiveness means the willingness to help customers and provide prompt service; assurance characterizes the knowledge and courtesy of employees

and their ability to inspire trust and confidence; and empathy demonstrates the level of caring and individualized attention the firm provides to its customers. The differences between SERVQUAL model and SERVPERF model in terms of dimensionality lie in how they are measured. While a certain dimension of the SERVQUAL model is measured by the gap or difference score between expectation and perception of the dimension, an individual dimension of the SERVPERF model is evaluated solely based perception of the dimension.

To the best of our knowledge, almost all the service quality research in the hotel industry are based on the SERVQUAL model or its variations ([22], [27]). As such, they suffer inherently from the pitfalls of the SERVQUAL model, especially by employing survey methodology measuring users' perceptions and expectations. In the current study, we apply SERVPERF model in order to catch the content of online user reviews more accurately.

Online User Reviews and Hotel Industry

Social media provides consumers with a platform for interactivity, and interactivity leads to consumer empowerment by providing the ability to make his or her voice heard [58]. The importance of online user reviews, as a particular format of social media and a major form of electronic (digital) word-of-mouth (eWOM), has been extensively documented in recent literature (e.g., [6], [8], [12], [13], [21]). While a growing body of studies have shown that the influence of user reviews is important for experience goods such as books, CDs, and electronics (e.g., [2], [21], [37]), consumer generated reviews of travel and hospitality services have been found to be a particularly critical information resource for travelers [51].

The explosive growth of hotel room bookings through e-distribution channels ([48]) has naturally coincided with a corresponding unprecedented increase of online reviews. [47] suggested that hotels need to actively embrace the concept of social networks and user-generated content to monitor reviews and manage online reputation, as faceless reviewers are rapidly becoming the travel opinion leaders of the electronic age [41]. Studies on online consumer review sites (i.e., TripAdvisor or Yelp) are relatively limited [47], although travel-related online review sites were found to comprise over one-quarter of social media websites found on the Internet. [38], in an investigation of hospitality marketing academic research published between 2008 and 2010, indicated that *"despite the explosion of social media and the emergent Web 2.0 phenomenon in recent years, very little attention has been given to marketing applications of these [online review] phenomena with the hospitality field"* (p.7).

Our study differs from and complements the extant literature on service quality and online user reviews in several perspectives. Drawing upon the SERVQUAL and SERVPERF models, we employ the five major dimensions as the foundation to guide our sentiment analysis. This approach overcomes the major drawbacks of previous operationalization of those five dimensions primarily using the self-reporting survey items; instead we derive user opinions directly from the content of their reviews. Our study is among the first to invest the effect of textual content of online user reviews in the hospitality industry. In addition, our study is also among the first to measure and evaluate the effect of multifaceted effects of online user reviews. Furthermore, our study integrate the sentiment analysis technique with the econometrics modeling technique in order to synergize the power of examining both the qualitative content and testing the quantitative impact of online user reviews. Lastly, we view the measurement of service quality and performance as a dynamic process in contrast to previous literature that largely solicit user opinions and perceptions in one snapshot. Our extensive and longitudinal dataset and dynamic modeling approach allow us to provide a better understanding on this dynamic process.

3. Data and Sentiment Analysis

Research Context

Our online user reviews data were collected by a leading provider of social media monitoring tools for the hospitality industry. The data consists of all online user reviews posted in various online venues from 1999-2011 for 86 hotels in the Washington D.C. area. This substantive data sample consists of total 70,103 online user reviews. Besides collecting overall hotel characteristics, for each entry of user review, the company collects the review body (review content), review rating (1-5), review site, date of review, author, author location, trip type, and hotel responses if available. As the company launched in early 2010, TripAdvisor popularity ranking started to be collected in February 2010 on a daily basis. Table 1 summarizes the hotels and reviews according to their chain scale segment categorization.

Table 1. Summary of Hotel Characteristics

Chain Scale Segment	N	Number of Reviews (%)
Luxury	17	12,937 (18.46)
Upper upscale	22	22,338 (31.86)
Upscale	8	4,861 (6.93)
Upper midscale	4	2,546 (3.63)
Midscale	1	326 (0.47)

Economy	3	1,185 (1.69)
Independent	31	25,910 (36.96)
Total	86	70103

Our data show that online user reviews really start getting popular after 2006. TripAdvisor, Hotels.com, and Priceline seems to be the leading site that users post reviews, which also indicate those three websites attract most online user traffic. Volume of online user reviews for hotels start picking up after year 2006, and TripAdvisor was not the leader initially, but developed and catching up very fast to become the frontrunner undisputedly in the market. Though there are up and downs for other sites over the years.

Sentiment Analysis and Procedures

We employ the sentiment analysis technique to take a more granular look at the review content in order to more efficiently and accurately measure the service quality based on user feedback. Sentiment analysis is the computational detection and study of opinions, sentiments, emotions, and subjectivities in text ([36], [39], [52]). To accomplish the goal of mining opinions, the sentiment analysis technique is often split into two consecutive tasks: detecting which text segments (e.g., sentences) contain sentiments, and determine the polarity and even strength of that sentiment [52]. Thus, sentiment analysis' main purpose is to determine the attitude of a speaker or a writer on some specific topics. The use of sentiment analysis and related approaches has gained great popularity in the past decade due to several factors, including the advance of machine learning methods in natural language processing and information retrieval, the availability of large and rich datasets for machine learning algorithms to be trained on, and the realization of many commercial intelligence applications ([11], [52]).

There are three key processes involved in the sentiment analysis procedures: pretreatment, SERVPERF dimension classification and sentiment polarity classification. In the pretreatment process, we clean up the raw customer reviews and then store them as the refined corpus in a computable format. In this research, first, we clean the raw content by deleting blank records and duplicated ones. Next, sentences are recognized from piece of review and all the words in each sentence are normalized before storing. In the second process, we decide if the sentences refined by the first step fall into one of the specific service quality dimensions. During this process, both domain knowledge and the machine learning algorithm set will provide effective support for natural language processing directed at the interpretation. Furthermore, we create the sentiment classifier in the third process and assign sentiment polarity for each unit of analysis,

in this case, a sentence. During this process, the associated sentiment training corpus will also be integrated into the support set.

Using text classification algorithm, we mine customers' opinion toward a hotel based on the five dimensions of perceived service quality (TANGIBLES, RELIABILITY, RESPONSIVENESS, ASSURANCE, and EMPATHY) from the content of reviews. Finally, a SERVPERF opinion matrix is derived to show customers' sentiment on each dimension (a score from -1 to 1), where a score of 1 means the customer has the most positive view of that respective dimension, and -1 denotes the most negative sentiment. The overall sentiment score of dimension i is calculated by the following formula:

$$S_i = \frac{N_{pi} - N_{ni}}{N_{pi} + N_{ni}} \quad (1)$$

where N_{pi} denotes the number of positive sentences in dimension i and N_{ni} denotes the number of negative sentences in dimension i . Some of previous study use ternary classification to represent sentiment polarity, positive, negative and neutral [52]. We only use positive and negative labels in this study based on two main reasons. The first reason is that sentence includes subjective expressions always imply either positive or negative feelings. Neutral is a fairly vague range, which is much less accurate to identify. The second consideration is from the machine learning perspective, there is no mature sentiment repertoires developed to efficiently and accurately identify neutral opinions.

The sentiment analysis is conducted using the Naïve Bayes (NB) process, which is a simple but effective classifier that has been used in numerous of information processing techniques such as image recognition, NLP, information retrieval, etc, based on the open-source Natural Language Toolkit (NLTK) ([16], [35], [44], [57]). In the first pretreatment process, after raw material cleanup and sentence tokenization,¹ we stored 575,228 sentences covers 64,806 pieces of reviews. Following the sentiment analysis procedure discussed beforehand, in the second process, we train the classification system with external knowledge supporting. The external knowledge, in our context, includes the five key words list to represent the five dimensions of SERVPERF. These keywords are selected manually by reading 500 reviews which were randomly chosen from 64,806 reviews. We choose the nouns to represent each dimension based on previous research [42, 71] and experts' domain knowledge. In addition, we import the Cornell movie-review dataset for sentiment signals in the third process. Moreover, we compute the accuracy of classifier on the test set

¹ Sentence tokenization describes the process of breaking a piece of text into individual sentences.

and use the F-measure to evaluate the performance based on precision and recall.

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

where TP is the number of true positive, FP is the number of false positives, and FN is the number of false negatives.² In this study, the proposed five dimension classification algorithm get 0.68 F-measure and positive-negative sentiment classification algorithm get 0.91 F-measure on the test set. On average, F-measure of quinary (five dimension classification case) and binary (positive-negative sentiment classification case) is around 0.6 and 0.8 respectively [52]. Therefore, our classification is fairly accurate.

The result for the five dimension classification is shown in Table 2. There are 443,568 sentences related to SERVPERF and number of positive sentences is significantly more than that of negative sentences. In addition, we notice the dimension of TANGIBLES has the largest number of sentences, accounts for almost 70% of the total counts, which suggest users place a lot of attention on the items of this dimension such as physical facilities, equipment, and personnel.

Table 2. Distribution of the Five Dimensions of SERVPERF

Dimension	# of Positive	# of Negative	Number of Sentences (%)
TANGIBLES	205,189	103,815	309,004 (69.66)
RELIABILITY	28,977	17,696	46,673 (10.52)
ASSURANCE	29,224	12,025	41,249 (9.3)
RESPONSIVE NESS	17,288	9,128	26,416 (5.96)
EMPATHY	16,540	3,686	20,226 (4.56)
Total			443,568

4. Empirical Modeling

As we are interested in examining whether the five SERVPERF dimension measures derived from the sentiment analysis drive online users' reviewing behavior, we construct a system of two *interdependent* equations: one equation with the weekly average user

review *rating* as the dependent variable and the other with weekly *number* of reviews as the dependent variable. We assume that in each time period (i.e., week), the errors in the two equations may be correlated, which implies that factors not included in our model could simultaneously influence both the number and rating of online user reviews for hotels. In addition, following extant research (e.g., [13], [21]), we use a log-linear model. To control for any hotel idiosyncratic factors that could influence user reviews, we include hotel fixed effects in the model by adding hotel-specific dummy variables. The advantage of fixed effects estimation is that it controls for intrinsic hotel characteristics, which inherently affect user reviews. In addition, fixed effects estimation also allows the error term to be arbitrarily correlated with other explanatory variables, thus making the estimation results more robust. The system of equations is specified as follows:

$$\begin{aligned} \ln(WeekRating_{it}) = & \alpha_{r0} \ln(WeekRatingNum_{it}) + \\ & \alpha_{r1} \ln(WeekRank_{i,t-1}) + \\ & \alpha_{r2} \ln(CumuRating_{it}) + \\ & X_{it}^c \beta_r^c + \mu_{ci} + \varepsilon_{cit} \end{aligned} \quad (5)$$

$$\begin{aligned} \ln(WeekRatingNum_{it}) = & \alpha_{n0} \ln(WeekRating_{it}) + \\ & \alpha_{n1} \ln(WeekRank_{i,t-1}) + \\ & \alpha_{n2} \ln(CumuRatingNum_{it}) + \\ & X_{it}^c \beta_n^c + \rho_{ci} + \sigma_{cit} \end{aligned} \quad (6)$$

Eq. (5) reflects the weekly average review rating and Eq. (6) displays the number of weekly online user reviews. Let $i = 1, \dots, N$ index the hotels. X_{it}^c is a vector of ten independent variables including the log value of the *count* of positive and negative sentences for each of the five dimensions (i.e. TANGIBLES, RELIABILITY, RESPONSIVENESS, ASSURANCE, and EMPATHY) in reviews for hotel i at week t . $CumuRating_{it}$ in Eq. (5) represents the cumulative average user review grade of hotel i by week t . Since online travel sites often provide average user review grade besides the hotel listing, it is the most noticeable information on the website. While $CumuRating_{it}$ measures the overall evaluation of the hotel experience from the users, $WeekRating_{it}$ reflects the most recent and concurrent valence of online user WOM information for the hotel. $CumuRatingNum_{it}$ in Eq. (6) represents the cumulative number of online user reviews for hotel i by week t , which indicate the total volume of WOM information for that hotel.

² Both True positive and True negative mean the correct outcomes, false positive represent the Type I error and False negative means Type II error. In quinary classification, we will calculate the precision/recall for each dimension then take average to get the precision/ recall for quinary classifier.

In both equations, we include the previous week's TripAdvisor popularity index (ranking)³ ($\text{Ln}(\text{WeekRank}_{i,t-1})$) to control for the most recent travelers' satisfaction level. Hotels are categorized in the same market based on their geographic location. For example, there are currently 127 hotels classified in the Washington D.C. area, then the ranking will be displayed as “*Ranked #2 of 127 hotels in Washington DC*” on TripAdvisor webpage. TripAdvisor is a prominent example of an eWOM platform within the travel industry [23]. TripAdvisor has harvested notable success in a short period and attracts millions of global visitors on a daily basis. Many hotels aspire to achieve top ranking in TripAdvisor's popularity index, which instantly signals a hotel's level of quality. The popularity ranking along with the cumulative average user review rating and number of reviews are the most prominently displayed information on TripAdvisor website. μ_{ci} and ρ_{ci} denote the hotel-specific fixed effects that capture the idiosyncratic characteristics associated with each hotel, such as its size, location, brand name, star level, and style. The fixed effects capture all non-time varying, unobserved heterogeneity of each hotel; thus, we are able to control for unobserved differences across hotels.

In Eqs. (7) and (8), we use the overall sentiment score for each of the five dimensions instead of the count of the positive and negative sentences. X_{it}^s is a vector of five covariates including the sentiment score for each of the five dimensions.

$$\begin{aligned} \text{Ln}(\text{WeekRating}_{it}) = & \gamma_{r0} \text{Ln}(\text{WeekRatingNum}_{it}) + \\ & \gamma_{r1} \text{Ln}(\text{WeekRank}_{i,t-1}) + \\ & \gamma_{r2} \text{Ln}(\text{CumRating}_{it}) + \\ & X_{it}^s \delta_r^c + \mu_{si} + \varepsilon_{sit} \end{aligned} \quad (7)$$

$$\begin{aligned} \text{Ln}(\text{WeekRatingNum}_{it}) = & \gamma_{n0} \text{Ln}(\text{WeekRating}_{it}) \\ & + \gamma_{n1} \text{Ln}(\text{WeekRank}_{i,t-1}) \\ & + \gamma_{n2} \text{Ln}(\text{CumRatingNum}_{it}) \\ & + X_{it}^s \delta_n^c + \rho_{si} + \sigma_{sit} \end{aligned} \quad (8)$$

5. Results and Discussion

The three-stage least-square (3SLS) procedure was used to simultaneously estimate the system of two equations to consider both the endogeneity of the dependent variables and the correlation of the two

equations. The 3SLS estimation results for Eqs. (5) and (6) are reported in Table 3.

For the user review rating equation in Table 3, the coefficient of $\text{Ln}(\text{WeekRank}_{i,t-1})$ is significantly positive, suggesting that higher TripAdvisor ranking (smaller number) in the previous week is correlated to lower user review rating in the following week. This result indicates that a better TripAdvisor ranking actually may lead to lower user review ratings. Though a bit counterintuitive, this finding echoes a recent work by [26] that models the strategic implications of WOM. “Underpromising and overdelivering” is the common advice given to firms for managing customers' quality expectations. [26] shows that under the influence of WOM, it is likely that firm can strategically set consumer expectations toward either direction. Especially when repeat purchases are critical, as it is in the lodging industry, it is important for companies to consider consumers with adaptive anticipation to set expectations in order to optimally influence demand generation ([30], [31]).

Travelers' quality expectation toward a hotel can be easily formed by reading online user reviews and by looking at the TripAdvisor popularity index. Previous literature on herding and informational cascades (e.g., [14]) also suggested that online users' choices of products exhibit distinct jumps and drops with changes in sales ranking, and most popular products could easily drive more consumers to follow. In our context, when top-ranked hotels do not meet users' high expectations, they are more likely to share the negative (i.e., lower than expectation) experience by posting negative reviews.

The coefficient of $\text{Ln}(\text{CumRating}_{it})$ is significantly positive, indicating that higher cumulative average user review rating is correlated to higher current average user review rating. It is noteworthy to see the differential (in fact different direction) impact of cumulative user rating and popularity ranking. Both pieces of information, when available, are often placed in the most prominent locations on many websites, assuming they would somewhat complement and reinforce each other. But our results suggest that companies may need to take another look at this strategy when trying to maximize informational effects of online user-generated content.

In terms of the effect of the five SERVPERF dimensions, TANGIBLES seem to have the strongest impact on user review ratings. The positive number of mentions for this dimension has strong positive effect on user review ratings, while the negative number of mentions has strong negative impact. The impact of positive mentions (the coefficient is 0.02) is much stronger than that of negative mentions (the coefficient is -0.01). In contrast to the common wisdom that negative reviews may have stronger impact than

³ The TripAdvisor Popularity Index (ranking) incorporates Traveler Content to determine traveler satisfaction. Emphasis is placed on the most recent information. TripAdvisor calculate the Popularity Index using a sophisticated algorithm.

positive reviews [19], our result suggests that this may not be the case when delving deeper into the review content. This indicates that users may place different weight on hotel attributes when reading the reviews. Our results also show that such differential impact between positive and negative reviews is not only present within the same dimensions, but also across the different dimensions. The coefficient of the number of positive mentions of RELIABILITIES is positive and marginally significant in influencing the review ratings, but the number of negative mentions is not significant. For RESPONSIVENESS, only the negative number of mentions has a strong negative impact, but the number of positive mentions has no significant impact. For ASSURANCE, however, only the number of positive mentions has a strong positive impact. Neither the number of positive nor negative mentions of EMPATHY has any significant impact on review ratings.

Table 3. Fixed Effects 3SLS Estimation Results

Variable	Coefficient	Std. Err.
Eq. (5): User review rating equation		
$\ln(\text{WeekRatingNum}_{it})$	-0.07	0.04
$\ln(\text{WeekRank}_{i,t-1})$	0.03***	0.01
$\ln(\text{CumuRating}_{it})$	1.65***	0.29
$\ln(\text{TanPosNum}_{it})$	0.02***	0.004
$\ln(\text{TanNegNum}_{it})$	-0.01***	0.001
$\ln(\text{ReliPosNum}_{it})$	0.002*	0.001
$\ln(\text{ReliNegNum}_{it})$	0.001	0.001
$\ln(\text{ResPosNum}_{it})$	0.001	0.001
$\ln(\text{ResNegNum}_{it})$	-0.002***	0.001
$\ln(\text{AssuPosNum}_{it})$	0.002**	0.001
$\ln(\text{AssuNegNum}_{it})$	-0.001	0.001
$\ln(\text{EmpPosNum}_{it})$	0.01	0.001
$\ln(\text{EmpNegNum}_{it})$	0.001	0.001
Constant	-0.12	0.41
Eq. (6): User review volume equation		
$\ln(\text{WeekRating})$	-0.11	0.4
$\ln(\text{WeekRank}_{i,t-1})$	0.02	0.02
$\ln(\text{CumuRatingNum}_{it})$	0.35***	0.03
$\ln(\text{TanPosNum}_{it})$	0.07***	0.01
$\ln(\text{TanNegNum}_{it})$	0.02***	0.003
$\ln(\text{ReliPosNum}_{it})$	0.02***	0.001
$\ln(\text{ReliNegNum}_{it})$	0.01***	0.001
$\ln(\text{ResPosNum}_{it})$	0.01***	0.001
$\ln(\text{ResNegNum}_{it})$	0.007***	0.002
$\ln(\text{AssuPosNum}_{it})$	0.02***	0.001

$\ln(\text{AssuNegNum}_{it})$	0.01***	0.001
$\ln(\text{EmpPosNum}_{it})$	0.02***	0.002
$\ln(\text{EmpNegNum}_{it})$	0.01***	0.001
Constant	-0.19	0.46

N = 5045 Group = 86

Note: ***p < .01; **p < .05; *p < .10; Hotel dummies (fixed effects for each of the 86 hotels) are not reported.

For the user review volume equation, the coefficient of $\ln(\text{CumuRatingNum}_{it})$ is significantly positive, showing a strong correlation between cumulative number of reviews and the most recent number of reviews. This suggests a strong effect of past overall volume of WOM in generating more volume of WOM. This result is related to the recent work on the detailed look at the self selection, temporal, and sequential dynamics of online user reviews ([20], [37]). Different from the extant work, which usually investigate the one-time purchase of experience good such as books and cameras, the choice of hotels can be repeated and adaptive. Thus the examination of the online user reviews for hotels is not confounded by the normal product adoption and diffusion process, which biased many previous works on studying the dynamics of online user reviews. Also in contrast to the findings of extant literature that usually identifying a diminishing trend of reviews, our results suggest more previous reviews actually lead to more future reviews. This finding confirms the importance of examining different industries, and calls for significant future research on identifying and conceptualizing online user content generation behavior.

The results shown in Table 3 suggest that all ten sentiment variables have strong positive impact on driving number of reviews, regardless of positive or negative number of mentions. This finding suggests that online users review behavior in the hotel industry is significantly driven by their experience in these five major areas. This finding demonstrates that the five dimensions defined and measured in our study indeed provide a comprehensive measurement of service performance from different perspectives. More importantly, the results validate the importance and power of the novel text mining and sentiment analysis technique in order to more efficiently and accurately evaluate online user-generated content.

After separating the effect of positive and negative mentions of the five dimensions of SERVPERF, we would like to consider the overall impact of each dimension by estimating Eqs. (7) and (8) by using the sentiment score of each of the five dimensions. The 3SLS estimation results for Eqs. (7) and (8) are reported in Table 4.

The magnitude and significance level of the coefficients of $\ln(\text{WeekRank}_{i,t-1})$ and

$\ln(\text{CumuRating}_{it})$ in the estimation results of Eq. (7) remain qualitatively similar to the estimation results of Eq. (5). The overall sentiment of TANGIBLES has the strongest and largest impact on user review ratings, which suggests the more positive reviews in this dimension contributes to the most of user review ratings. The coefficients of the RESPONSIVENESS and ASSURANCE are also significantly positive, but the magnitude is much smaller than that of TANGIBLES. When we look at the estimation results of Eq. (5) and Eq. (7) together, the importance of investigating the review content at both sentiment variation and overall sentiment levels are not negligible. For example, if only the results of Eq. (5) are provided, the conclusion could be that effort should be invested to increase the RESPONSIVENESS and ASSURANCE service in order to improve the user review ratings. However, by also considering the results in Eq. (7), managers should pay more attention to improve the negative feedback related the RESPONSIVENESS, and be more aware of and maintain the positive services of the ASSURANCE.

In Eq. (8), the coefficient of $\ln(\text{CumuRatingNum}_{it})$ remains significantly positive as it does in Eq. (6). The coefficient of the overall sentiment of TANGIBLES is not significant, whereas the coefficients of the sentiment score of the other four dimensions are all significantly negative. This finding suggest that overall negative reviews would drive higher number of reviews, consist with extant literature and anecdotal evidence that users are more likely to share their negative experiences in posting reviews and feedback online. The insignificance of the sentiment of TANGIBLES is in fact also consistent with the estimation results of Eqs. (5) and (6) that positive sentiment on this dimension has larger impact than negative sentiment as we discussed earlier.

To summarize, in light of the rapid growth of the social media phenomenon and its undeniable impact on consumers and businesses, we endeavor to pioneer the social media research in the hotel industry focusing on measuring the service quality and performance by mining the textual content of online user reviews. First, we extend the SERVPERF model by analyzing the more objective data gathered from online customers' review content on multiple dimensions of hotel service quality applying the sentiment analysis technique. We provide evidence that online customer reviews form service quality perceptions on the basis of their differential evaluations of all five dimensions of SERVPERF. This result demonstrate the advantage of using the novel automated text mining approach to more accurately and efficiently measure consumer opinions from the large amount of online user-generated content. Second, we further investigate the

influence of different dimensions of derived from the sentiment analysis on consumers' reviewing and content generating behavior. Our findings indicate different dimensions have different impact in shaping consumers' overall evaluation, and in driving consumers' review posting behavior. In addition, our results indicate that more previous reviews actually lead to more future reviews, showing the strong clout of the social media impact, and the importance to also understand the uniqueness of applying social media marketing strategy in different industries.

Table 4. Fixed Effects 3SLS Estimation Results

Variable	Coefficient	Std. Err.
Eq. (7): User review rating equation		
$\ln(\text{WeekRatingNum}_{it})$	-0.01	0.02
$\ln(\text{WeekRank}_{i,t-1})$	0.03***	0.01
$\ln(\text{CumuRating}_{it})$	1.59***	0.29
TanSenti_{it}	0.11***	0.01
ReliSenti_{it}	-0.003	0.005
ResSenti_{it}	0.01***	0.005
AssuSenti_{it}	0.01**	0.005
EmpSenti_{it}	-0.01	0.01
Constant	-1.02	0.47
Eq. (8): User review volume		
$\ln(\text{WeekRating})$	-0.23	0.61
$\ln(\text{WeekRank}_{i,t-1})$	-0.02	0.03
$\ln(\text{CumuRatingNum}_{it})$	0.76***	0.05
TanSenti_{it}	-0.02	0.07
ReliSenti_{it}	-0.17***	0.01
ResSenti_{it}	-0.15***	0.02
AssuSenti_{it}	-0.10***	0.02
EmpSenti_{it}	-0.15***	0.02
Constant	-2.33	0.98
N = 5045 Group = 86		

Note: ***p < .01; **p < .05; *p < .10; Hotel dummies (fixed effects for each of the 86 hotels) are not reported.

One limitation of this study has to do with the generalization issue as the study is anchored at the hotel industry. In a future study, we intend to provide a more comprehensive view of measuring and evaluating service quality by mining social media content and performing multiple tests and across various industries. Another methodological limitation associated with the current study is that we used sentence-level analysis and assigned each sentence to one dimension. This approach ignores some cases where content of one sentence may cover more than one dimension. More sophisticated and advanced sentiment analysis

technique is expected to be developed and applied in our future research endeavors in examining online user-generated content. Other future studies, as we allude to in the aforementioned discussions, includes taking a detailed look at the effect of various information aggregation and display strategy on consumer behavior, examining the effect of different strategies of managing and using user-generated content, and establishing the direct link of different dimensions of user-generated content with product and service sales as well as overall firm performance.

6. References

- [1] Al Khattab, S.A. and Aldehayyat, J.S., "Perceptions of Service Quality in Jordanian Hotels," *International Journal of Business and Management* (6:7), 2011, p226.
- [2] Archak, N., Ghose, A., and Ipeirotis, P.G., "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (56:8), 2011, pp. 1485-1509.
- [3] Brady, M.K., Cronin, J.J., and Brand, R.R., "Performance-only Measurement of Service Quality: A Replication and Extension," *Journal of Business Research* (55:1), 2002, pp.17-31.
- [4] Brady, M.K., and Cronin, J.J., "Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach," *The Journal of Marketing* (65:3), 2001, pp. 34-49.
- [5] Brown, T.J., Churchill, G.A., and Peter, J.P., "Improving the Measurement of Service Quality," *Journal of Retailing* (69:1), 1993, pp.127-139.
- [6] Cao, Q., Duan, W., and Gan, Q., "Exploring Determinants of Voting for the 'Helpfulness' of Online User Reviews: A Text Mining Approach," *Decision Support Systems* (50:2), 2011, pp.511-521.
- [7] Carrillat, F.A., Jaramillo, F., and Mulki, J.P., "The Validity of the SERVQUAL and SERVPERF Scales: A Meta-analytic View of 17 Years of Research across Five Continents," *International Journal of Service Industry Management* (18:5), 2007, pp.472-490.
- [8] Chevalier, J. A., and Mayzli, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Marketing Science* (43:3), 2006, pp. 345-354.
- [9] Cronin, J.J., Brady, M.K., and Hult, G.T.M., "Assessing the Effects of Quality, Value, and Customer Satisfaction on Consumer Behavioral Intentions in Service Environments," *Journal of retailing* (76:2), 2001, pp. 193-218.
- [10] Cronin, J.J., and Taylor, S.A., "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing* (56:3), 2003, pp. 55-68.
- [11] Das, S., and Chen, M., "Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web," in *EFA 2001 Barcelona Meetings*, 2001.
- [12] Duan, W., B. Gu, and Whinston, A.B., "Do Online Reviews Matter? – An Empirical Investigation of Panel Data," *Decision Support Systems* (45:4), 2008, pp. 233-42.
- [13] Duan, W., B. Gu, and Whinston, A.B. "The Dynamics of Online Word-of-mouth and Product Sales—An Empirical Investigation of the Movie Industry," *Journal of Retailing* (84:2), 2008, pp. 233-242.
- [14] Duan, W., B. Gu, and Whinston, A.B., "Informational Cascades and Software Adoption on the Internet: An Empirical Investigation," *MIS Quarterly* (33:1), 2009, pp. 23-48.
- [15] Van Dyke, T.P., Kappelman, L.A., and Prybutok, V.R., "Measuring Information Systems Service Quality: Concerns on the Use of the SERVQUAL Questionnaire," *MIS quarterly* (21) 2, 1997, pp. 195-208.
- [16] Escudero, G., Marquez, L., and Rigau, G., "Naive Bayes and Exemplar-based Approaches to Word Sense Disambiguation Revisited," in *Proceedings of the 14th European Conference on Artificial Intelligence*. Berlin, Germany, 2007.
- [17] Forman, C., Ghose, A., and Wiesenfeld, B., "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19) 3, 2008, pp. 291-313.
- [18] Ghose, A., and Panagiotis, I., "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering* (23:10), 2011, pp. 1498-1512.
- [19] Godes, D., and Mayzlin, D., "Using Online Conversations to Study Word-of-mouth Communication," *Marketing Science* (23:4), 2004, pp. 545-560.
- [20] Godes, D., and Silva, J.C., "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science* (43:3), 2011, pp. 345-354.
- [21] Gu, B., Park, J., and Konana, P., "The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products," *Information Systems Research* (23:1), 2011, pp. 182-196.
- [22] Heung, V.C.S., MY Wong, and Q. Hailin, "Airport-restaurant service quality in Hong Kong: an application of SERVQUAL," *Cornell Hotel and Restaurant Administration Quarterly* (41:3), 2000, pp.86-96.
- [23] Jeacle, I., and Carter, C., "In TripAdvisor We Trust: Rankings, Calculative Regimes and Abstract Systems," *Accounting, Organizations and Society* (36:4-5), 2011, pp. 293-309.
- [24] Jiang, J.J., Klein, G., and Crampton, S.M., "A Note on SERVQUAL Reliability and Validity in Information System Service Quality Measurement," *Decision Sciences* (31:3), 2000, pp.725-44.
- [25] Jones, Ron, "Social Media Marketing 101, Part 1," 2009, Accessed March 20, 2012, <http://searchenginewatch.com/article/2064413/Social-Media-Marketing-101-Part-1>.
- [26] Joshi, Y., and Musalem, A., "Underpromising and Overdelivering: Strategic Implications of Word of Mouth." Robert H. Smith School Research Paper No. RHS-06-142, 2009.
- [27] Juwaheer, T.D., "Exploring International Tourists' Perceptions of Hotel Operations by Using a Modified SERVQUAL Approach—A Case Study of Mauritius," *Managing Service Quality* (14:5), 2004, pp. 350-64.

- [28] Kaplan, A.M., and Haenlein, M., "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business horizons* (53:1), 2010, pp. 59–68.
- [29] Kilbourne, W.E., Duffy, J.A., Duffy, M., and Giarchi, G., "The Applicability of SERVQUAL in Cross-national Measurements of Health-care Quality," *Journal of Services Marketing* (18:7), 2004, pp.524–33.
- [30] Kopalle, P.K., and Assunção, J.L., "When (not) to Indulge in 'Puffery': The role of Consumer Expectations and Brand Goodwill in Determining Advertised and Actual Product Quality," *Managerial and Decision Economics* (21:6), 2000, pp.223–241.
- [31] Kopalle, P.K., and Lehmann, D.R., "Setting Quality Expectations When Entering A Market: What Should the Promise Be?," *Marketing Science* (25:1) 2006, pp.8–24.
- [32] Lam, S. S. K., "SERVQUAL: A Tool for Measuring Patients' Opinions of Hospital Service Quality in Hong Kong," *Total Quality Management* (8:4), 1997, pp.145–52.
- [33] Lam, T. K. P., "Making Sense of SERVQUAL's Dimensions to the Chinese Customers in Macau," *Journal of Market-Focused Management* (5:1), 2007, pp.43–58.
- [34] Lee, M., and Ulgado, F.M., "Consumer Evaluations of Fast-food Services: A Cross-national Comparison," *Journal of Services Marketing* (11:1), 1997, pp.39–52.
- [35] Lewis, D., "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," in *Proceedings of 10th European Conference on Machine Learning*, Chemnitz, Germany, 1998, pp. 4- 15.
- [36] Li, N., and Wu, D. D., "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast," *Decision Support Systems*, (48:2), 2010, pp. 354–368.
- [37] Li, X., and Hitt, L., "Self Selection and Information Role of Online Product Reviews," *Information Systems Research* (19: 4), 2008, pp. 456-474.
- [38] Line, N., and Runyan, R., "Hospitality Marketing Research: Recent Trends and Future Directions," *International Journal of Hospitality Management* (31:2), 2007, pp. 477-488.
- [39] Liu, B., "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing*, N. Indurkha and F. J. Damerau (eds.), 2010.
- [40] Liu, Y., "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), 2006, pp. 74–89.
- [41] Litvin, S., Goldsmith, R. E. and Pan., B., "Electronic Word-of-mouth in Hospitality and Tourism Management," *Tourism Management* (29:3), 2008, pp. 458-468.
- [42] Manickas, P.A., and Shea, L.J., "Hotel Complaint Behavior and Resolution: A Content Analysis," *Journal of Travel Research* (36:2), 1997, pp. 68–73.
- [43] McCarthy, L., Stock, D., and Verma, R., "How Travelers Use Online and Social Media Channels to Make Hotel-choice Decisions," *Cornell Hospitality Report* (10:18), 2010, pp. 4-18.
- [44] Nigam, K., and Ghani, R., "Analyzing the Effectiveness and Applicability of Co-training," in *Proceedings of the ninth international conference on Information and knowledge management*, McLean, VA, 2000, pp. 86–93.
- [45] Netzer, O., Feldman, R., Fresko, M., and Goldenberg, J., "Mine Your own Business: Market Structure Surveillance through Text Mining," Working Paper, Tech. rep., Columbia University, 2011.
- [46] Noone, B.M., McGuire, K.A., and Rohlf, K.V., "Social Media Meets Hotel Revenue Management: Opportunities, Issues and Unanswered Questions," *Journal of Revenue & Pricing Management* (10:4), 2011, pp. 293–305.
- [47] O'Connor, P., "Managing a Hotel's Image on TripAdvisor," *Journal of Hospitality Marketing &Management* (19:7), 2010, pp. 754-772.
- [48] O'Connor, P., and Murphy, J., "Hotel Yield Management Practices across Multiple Electronic Distribution Channels," *Information Technology & Tourism* (10:2), 2008, pp. 161-172.
- [49] O'Neill, M., and Palmer, A., "An Exploratory Study of the Effects of Experience on Consumer Perceptions of the Service Quality Construct," *Managing Service Quality* (13:3), 2003, pp. 187–196.
- [50] O'Neill, M.A., and Palmer, A., "Wine Production and Tourism," *Cornell Hotel and Restaurant Administration Quarterly* (45:3), 2004, pp. 269–84.
- [51] Pan, B., MacLaurin, T. and Crotts, J., "Travel Blogs and the Implications for Destination Marketing," *Journal of Travel Research* (46:1), 2007, pp. 35–45.
- [52] Pang, B., and Lee, L., "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Barcelona, Spain, 2004, pp. 271-278.
- [53] Parasuraman, A., Berry, L.L., and Zeithaml, V.A., "Refinement and Reassessment of the SERVQUAL Scale," *Journal of retailing* (67:4), 1991, pp. 420-450.
- [54] Parasuraman, A., Zeithaml, V.A., and Berry, L. L., "SERVQUAL: A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing* (64:1), 1998, pp. 12-40.
- [55] Parasuraman, A., Zeithaml, V.A., and Berry, L.L., "A Conceptual Model of Service Quality and Its Implications for Future Research," *The Journal of Marketing* (49), 1985, pp. 41–50.
- [56] Parasuraman, A., Zeithaml, V.A., and Berry, L.L., "Alternative Scales for Measuring Service Quality: A Comparative Assessment Based on Psychometric and Diagnostic Criteria," *Journal of Retailing* (70:3), 1994, pp. 201–30.
- [57] Pedersen, T., "A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation," in *Proceedings of the 1st North American chapter of the Association for Computational Linguistics Conference*, Stroudsburg, PA, 2000, pp. 63–69.
- [58] Pehlivan, E., Sarican, F., and Berthon, P., "Mining Messages: Exploring Consumer Response to Consumer-vs. Firm-generated Ads," *Journal of Consumer Behaviour* (10:6), 2011, pp. 313–21.
- [59] Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *The Journal of Political Economy* (82:1), 1974, pp. 34–55.