Tourism Management 58 (2017) 51–65

Contents lists available at ScienceDirect

Tourism Management

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

A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism

Zheng Xiang [a,*], Qianzhou Du [b], Yufeng Ma [c], Weiguo Fan [d]

[a] Department of Hospitality and Tourism Management, Pamplin College of Business, Virginia Tech, Blacksburg, VA 24061, USA
[b] Department of Business Information Technology, Pamplin College of Business, Virginia Tech, Blacksburg, VA 24061, USA
[c] Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA
[d] Center for Business Intelligence & Analytics, Department of Accounting and Information Systems, Pamplin College of Business, Virginia Tech, Blacksburg, VA 24061, USA

HIGHLIGHTS

- We applied text analytics to compare three major online review platforms, namely, TripAdvisor, Expedia, and Yelp.
- Findings show discrepancies in the representation of hotel product on these platforms.
- Information quality, measured by linguistic and semantic features, sentiment, rating, and usefulness, varies considerably.
- This study is the first to comparatively explore data quality in social media studies in hospitality and tourism.
- This study highlights methodological challenges and contributes to the theoretical development of social media analytics.

ARTICLE INFO

Article history:
Received 1 April 2016
Received in revised form
1 September 2016
Accepted 6 October 2016

Keywords:
Online reviews
Hotel industry
Information quality
Social media analytics
Text analytics
Machine learning

ABSTRACT

Online consumer reviews have been studied for various research problems in hospitality and tourism. However, existing studies using review data tend to rely on a single data source and data quality is largely anecdotal. This greatly limits the generalizability and contribution of social media analytics research. Through text analytics this study comparatively examines three major online review platforms, namely TripAdvisor, Expedia, and Yelp, in terms of information quality related to online reviews about the entire hotel population in Manhattan, New York City. The findings show that there are huge discrepancies in the representation of the hotel industry on these platforms. Particularly, online reviews vary considerably in terms of their linguistic characteristics, semantic features, sentiment, rating, usefulness as well as the relationships between these features. This study offers a basis for understanding the methodological challenges and identifies several research directions for social media analytics in hospitality and tourism.

1. Introduction

There is a growing literature on social media analytics that combines Web crawling, computational linguistics, machine learning and statistical techniques to collect, analyze, and interpret the so-called big data for business purposes such as tracking trending topics and popular sentiments as well as identifying opinions and beliefs about products (Fan & Gordon, 2014; Lazer et al., 2009). Particularly, online consumer reviews, widely considered a rich data source that reflects consumer experiences and evaluation of products, have been studied to understand a range of research problems in hospitality and tourism (Schuckert, Liu, & Law, 2015b). Studies using online reviews usually employ a sample of review (and related) data, large or small, to extract features or measures that allow the researcher to detect, describe or predict patterns that are meaningful from theoretical or practical perspectives. This literature complements conventional approaches that primarily rely on surveys, personal interviews and other communication-based methods and represents a promising research direction by taking advantage of the abundant, readily available data resources (Xiang, Schwartz, Gerdes, & Uysal, 2015). While this line of research has generated novel insights into hospitality and tourism management, existing studies are limited in that 1) they tend to use a single data source for online reviews and 2) the quality of data is

* Corresponding author.
E-mail addresses: philxz@vt.edu (Z. Xiang), qiand12@vt.edu (Q. Du), yufengma@vt.edu (Y. Ma), wfan@vt.edu (W. Fan).

http://dx.doi.org/10.1016/j.tourman.2016.10.001
0261-5177/

largely anecdotal and often based upon the popularity of the websites from which the data were collected, which substantially limits their generalizability and contribution to knowledge.

With this in mind, this study comparatively examines three major online review platforms, namely TripAdvisor, Expedia, and Yelp, in terms of information quality related to online reviews in these websites with the goal to provide a basis for understanding the methodological challenges and for identifying opportunities for the development of social media analytics in hospitality and tourism. The rest of the paper is organized as follows: the next section, Research Background, reviews related literature to provide the motivations for the present study. In Research Design and Analytical Framework, we outline our methodological approaches and describe key measures and methods used to assess the three platforms with specific research questions. In Data Collection and Analysis, we describe the data collection process and explain, in details, the text analytics procedures to develop key metrics to describe review characteristics as well as statistical analyses conducted to compare and contrast the three platforms. Then, research findings are presented followed by a discussion on implications for both research and practice. Finally, conclusions are drawn, and limitations and directions for future research are discussed.

2. Research background

Big data analytics has been touted as a new research paradigm that utilizes diverse sources of data and analytical tools to make inferences and predictions about reality (Boyd & Crawford, 2012; Mayer-Schönberger & Cukier, 2013). Particularly, with increasingly powerful natural language processing and machine learning capabilities, textual contents from the Web provide a huge shared cognitive and cultural context and, thus, have been analyzed in many application domains (Halevy, Norvig, & Pereira, 2009). However, in recent years there have been growing criticism and concerns about the data-driven approach especially those using online user-generated contents as research data. For example, Ekbia et al. (2015) discuss some of the epistemological dilemmas in existing big data analytics, including the validity of claims about causal relationships, as opposed to mere statistical correlations, within the data. Others (e.g., Frické, 2015) challenge the nature of inductive reasoning in big data analytics and suggest that there are potential hazards in making generalizable claims. Particularly, Ruths and Pfeffer (2014) argue that studies using social media data should be aware of a number of validity problems such as platform biases (e.g., platform design, user base, and platform specific behavior), data availability biases, and data authenticity issues. Tufekci (2014) specifically highlights the conceptual and methodological challenges in social media studies, particularly sampling biases arising from using a single platform as data source due to the sociocultural complexity of user behavior and unrepresentative sampling frame, which may complicate the interpretation of research findings. Importantly, she argues that social media platforms are comparable to specimens in biological research wherein they are selected for certain characteristics suitable for laboratory examinations at the expense of illuminating other potentially important features. As such, Frické (2015) suggests that correlations found in many of existing big data studies might only be considered "candidate" solutions to the problems at hand. Ruths and Pfeffer (2014), among others, call for the use of a variety of triangulation approaches for big data analytics such as applying the same methods to examine the performance on two or more distinct data sets when studying a new social phenomenon.

The online eco-system in hospitality and tourism is vast, complex, and diverse; so are online review platforms, which range from community-based sites such as LonelyPlanet, Tripadvisor and Yelp to transaction-based online travel agencies (aka OTAs) like Expedia and Bookings.com where reviews are incorporated as electronic word-of-mouth (Gligorijevic, 2016). Although they may all be considered part of social media with the common goal to assist consumer decision making by providing trusted, shared social knowledge, these platforms are complex sociocultural and economic systems that reflect different business models, different technological affordances, and different user segments/bases and power distribution in the online eco-system (Jeacle & Carter, 2011; Scott & Orlikowski, 2012). For example, TripAdvisor, by incorporating a variety of user data and information tools, represents various actors, resources and, importantly, business models through its website (Yoo, Sigala, & Gretzel, 2016). Recent market dynamics such as Expedia's takeover of Travelocity and Orbitz have created a new power structure within the online eco-system with the emergence of a potentially dominant social knowledge base (see http://time.com/money/3707551/expedia-orbitz-impact-travelers). From the business viewpoint, online reviews including their peripheral cues, such as user-supplied photos and the reviewer's personal information, are intended as means of persuasive communication in order to build credibility and influence user behavior (Sparks, Perkins, & Buckley, 2013; Zhang, Zhang, & Yang, 2016). Therefore, the selection, ranking, and display of online reviews reflect the platform's strategy to maximize these effects on its targeted audience. Also, the contribution of online reviews is a self-selection process, which contributes to quality differentiation over the life cycle of product reviews (Li & Hitt, 2008). Mkono and Tribe (2016), in a recent study, show that users on travel-related social media sites not only are product evaluators but also may play additional, important roles such as online activist, troll, social critic, information seeker and socialite. Furthermore, these issues are confounded with the long-standing concerns about the authenticity of online reviews (Luca & Zervas, 2015). Therefore, social media research using online review data must be cognizant of the nuances in these data sources in order to make conscious, appropriate methodological decisions when considering the representativeness and quality of the data.

Hospitality and tourism appears to be an ideal application field of social media analytics with tremendous growth and potential. For example, a recent study (Lu & Stepchenkova, 2015) cited over 100 papers primarily focusing on user-generated contents published in hospitality and tourism journals in the previous 10 years. Schuckert et al. (2015b) cited 50 articles related to online reviews published within and outside the field of hospitality and tourism, indicating the growing interest in understanding the impact of online reviews. For the present study, we examined the literature that specifically used online reviews as data in hospitality and tourism. Table 1 lists a sample of recent publications in six leading tourism and hospitality journals, namely *Tourism Management*, *Annals of Tourism Research*, *Journal of Travel Research*, *International Journal of Hospitality Management*, *Cornell Hospitality Quarterly*, and *Journal of Tourism and Hospitality Research*. While there is a considerable amount of publications elsewhere within and outside hospitality and tourism and, therefore, this compilation by no means represents a full picture of the literature, these journals were selected as the sampling frame because of their high influence in the field (McKercher, Law, & Lam, 2006). As shown in the table, these studies all collected and analyzed online reviews (and associated information) to address a variety of research problems such as travel motivation (e.g., Pearce & Wu, 2015), opinions and sentiments related to hospitality products (e.g., Crotts, Mason, & Davis, 2009; Levy, Duan, & Boo, 2013; Xiang et al., 2015), impact of online reviews on hotel business performance (e.g., Melián-González, Bulchand-Gidumal, & López-Valcárcel, 2013; Xie, Zhang, & Zhang, 2014; Ye, Law, & Gu, 2009), and the nature and utilities of online reviews as data (e.g., Fang, Ye,

Table 1
Sample of recent literature on analytics using online reviews in hospitality and tourism.

No.	Data Source	Data and Sampling	Study Variables & Analysis	Product	Reference
1	Booking.com	All available 1440 Spanish coastal hotels that have 186 K reviews.	Examines the distribution of scoring scale in Booking.com.	Hotels	Mellinas, María-Dolores, and García (2015)
2	Ctrip	All available 3.6 K reviews (between 02/2007-01/2008) for 248 hotels in three cities in China.	Uses a log-linear regression model to examine the influence of online reviews (volume and rating) on the number of hotel bookings.	Hotels	Ye et al. (2009)
3	Daodao.com	44 K online reviews covering 774 star-rated hotels in Beijing, China (using avail. price info as filter).	Uses linear regression to examine the influence of price on customers' perceptions of service quality and value.	Hotels	Ye, Li, Wang, and Law (2014)
4	Dianping	Reviews (number not specified) of 1242 restaurants containing overall ratings in Beijing, China.	Uses an econometric model with thematic analysis, sentiment, and volume to examine differences between consumer-generated reviews and expert reviews.	Restaurants	Zhang, Ye, Law, and Li (2010)
5	Expedia	61 K consumer ratings and comments of 11 K hotels in the U.S (approx. 1/3 of population).	Uses factor analysis based upon lexicon and linear regression to identify guest experience-related factors that influence hotel satisfaction.	Hotels	Xiang et al. (2015)
6	Flyertalk	1.5 K comments from members of five major hotel programs using theoretical sampling.	Uses content analysis to identify communication-based core categories that influence hotel communication programs.	Hotels	Berezan, Raab, Tanford, and Kim (2015)
7	London-eating.co.uk	2.5 K customer comments for 300 restaurants in London, UK.	Uses content analysis to identify salient factors that influence restaurant evaluation.	Restaurants	Pantelidis (2010)
8	Qunar	33 K expert reviews and 36 K customer reviews (covering all hotel classes) across 197 cities in China.	Uses Bayesian logit model to examine relationship between volume of expert reviews and rating	Hotels	Zhang et al. (2016)
9	Toprural	4.3 K establishments (60% of population) and 30 K messages (all available).	Uses regression analysis to examine relationships between price and advertising expenditures and review rating, volume and performance.	Rural lodging	Nieto, Hernández-Maestro, and Muñoz-Gallego (2014)
10	TripAdvisor	>= 16 K words of reviews were collected for each property (more than 400 words per review) on three 5-star hotels.	Uses combination of content analysis and linguistic analysis to identify features of hotel product in travel reviews that lead to customer delight for hotels in a competitive set.	Hotels	Crotts et al. (2009)
11	TripAdvisor	373 reviews (keyword sampling) related to Costa Rica lodging.	Uses exploratory content analysis and linear regression to identify factors that influence ecotourists' satisfaction with ecolodge stays.	Ecolodges	Lu and Stepchenkova (2012)
12	TripAdvisor	26 K hotel reviews of 17 K hotels in 249 tourist areas in Europe (min. 10 hotels/area and min 10 reviews/hotel as cutoff).	Examines relationship between valence and volume of online reviews.	Hotels	Melián-González et al. (2013)
13	TripAdvisor	5 K reviews of 843 individual hotels in five major hotel markets in Texas, USA.	Examines relationships between sentiment, rating, volume and variation of reviews and hotel performance.	Hotels	Xie et al. (2014)
14	TripAdvisor	42 K ratings related to online reviews on 185 hotels in Hong Kong.	Examines the relationships between overall rating and specific ratings to detect fake reviews.	Hotels	Schuckert, Liu, and Law (2015a)
15	TripAdvisor	350 reviews about a specific entertainment performance at an attraction site in China.	Uses exploratory content analysis to examine factors that influence tourists' satisfaction at performance-based tourist attractions.	Attractions	Pearce and Wu (2015)
16	TripAdvisor	20 K reviews for 106 attractions in New Orleans, USA (min. 3 review/site as cutoff)	Uses readability, reviewer characteristics, rating, and usefulness related to reviews in econometric models to examine what drives perceived helpfulness of reviews.	Attractions	Fang et al. (2016)
17	TripAdvisor	40 K ratings for hotels randomly selected from 20 destinations (min. 100 ratings)	Examines relationships between geographic region, user information (e.g., star value, textual description and date of posting) and rating patterns.	Hotels	Banerjee and Chua (2016)
18	Yelp	35 restaurants in London with 2.5 K reviews and 10 restaurants in New York with 2.6 K reviews	Uses reviewer characteristics, rating, length, readability, and perceived enjoyment in a regression model to understand what leads to perceived usefulness of reviews.	Restaurants	Liu and Park (2015)
19	Yelp	35 restaurants in London with 2.K reviews and 10 restaurants in New York with 2.6 K reviews	Uses a count model to examine relationship between sentiment and rating (findings show the asymmetric relationship)	Restaurants	Park and Nicolau (2015)
20	69 sources provided by a third-party website	235 Swiss hotels (12% of population) for the period 2008–2010 with 60 K reviews.	Uses artificial neural network model with ten input variables to investigate the relationships among user generated online reviews, hotel characteristics, and Revpar.	Hotels	Phillips, Zigan, Silva, and Schegg (2015)
21	TripAdvisor and 3 other sites	543 travel reviews by consumers with mobility challenges using keywords sampling methods.	Uses mixed methods (including quantitative and qualitative content analysis) to identify lodging attributes that influence satisfaction.	Rural lodging	Zhang and Cole (2016)
22	TripAdvisor and 5 other sites	1.9 K one-star reviews from 10 review websites as well as 225 management responses from 86 hotels in Washington, DC.	Uses content analysis to examine issues related to hotel complaints, which were then analyzed by hotel characteristics, including chain-scale segments, and reviewer characteristics, including purpose of travel and geographic location	Hotels	Levy et al. (2013)

Note: List is ordered alphabetically by name of data source. If a source was used in multiple studies, the studies are chronologically ordered. Three studies using multiple data sources (No. 20–22) are listed at the end of table.

Kucukusta, & Law, 2016; Liu & Park, 2015; Park & Nicolau, 2015; Zhang et al., 2016). Although they are different from each other in terms of study purpose and methodology, these studies fit into the general definition of social media analytics, which is basically concerned with using analytical tools and frameworks to collect, analyze, summarize, and interpret social media data to extract useful patterns and insights (Fan & Gordon, 2014).

In these studies, data were obtained from a variety of sources including the dominant social media sites such as TripAdvisor, Daodao (TripAdvisor's Chinese outlet) and Yelp, OTA sites such as Expedia and Ctrip, and specialty sites such as Flyertalk and Toprural. Overall, this sample of literature reflects the growing breadth and diversity of social media analytics research in hospitality and tourism; also, there are several observations specifically relevant to the present study: First, in terms of sample size these studies range from a few hundred to a few hundred thousand reviews or other features and, thus, they may not all be considered big data research. Nonetheless, these studies are primarily data-driven with the aim to identify novel patterns in the data in order to develop generalizable understandings about the phenomenon at hand. Second, there are a variety of measures collected and derived from online reviews including length of review, lexicon (words), topics (themes), sentiment (valence), rating, readability, usefulness, and peripheral features such as reviewer identity and characteristics. These measures cover linguistic characteristics and content of online reviews as well as variables that are associated with (e.g., rating), or impacted by (e.g., helpfulness), online reviews. Third, in terms of analysis a variety of analytical methods were used including content analysis, text mining, machine learning (e.g., artificial neural network analysis), multivariate regression, econometric modeling, or various combinations of these techniques.

Finally, and perhaps most importantly, most of these studies utilized only a single source of data usually based upon the popularity of the website. For example, several studies used TripAdvisor, which appears to be the "premier" sampling field, by citing the fact that it is the largest travel-related review site in the world (e.g., Banerjee & Chua, 2016; Pearce & Wu, 2015; Xie et al., 2014). Other sources were adopted oftentimes for similar reasons or based upon the popularity of the website. For example, in the Xiang et al. study (2015) Expedia data were used because the company requires reviewers to make at least one transaction through its website before being allowed to contribute a review to the website, which, presumably, prevents hospitality businesses or marketers to post inauthentic reviews. In terms of sampling methods some studies (e.g., Crotts et al., 2009; Fang et al., 2016; Ye et al., 2014) adopted certain rules of thumbs (e.g., using a minimum length of reviews or a threshold number of reviews per case), while others (e.g., Mellinas et al., 2015; Phillips et al., 2015; Ye et al., 2009) used all available data. In a few rare cases involving multiple sources (e.g., Levy et al., 2013; Ye et al., 2014; Zhang & Cole, 2016), data were aggregated and then analyzed without assessment of the potentially unique contributions from each of these sources. It is also noticeable that there seems to be a growing interest in understanding review data in terms of its representativeness. For example, in a research note Mellinas et al. (2015) suggested that the way the scoring scale in Booking.com is displayed could be misleading and researchers must treat this kind of data with caution. Using a Yelp data set Park and Nicolau (2015) showed that there exists an asymmetric relationship between sentiment and rating. However, in general there has been little discussion and evaluation of the quality of review data, especially how these data truly reflect the hospitality and tourism industry as well as consumer experience, which we believe could substantially limit the generalizability of social media analytics research in hospitality and tourism.

3. Research design and analytical framework

In order to understand the methodological challenges related to data quality in social media analytics, we devised a study to assess information quality related to online reviews on three major platforms, namely TripAdvisor, Expedia, and Yelp. The rationale for selecting these platforms was three-fold: 1) they are widely used by online consumers (Gligorijevic, 2016; Yoo et al., 2016); 2) each of them represents a fairly unique "species" of review platforms (i.e., TripAdvisor as the largest virtual travel community, Expedia the largest OTA, and Yelp the largest online community for small, local businesses); and, 3) they have been frequently used as primary data sources in academic literature both within and outside the hospitality and tourism field. We chose to compare all existing customer reviews and other peripheral information extracted from these platforms for the entire population of hotel properties in Manhattan, New York City in the US. The hotel sector was chosen as the study context because the product is fairly standardized at different service levels, which may allow us to observe discrepancies and nuances between these platforms. Manhattan was chosen because of its high number of hotel properties located in a relatively small geographical region with a wide variation in service levels (from budget to luxury) and service types (e.g., leisure vs. business).

To compare and contrast the information quality of these platforms, we applied an analytical framework with the focus on a set of review-related measures (see Fig. 1). The concept of information quality has been defined in different ways in the information systems literature (e.g., DeLone & McLean, 1992). Recently, within the eWOM contexts it has been operationalized as the quality of review content, linguistic characteristics, and peripheral cues that represent relevancy, sufficiency, currency, consistency, credibility and usefulness (see Filieri, Alguezaui, & McLeay, 2015 for a comprehensive review). We followed this general schema to include a set of key measures widely used in recent social media studies, particularly the text analytics literature (e.g., Abrahams, Fan, Wang, Zhang, & Jiao, 2015; Korfiatis, Garcia-Bariocanal, & Sanchez-Alonso, 2012; Mudambi & Schuff, 2010; Wang, Liu, & Fan, 2011). As can be seen in Fig. 1, a review can be seen as consisting of four basic components including linguistic features, semantic features, sentiment, and its source (the reviewer information). Linguistic features refer to characteristics related to the review textual content such as appropriate amount of data, ease of understanding, timeliness, relevancy, and completeness, etc., which are used to measure argument quality in the information science and

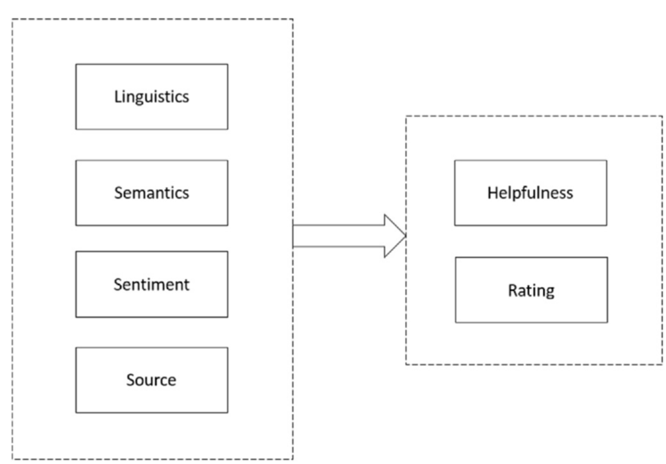

Fig. 1. Analytical Framework to Assess Information Quality of Online Reviews. (Note: the purpose of this framework was not to validate the theoretical structure; rather, it was intended primarily for understanding the commonalities and differences among the three platforms).

communications literature (e.g., Bailey & Pearson, 1983). Within the context of the present study, we considered review length (i.e., word count) and readability as key measures of linguistic features of a review (e.g., Fang et al., 2016). Another measure of information quality is semantic features, i.e., words, topics and semantic relationships between linguistic entities, which in the present study were operationalized as linguistic entities (tokens) and their latent dimensions (topics). Sentiment, which measures the valence (positive/negative) of an opinion, is an important feature widely used in text analytics (Pang & Lee, 2008). Review source represents the credibility of the information provider (Brinol & Petty, 2009) and has been adopted in social media studies to measure its impact on review user perception and behavior (e.g., Filieri et al., 2015; Sparks et al., 2013). Finally, rating and helpfulness have been extensively studied in the social media literature. Rating is the review provider's overall, numeric evaluation of the product and actual experience, which reflects the level of satisfaction with the product (Park & Nicolau, 2015; Xiang et al., 2015). Review helpfulness seems to be the “ultimate” measure of review quality since it represents a direct response by users who have read the review (e.g., Fang et al., 2016; Liu & Park, 2015; Mudambi & Schuff, 2010; Wang et al., 2011).

Within the study context, it was only possible to compare features that are commonly shared by all three sites, because, due to design differences, not all of the above-mentioned features were available on all of these platforms. For example, peripheral cues such as source credibility and reviewer-provided photos could be critical contributors to perceived review helpfulness; however, there are discrepancies in the provision and presentation of these types of information. For instance, Expedia does not provide detailed, trackable information about reviewers. In terms of users' response to reviews, Yelp offers two more options, namely “funny” and “cool”, besides “useful”. In this case, helpfulness (“useful” in Yelp's case) was used as the measure to compare these three platforms. With this in mind, we formulated the following research question to guide the analytical process:

Q_1. *Are there differences between TripAdvisor, Expedia, and Yelp in terms of review linguistic features, semantic features, sentiment, rating, and review helpfulness?*

As suggested by Ruths and Pfeffer (2014), the performance of theoretical relationships should be examined using multiple data sets in big data analytics. Therefore, in addition to individually comparing these measures of information quality, we also would like to assess the relationships between some of these key measures and rating and helpfulness in order to understand the intricacies between these variables and, particularly, whether and how the three platforms could be structurally similar or different. In recent text analytics literature, sentiment has been found to be highly associated with cues such as product rating (e.g., Park & Nicolau, 2015) and also a strong predictor of information quality (e.g., Jeong, Mankad, Gavirneni, & Verma, 2016). Rating (i.e., as indication of satisfaction) has been found to be a function of various characteristics of a review including latent dimensions revealed in the textual content of the reviews (Chua & Banerjee, 2015; Xiang et al., 2015). Schuckert et al. (2015a) showed that the consistency between overall rating and specific ratings on different product attributes can be used as a means to detect inauthentic reviews. As such, it was believed the relationships between rating and reviews' linguistic features, semantic features and sentiment are indications of internal consistency for information quality. For example, if a reviewer gave a hotel a rating of five while the review content was negative, it implies that either the review content or the rating was not a truthful reflection of his/her evaluation of the product. In addition, perceived helpfulness of reviews has also been found to be influenced by factors such as reviewer information (source), review content, linguistic features (e.g., length and readability), as well as sentiment (e.g., Chua & Banerjee, 2015; Fang et al., 2016; Korfiatis et al., 2012; Liu & Park, 2015; Mudambi & Schuff, 2010). Following this, a second set of research questions were formulated:

Q_{2a}. *Are there differences between TripAdvisor, Expedia, and Yelp in terms of relationships between rating and review linguistic features, semantic features, and sentiment?*

Q_{2b}. *Are there differences between TripAdvisor, Expedia, and Yelp in terms of relationships between review helpfulness and review linguistic features, semantic features and sentiment?*

4. Data collection and analysis

We applied the social media analytics procedure (e.g., Abrahams et al., 2015; Fan & Gordon, 2014) to answer the above research questions. We first collected relevant data from the three platforms. Then, the unstructured data were pre-processed and key metrics including online reviews' linguistic features, sentiment, semantic features, and perceived helpfulness were developed and compared among the platforms. Finally, a set of regression analyses were conducted to examine the relationships between these measures.

4.1. Data collection

Data collection took place in late 2015 on all searchable hotel properties in Manhattan, NYC in TripAdvisror, Expedia, and Yelp. Web crawlers written in the Python and Java programming languages were used to mimic a user's access to the system by specifying the travel destination and following all the links of hotel properties displayed as search results to download relevant information. Several types of data were collected including the name of the hotel property, its address, hotel class, all of its customer reviews, user responses (usefulness or helpfulness), and the overall rating. As shown in Table 2, in total, along with other types of information we collected approx. 439 k reviews from TripAdvisor, 481 k from Expedia, and 31 k from Yelp for a total of approx. 500 hotel properties (number of properties is platform specific). We used the Language Detection Package (i.e., langdetect) in Python (see https://www.python.org) to detect the languages that reviews were written in. After removing all non-English reviews, there were on average 991 reviews per property in TripAdvisor, 752 in Expedia, and 53 in Yelp. Although Expedia had the largest amount of reviews in total, TripAdvisor had the highest number of reviews per hotel property. Apparently, Yelp yielded a substantially smaller number of reviews compared to the other two. Reviews written in English served as the basis for the subsequent analysis on information quality.

4.2. Data analysis

4.2.1. Data pre-processing

All English language-based reviews collected from the three platforms were pre-processed using two basic procedures: tokenization and stop words removal. Tokenization is a form of lexical analysis whereby a stream of text is broken up into words, phrases, or other meaningful elements called tokens. In this study, each review was broken up into a vector of unigram-based tokens using a function called RegexTokenizer in Python's 'nltk.tokenize' Package. Stop words are words that do not contribute to the meanings of the text and are usually filtered out before the processing of natural language data. For this particular study, we applied an existing stop word list consisting of 429 English words (see http://www.lextek.com/manuals/onix/stopwords1.html), which has been widely applied in text mining and analytics. These two processes resulted in considerable reduction of the text corpus; that is, in terms of total token frequency there was approx. 40% left in both TripAdvisor and Yelp data sets while 43% in Expedia. Then, basic linguistic features

Table 2
Summary of the main data set.

Review Platform	N of Hotels	N of Reviews	N of English Reviews (percentage)	N of English Reviews per Hotel
TripAdvisor	443	438,890	438,826 (99.99%)	991
Expedia	467	480,589	351,182 (73.07%)	752
Yelp	581	30,816	30,770 (99.85%)	53

such as review length were computed for each review. We used the Python Package textstat0.2, which was developed based on the Flesch Reading Ease formula (Flesch, 1948) using word length and sentence length as core measure plus other weighting factors, to compute a readability score for each review. A higher score represents higher degree of readability and vice versa.

4.2.2. Development of key metrics

4.2.2.1. Review topics identification. Since the vector space represented by all the linguistic tokens was huge and difficult to describe and interpret, we used topic modeling to reduce the space to a manageable number of potentially meaningful dimensions (Griffiths & Steyvers, 2004). We applied the Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003) to the review text to discover the main topics related to consumers' experience and evaluation of hotel product. In a nutshell, The LDA model assumes that there exists a hidden structure consisting of a set of topics in the whole text corpus. The LDA algorithm uses the co-occurrence of observed words in different documents to infer this hidden structure. Mathematically, the model calculates the posterior distribution of the unobserved variables in the collection of documents. Given a set of training documents, LDA returns two main outputs. The first one is the list of topics associated with a set of words, which presumably contribute to this topic through their weights. The second output is a list of documents with a vector of weight values showing the probability of a document containing a specific topic. With this we identified the most salient topics within the review text and also computed and assigned topic scores to each review (i.e., to represent the likelihood of a topic to occur in a specific review). Since the review corpus contained potentially a large number of possible topics, we used the Elbow Method (Ketchen & Shook, 1996) to examine the perplexity values in order to determine the appropriate number of topics. Perplexity is commonly used in language modeling to test the fitness of a text model given a training corpus. A lower perplexity score indicates better generalization performance in new documents. Like in cluster analysis, the Elbow Method looks at the percentage of variance explained as a function of the number of topics. With the outputs of topic modeling we assigned each review a number of topic scores representing the review's likelihood of containing a specific topic.

4.2.2.2. Review sentiment identification. Sentiment analysis is a text mining procedure to discover emotive content in texts (Pang & Lee, 2008). To develop a sentiment score for each review, we started out by applying two existing lexicons of positive and negative word senses that had been developed in other product domains to the review corpus. However, the results were inconsistent and unreliable. Therefore, we decided to develop our own lexicon of positive and negative word senses that is presumably more suitable to the hospitality product. Following the sentiment detection procedure outlined in Abrahams et al. (2015), we first randomly selected 10,000 reviews as the training data set, out of which we further generated 2000 reviews based on polarized ratings. We then used domain experts to manually label these 2000 reviews into a positive and a negative set. Based upon this, we were able to generate a lexicon of "smoke" words which were then used as the sentiment classifiers to apply to the entire review corpus. After training the classifiers using the Naïve Bayes method and running the 10-fold cross validation (a typical performance test in sentiment analysis), the classification results seemed quite satisfactory with precision and recall rates for predicting both negative and positive reviews higher than 95%. Then, each review was assigned a sentiment score between 0 and 1 with 0 and 1 representing the two extremes of sentiment, i.e., negativity and positivity, respectively.

4.2.2.3. Review helpfulness score development. Review helpfulness is a key indicator of review quality; however, only a small portion of reviews had been rated as helpful on the three platforms. In order to compare the three platforms based on this key metric, we devised a machine learning procedure to "simulate" a helpfulness score for each review using the centroid-based summarization approach developed by Radev, Jing, Stys, and Tam (2004). First, all English-based reviews from the three platforms were merged together. Then, the top 3000 tokens with the highest frequencies were selected to form the shared dictionary and build the bag of words corpus. For each review, the 3000-token vector was transformed into the TF-IDF representation. TF-IDF, abbreviation for term frequency–inverse document frequency, is a numerical value used in text mining to reflect how important a word is to a document in a collection or corpus. Based upon this token-document TF-IDF matrix, reviews marked "helpful" were extracted to form the helpfulness corpus and its mean vector was calculated to represent the semantic centroid of review helpfulness. With this, cosine similarity between TF-IDF representation of each review and the centroid is computed and its value was assigned as the helpfulness score to each review (0–1).

4.2.3. Examine the relationships between key measures

Using the metrics extracted and developed with the methods described above, we ran two linear regression models to compare the three platforms. First, we assessed the relationships between rating and review characteristics including review topics and sentiment. The main rationale was that the correlations between these features should indicate the internal consistency of a review. The second regression analysis assessed the relationships between the "simulated" review helpfulness score and review characteristics including linguistic features, review topics, as well as sentiment. Both analyses were intended to observe the models' performances and also to detect any structural differences and inconsistencies across the three platforms.

Both regression analyses were run in the JMP statistical software by SAS (see http://www.jmp.com). Since some of the independent variables were constructed in different scales, we applied scale transformation to normalize their values. For example, review topic scores were log-transformed due to their low probability values. Also, multicollinearity was found among some of the variables after the initial runs, which made the models unstable and results difficult to interpret. We then conducted a series of centralization operations (i.e., obtaining a new score by using the difference between the original score and the mean). By doing so, we ensured the variance inflation factor (VIF) scores for all predictors in the regression models were below 10, which effectively removed the multicollinearity problem.

5. Findings

In this section, we first present the diagnostic analysis in terms of the extent to which the three platforms represent the hotel product from the supply side perspective. Then, we describe the characteristics of online reviews using the metrics we developed to measure information quality of these platforms. Finally, we present the results of regression analyses assessing the relationships between review characteristics and both rating and helpfulness.

5.1. Representation of the hotel industry

We examined the number of hotel properties in these websites in terms of hotel class and brand. Among them only Expedia has a star designation system (from one-star to five-star) for its listed properties; therefore, we applied Expedia's star rating system to TripAdvisor and Yelp by matching hotel names listed in Expedia. As can be seen in Table 3, there were huge discrepancies between the three platforms in terms of number and proportion of hotel properties in different service classes. Compared to Expedia, there were 54 fewer hotels in TripAdvisor and 182 in Yelp with a star rating, suggesting that there are substantial differences in listed properties on these platforms.

Table 4 lists hotel properties under major brands on these three platforms. As can be seen, TripAdvisor and Expedia were very similar with only one short in Expedia for the InterContinental Hotels Group (IHG) brands. Yelp apparently showed most discrepancies in that the total number of brand properties was substantially higher than the number of either Expedia or TripAdvisor. This might be caused by inaccurate hotel names in the system, which led to duplicates in the downloaded data set. From the consumers' standpoint, this could be a big obstacle when they are searching for hotel brands on the website. It is also noteworthy that Hilton brands were under-represented (as shown in percentage) in Yelp, while there was a considerable higher percentage of other brands, compared to the other two. This might be attributable to the Yelp business model as a platform for local, small businesses.

Table 3
Number of properties per hotel class on three platforms.

Class	TripAdvisor		Expedia		Yelp	
	N	Percent	N	Percent	N	Percent
One-star	6	1.5%	4	0.9%	2	0.7%
Two-star	49	12.3%	60	13.3%	44	16.3%
Three-star	143	35.9%	183	40.5%	92	34.1%
Four-star	163	41.0%	172	38.1%	110	40.7%
Five-star	37	9.3%	33	7.3%	22	8.1%
Total	398 (443)*	100%	452 (467)	100%	270 (581)	100%

*: Number in parenthesis indicates total number of hotel properties with or without star rating.

Table 4
Hotel brands on three platforms.

Brand	TripAdvisor		Expedia		Yelp	
	N	Percent	N	Percent	N	Percent
Hilton	34	25.4%	34	25.6%	45	19.9%
Marriott	33	24.6%	33	24.8%	56	24.8%
IHG	20	14.9%	19	14.3%	32	14.2%
Choice Hotels	12	9.0%	12	9.0%	21	9.3%
Starwood	10	7.5%	10	7.5%	20	8.8%
Wyndham	10	7.5%	10	7.5%	14	6.2%
Hyatt	8	6.0%	8	6.0%	11	4.9%
Others*	7	5.2%	7	5.3%	27	11.9%
Total	134	100%	133	100%	226	100%

*: Distributed among Best Western, Carlson, La Quinta, Vantage, Extended Stay.

5.2. Characteristics of online reviews

Table 5 lists a number of basic attributes of reviews on three platforms. The data set consisted of reviews dated back to as early as 15 years ago (the oldest reviews in the data set were created in late 2001 on TripAdvisor, late 2004 on Yelp, and 2005 on Expedia, respectively). On average, reviews in TripAdvisor and Yelp were considerably older than those in Expedia, which can likely be attributed to Expedia's relatively late incorporation of social media contents into its transaction-based business focus. Average length of reviews (measured by number of tokens after data preprocessing) was similar between TripAdvsior and Yelp and much higher than that of Expedia. In fact, data extracted from Expedia contained a considerable number ($N = 603$) of "empty" reviews (i.e., length $= 0$). Three platforms were similar in terms of review readability. Average rating was similar between TripAdvsior and Expedia while much lower in Yelp. Finally, average number of helpfulness responses per review was much higher on TripAdvisor and Yelp than Expedia. Overall, TripAdvisor and Yelp appeared to have richer information in relation to online reviews than Expedia, while Yelp seemed to be unique in that it is likely to attract consumers to voice their dissatisfaction or complaints.

5.2.1. Review length

We plotted the distributions of review length on the three platforms alongside each other in Fig. 2 with the X axis representing a specific length and Y axis its percentage of reviews on a specific platform. As can be seen, review length in Expedia was substantially skewed toward the shorter end. In fact, the vast majority (87.5%) of reviews in Expedia had a length of no greater than 50 words, while only 57.4% in TripAdvisor and 54.8% in Yelp were equal or shorter. The contrast was even more drastic in that approx. 61% of reviews in Expedia contained no more than 25 words, while only 21.9% in TripAdvisor and 24.1% in Yelp were equal or shorter. In general, TripAdvisor and Yelp appeared to be similar albeit the distribution was "narrower" surrounding the mode in TripAdvisor while more spread out in Yelp.

5.2.2. Review topics

Table 6 shows the major topics identified from the aggregated review corpus using the Latent Dirichlet Allocation (LDA) topic modeling method. Using the perplexity scores to assess the goodness-of-fit of topic modeling, we arrived at a solution with five topics for simplicity while maintaining meaningful interpretation of the semantic space. Based upon their meanings these topics were manually labelled as Basic Service, Value, Landmarks & Attractions, Dining & Experience, and Core Product. The numeric values to the right of the linguistic tokens represent the posterior probabilities, which indicate the likelihood for a given token to belong to a specific topic (to save space only the topic 10 tokens are presented here). Note these posterior probabilities are very small because they were calculated against a large number of potentially available dimensions represented by the bag of tokens in the review text. Conceptually, these topics represent the common "themes" that ran through the overall review corpus. Understandably, these topics are

Table 5
Summary of basic characteristics of reviews on three platforms.

	TripAdvisor	Expedia	Yelp
Avg. age of reviews (days)	1066	654	1076
Avg. length (N of tokens)	56.4	24.1	59.4
Review readability[a]	75.7	75.3	79.4
Avg. rating	4.11	4.08	3.32
Avg. N of helpfulness response per review	1.11	0.15	1.46

[a] Range 0–100; 0 means most difficult and 100 easiest.

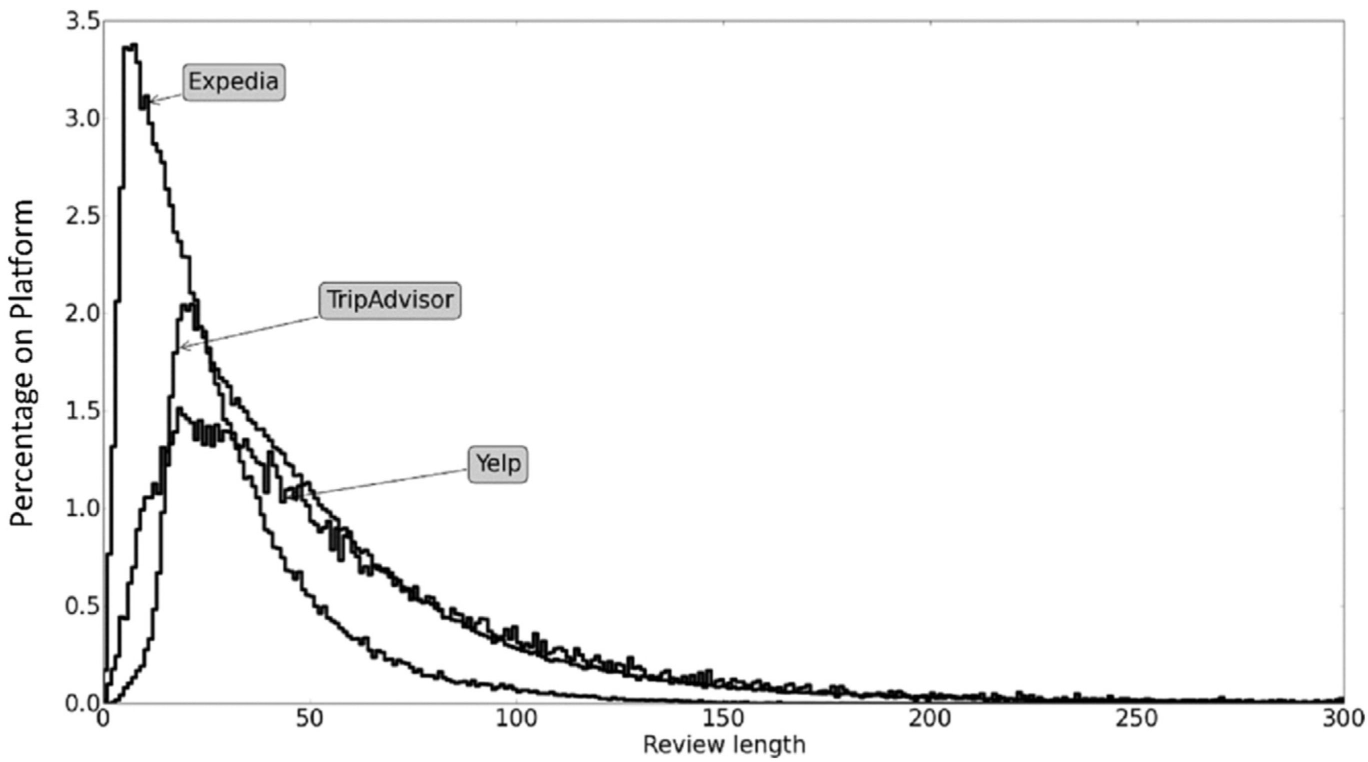

Fig. 2. Distribution of review length on three platforms.

destination specific in that they may not be the same if our data had been collected from a different city, particularly for Landmarks & Attractions (Topic 3) and Dining & Experience (Topic 4). Other topics appeared to be more generic, relevant to the hotel product.

To understand potential differences between the three platforms, we then examined the manifestation of these five topics on each of these websites. As shown in Fig. 3 (X axis represents the topics and Y the distribution of reviews on each platform as percentage), overall these five topics were fairly evenly distributed on three platforms. Topic 5, i.e., Core Product, manifested almost equally between the three websites, suggesting that hotel guest room and bathroom shared the same level of concern or mentioned equally in the reviews. However, other four topics manifested quite differently among these websites. Topic 1, i.e., Basic Service, was much more prominent in Yelp compared to TripAdvisor and Expedia. Topic 2, i.e., Value, appeared to be more prominent in Expedia than the other two, which seemed to fit with the transaction-based nature of Expedia. Topic 3, i.e., Landmarks & Attractions, was particularly less prominent on Yelp, which may be a reflection of its status as a platform for local businesses. Compared to TripAdvisor and Yelp, Expedia was lower on Topic 4, i.e., Dining & Experience. These different levels of manifestation of the common topics seemed to reflect these platforms' business orientations and user bases, and it seems that online reviews on these websites had different "flavors".

In order to understand whether, and the extent to which, these topics were related to how reviewers evaluate the product and their experiences, we plotted their distributions against satisfaction rating. As can be seen in Fig. 4, in reviews with rating of one (1 on the X-axis) the topic space was dominated by Basic Service (53.3%) followed by Core Product (21.6%), regardless of platform with a combined percentage of nearly 75%. In reviews with rating of two (2) there was almost the same pattern in that these two topics dominate the reviews with a combined percentage of 69.2%. This means that topics related to basic hotel services (i.e., front desk, staff, etc) and core product (the guest room and bathroom) were salient in reviews associated with lower ratings. On the other hand, topics related to value, landmarks and attractions, and dining and experience increased their share of the semantic space from reviews with lower ratings to those with higher ratings. Particularly noteworthy was that Topic 3 (Landmarks & Attractions) drastically increased from 5.8% (rating score of 1) to 18.2% (rating of 5) and

Table 6
Review Topics Identified using Latent Dirichlet Allocation (LDA).

Topic 1. Basic Service		Topic 2. Value		Topic 3. Landmarks & Attractions		Topic 4. Dining & Experience		Topic 5. Core Product	
desk	0.044	great	0.107	square	0.072	bar	0.030	room	0.081
front	0.040	location	0.080	times	0.060	view	0.018	free	0.027
room	0.040	staff	0.076	central	0.042	trip	0.018	bed	0.026
service	0.035	good	0.062	park	0.039	restaurant	0.017	small	0.022
air	0.011	breakfast	0.029	station	0.023	service	0.013	size	0.020
check	0.010	nice	0.021	building	0.019	experience	0.012	area	0.020
business	0.009	place	0.021	subway	0.018	visit	0.010	coffee	0.019
rate	0.008	excellent	0.018	empire	0.017	wonderful	0.010	nice	0.016
door	0.007	price	0.016	state	0.016	lovely	0.009	bathroom	0.015
customer	0.007	friendly	0.015	broadway	0.014	top	0.009	shower	0.014

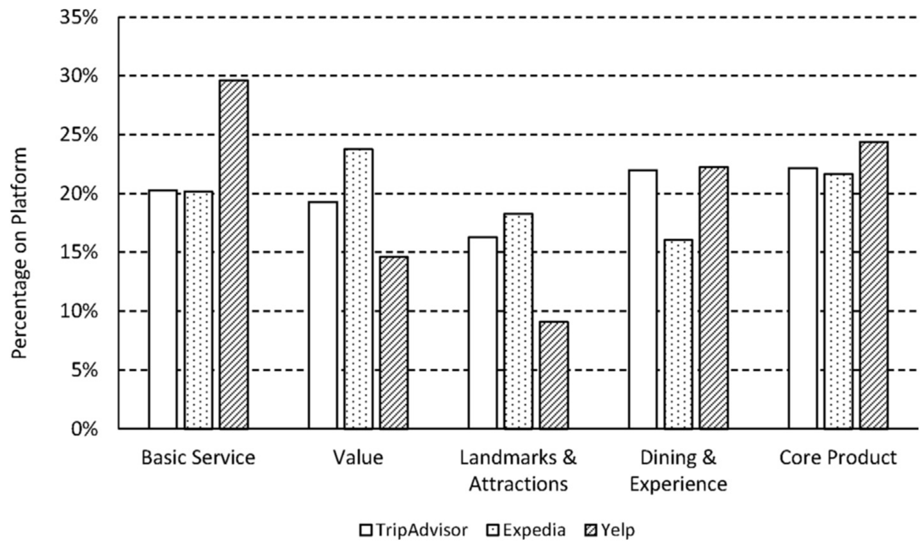

Fig. 3. Manifestation of five topics on three review platforms.

Topic 4 (Dining & Experience) from 11.0% (rating of 1) to 26.5% (rating of 5). This suggests that reviewers may have distinct mental models when writing reviews with either positive or negative sentiment. This also means that topics themselves are not neutral; that is, they implicitly reflect reviewers' evaluation of the product.

5.2.3. *Review sentiment*

Based upon the sentiment analysis, we computed and assigned a sentiment score to each review. Fig. 5 shows the distribution of reviews with their sentiment scores on three platforms (the X-axis represents the sentiment score, while the Y-axis indicates the percentage of reviews within a specific platform). As can be seen, to the right side of the X-axis are reviews with positive sentiment scores (<=1), while the negative reviews are aligned toward the left side (with 0 the lowest). On average, three platforms had an overall sentiment score of 0.68; however, Yelp was considerably lower (with a mean of 0.52) than the other two (TripAdvisor had a mean of 0.70 and Expedia 0.66). More interestingly, this graph shows almost an identical pattern in Expedia and TripAdvisor in that both of their distributions are to a great extent skewed toward the positive side. This indicates overall reviews in these two websites tend to have positive sentiments. Among these three, Yelp was unique with a near "saddle"-shaped distribution, which suggests that its reviewers were more polarized than the other two.

5.2.4. *Review helpfulness*

Based upon the semantic centroid identified using existing reviews that were rated helpful, or useful in the case of Yelp, we calculated a helpfulness score for each review by measuring the cosine similarity between the review's semantic space and the centroid. Fig. 6 shows the distribution of helpfulness scores of all English reviews on the three platforms. The X-axis represents the value of cosine similarity: the higher the value is, the more helpful a review would be rated. The Y-axis represents the percentage of reviews of a certain helpfulness score on a specific platform. As can be seen, the distribution exhibited similar normality on three platforms. However, their medians are different with Expedia the lowest at 0.14 (represented by approx. 1% of all of its reviews). Compared to the Expedia median, there were nearly two-thirds (65.6%) of reviews on Yelp and three quarters (74.5%) on TripAdvisor with a higher cosine similarity score, which suggests that there was considerable amount of reviews on these two platforms that would be rated helpful than Expedia.

5.3. *Results of regression analyses*

5.3.1. *Rating and review characteristics*

Table 7 shows the results of multivariate linear regression analysis to examine the relationships between rating and review

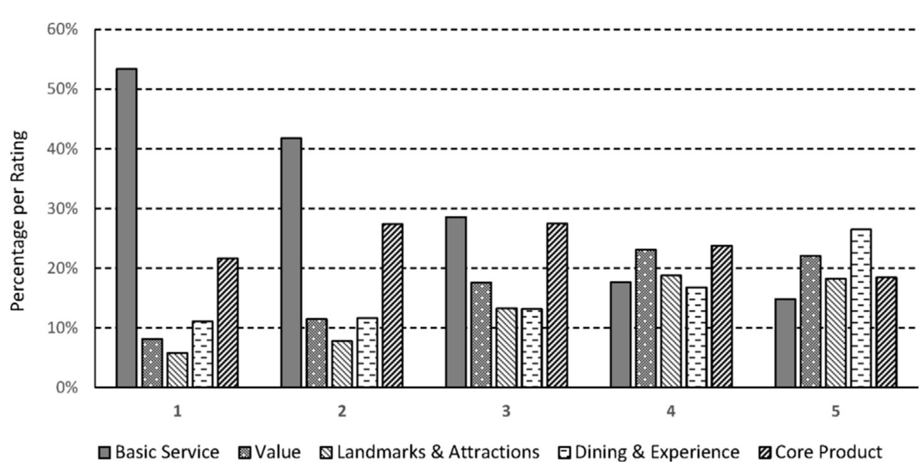

Fig. 4. Distribution of five topics plotted Against rating (overall).

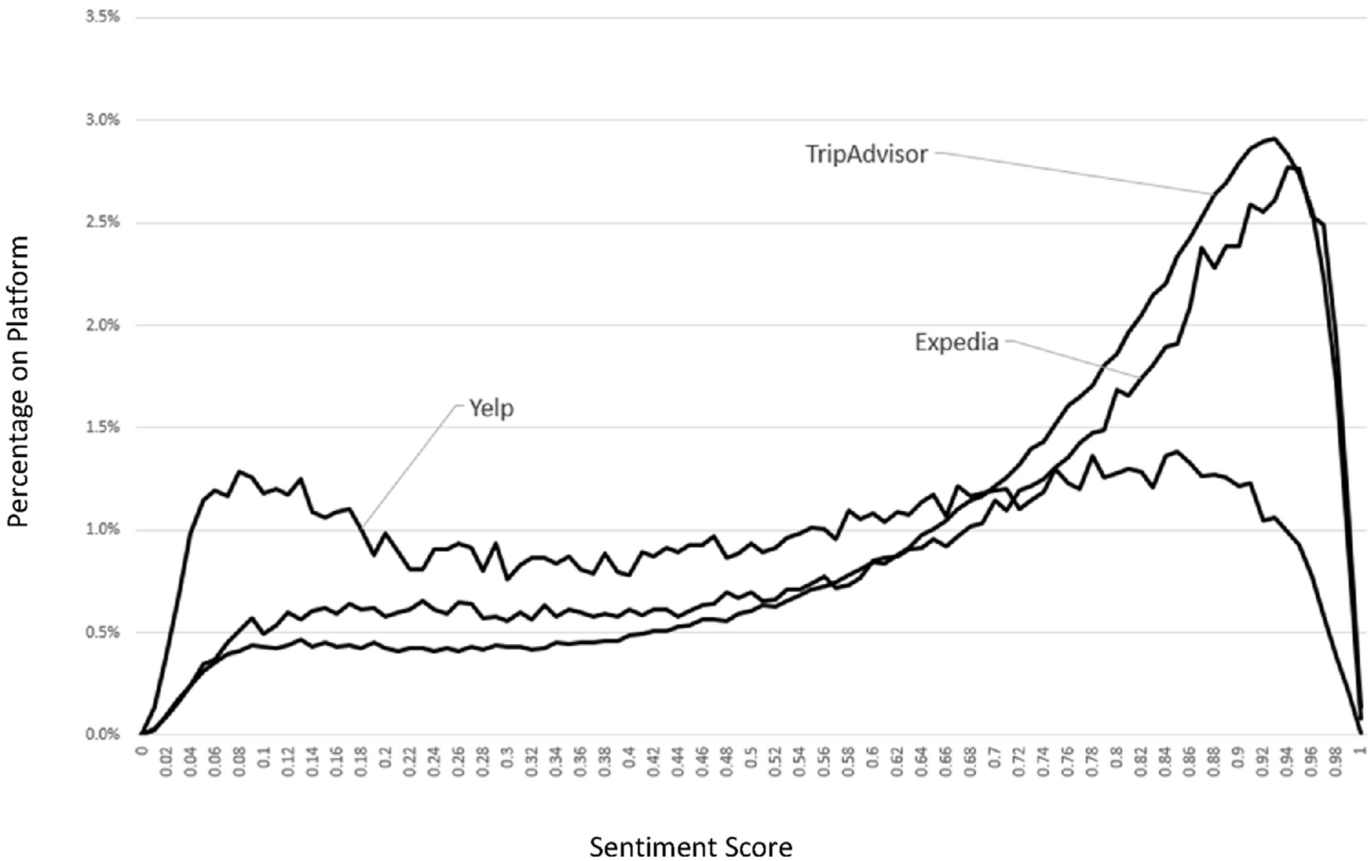

Fig. 5. Distribution of sentiment among all reviews on three review platforms.

characteristics including review topics and sentiment. The goal was to compare and contrast the three platforms in terms of the extent to which these characteristics contribute to rating as a measure of internal consistency. To do so, we added review features, one set at a time, to see the differences in their effects in explaining the variance in rating. To see the differences between the three platforms, we tested an overall model with a combined data set as well as an individual model based upon the data set for each of the platforms. We first introduced a couple of control variables, i.e., hotel class and brand, to see if the industry structure can be used to explain rating (Model 1). As can be seen, both hotel class and brand were significant in a positive way, suggesting that 1) the higher a hotel's class is, the more likely it has a higher rating; and, 2) whether a hotel is branded or not makes a difference. However, their overall contribution to rating was small with a combined Adjusted R Square of 0.036 (in the overall model).

Model 2 examined the contribution of review topics (measured as the likelihood of a specific topic to be contained in a review) to rating. The Adjusted R Square increased considerably (to 0.24 in the overall model). All topics were significant, while T1 (Basic Service) and T5 (Core Product) were negative and T2 (Value), T3 (Landmarks & Attractions), and T4 (Dining & Experience) all positive. While, given the large sample size, the level of significance was not surprising, these signs were quite revealing because they seemed to reflect the negative connotations of T1 and T5 (as also noted in Fig. 4). In Model 3 review sentiment was shown to be a strong predictor of rating in that its introduction more than doubled the Adjusted R Square in the overall model. We also examined the interaction effect between sentiment and topics on rating, which moderately improved the explanatory power. Among the interaction terms Sentiment * Basic Service appeared to be the strongest predictor (as shown in the size of its coefficient). It should be noted that some of the signs of coefficients for the topics flipped due to the introduction of dominant variables such as sentiment into the model. It is noteworthy that, besides review sentiment, review topics are also strong predictor of rating. This is consistent with the findings shown in Fig. 4; that is, words related to basic service and core products are not neutral.

While it is not surprising to see these review characteristics explain a large amount of variance in rating (Adjusted R Square = 0.56 in the overall model), interestingly, these three platforms yielded different performances. In terms of explanatory power Yelp appeared to the strongest followed by TripAdvisor and then Expedia. The relatively large coefficients of variables such as Sentiment (3.44), Sentiment * Basic Service (4.04), and Sentiment * Dining & Experience (1.29) in the Yelp model in Model 4, seemed to contribute to the explanatory power of its model. Most of the coefficients in the Expedia model were smaller than the other two except in the variable Sentiment * Value, which seemed to be consistent with the findings shown in Fig. 3 (i.e., Expedia has the flavor of “value”). Compared to the TripAdvisor model, the Expedia model had considerably lower explanatory power, which suggests that the topics contained in its reviews were not as consistent with the rating scores as TripAdvisor, although both platforms had almost identical patterns in review sentiment. In Model 2, TripAdvisor and Yelp were almost equal in Adjusted R Square and Yelp's performance improved considerably once sentiment was introduced (Model 3). This suggests that review topics were likely the variables that differentiated Expedia from the other two platforms, and review sentiment the variable that made Yelp unique compared to the other two.

5.3.2. *Review helpfulness and review characteristics*

Table 8 shows the results of the regression analysis examining

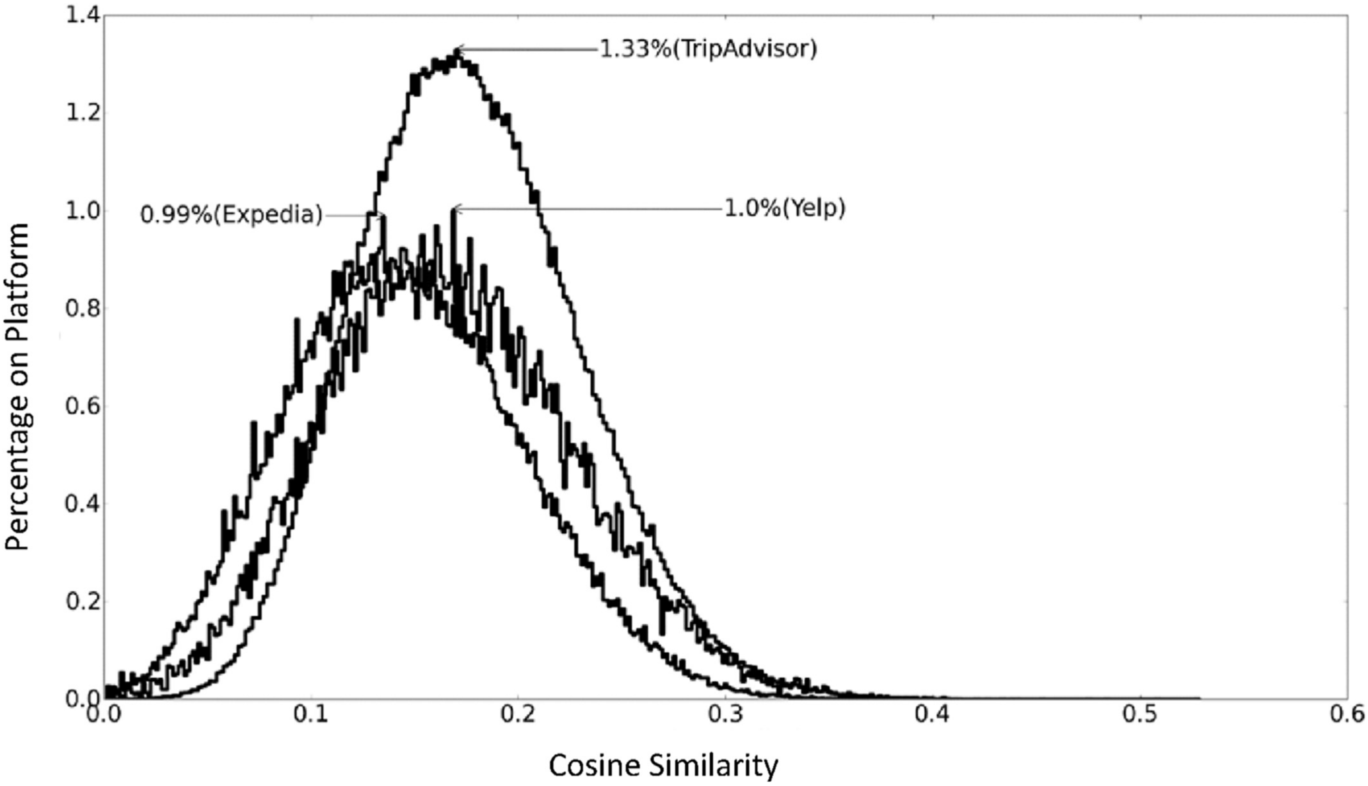

Fig. 6. Distribution of helpfulness score among all reviews on three review platforms.

the relationships between review helpfulness and review characteristics. Different from the analysis on rating, two new variables related to linguistic characteristics, i.e., review length and readability, were added (Models 6 and 7). As can be seen, although not substantial, both hotel class and brand contributed to review helpfulness. Among the linguistic features, review length seemed to a strong indicator of review helpfulness with an Adjusted R Square of 0.45 in the overall model (Model 6). By introducing topics into the regression model (Model 8) the explanatory power increased considerably to 0.54 in the overall model. Also noteworthy was the sign of the coefficients of these variables in that T1 (Basic Service), T4 (Dining & Experience), and T5 (Core Product) all seemed to be negatively correlated to helpfulness. Sentiment also seemed to be a strong predictor of helpfulness (Model 9), which explained over 10% more variance compared to Model 8, indicating the more positive the tone in the review, the more likely it will be perceived as helpful. The final model (Model 10), which included the interaction terms between sentiment and topics, reached an Adjusted R Square of 0.65 in the overall model. In general, the coefficients in these models were substantially smaller than those in the rating models because the dependent variable helpfulness was much smaller in scale (i.e., 0–1).

Table 7
Regression analysis examining relationships between rating and review characteristics.

	Model 1				Model 2				Model 3				Model 4			
	Overall	TPA	EXP	YLP	Overall	TPA	EXP	YLP	Overall	TPA	EXP	YLP	Overall	TPA	EXP	YLP
Class	0.22	0.23	0.22	0.04	0.17	0.18	0.20	0.02	0.09	0.09	0.13	0.003	0.09	0.09	0.14	0.004
Brand	0.05	0.03	0.09	0.05	0.12	0.11	0.10	0.16	0.06	0.06	0.07	**0.02****	0.05	0.04	0.07	0.003
T1					−1.95	−1.65	−1.96	−3.41	0.34	0.25	0.50	0.33	−0.41	−0.5	−0.31	**0.15****
T2					0.57	0.6	0.84	−0.15	0.007	−0.16	0.25	−0.11	−0.83	−0.9	−0.66	−0.96
T3					0.34	0.47	0.50	−0.49	−0.1	−0.23	**0.05***	−0.29	−0.96	−1.01	−0.89	−1.01
T4					0.85	1.03	0.83	**−0.12***	0.78	0.65	0.89	0.71	−0.25	−0.31	−0.13	**−0.18****
T5					−0.46	−0.21	−0.64	−0.98	0.38	0.23	0.48	0.62	−0.69	−0.78	−0.59	−0.45
Sentiment									3.1	3.22	2.78	3.77	2.89	2.95	2.69	3.44
Senti * T1													2.6	3.08	1.63	4.04
Senti * T2													−0.92	−0.42	−1.61	−0.45
Senti * T3													−0.32	−0.004	−0.52	**0.39****
Senti * T4													0.39	0.54	−0.02	1.29
Senti * T5													0.52	0.88	0.43	0.64
R^2	0.036	0.034	0.029	0.004	0.242	0.245	0.215	0.245	0.539	0.548	0.491	0.584	0.555	0.566	0.501	0.604
Adjusted R^2	0.036	0.034	0.029	0.004	0.242	0.245	0.215	0.245	0.539	0.548	0.491	0.584	0.555	0.566	0.501	0.604

Dependent variable: satisfaction rating (1–5).
TPA: TripAdvisor; EXP: Expedia; YLP: Yelp.
T1: Basic Service; T2: Value; T3: Landmarks/attractions; T4: Dining/experience; T5: Core product.
To save space, all un-bolded coefficients are significant at <0.001 level; otherwise, *: <0.1; **: <0.01.

Table 8
Regression analysis examining relationships between helpfulness and review characteristics.

	Model 5				Model 6				Model 7				Model 8
	Overall	TPA	EXP	YLP	Overall	TPA	EXP	YLP	Overall	TPA	EXP	YLP	Overall
Class	−0.000	−0.003	−0.001	−0.000	−0.000	−0.002	−0.001	0.000	**−0.000***	−0.002	−0.001	0.000	0.001
Brand	0.003	0.004	0.003	0.006	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.006	0.003
Length					0.050	0.050	0.050	0.050	0.050	0.05	0.050	0.050	0.060
Readability									−0.000	0.000	−0.001	**0.000****	−0.000
T1													−0.060
T2													0.007
T3													0.020
T4													−0.040
T5													−0.010
Sentiment													
Senti * T1													
Senti * T2													
Senti * T3													
Senti * T4													
Senti * T5													
R^2	0.001	0.003	0.001	0.002	0.453	0.466	0.291	0.499	0.453	0.466	0.293	0.499	0.537
Adjusted R^2	0.001	0.003	0.001	0.002	0.4526	0.466	0.291	0.499	0.453	0.466	0.293	0.499	0.537

Dependent variable: review helpfulness.
TPA: TripAdvisor; EXP: Expedia; YLP: Yelp.
T1: Basic Service; T2: Value; T3: Landmarks/attractions; T4: Dining/experience; T5: Core product.
To save space, all un-bolded coefficients are significant at <0.001 level; otherwise, *: <0.1; **: <0.01.
All coefficients are shown with three decimals due to their small values; also, a “0.000” indicates the coefficient is a minuscule, non-zero number.

Similar to the rating models, among the three platforms the Yelp data appeared to explain the highest amount of variance in the helpfulness score (with an Adjusted R Square of 0.69 in Model 10) followed by TripAdvisor (0.65) and Expedia (0.54). By examining the stepwise models, Model 6 seemed to suggest that review length was the strongest predictor in Yelp (0.50) compared to the other two. Review length was particularly weak in the Expedia model (0.29), which suggests that its variance in Expedia was not as strongly associated with helpfulness as in other websites. However, sentiment seemed to be a stronger predictor in Expedia because, by comparing Model 9 (which included sentiment) with Model 8, the increase of Adjusted R Squared in the Expedia model was almost 0.14, as opposed to roughly 0.10 in both TripAdvisor and Yelp. This seemed to suggest that, if review length was not perceived as indicator of helpfulness, a user may shift his/her attention to other review characteristics when evaluating the utilities of a review.

6. Discussion and implications

Motivated by the lack of understanding of data quality in social media-related studies, we applied a series of text analytics techniques to “dissect” three major review platforms in hospitality and tourism. This study shows that TripAdvisor, Expedia, and Yelp, while all incorporating consumer reviews as primary social knowledge, are indeed distinct from each other on a variety of aspects. In terms of representing the supply of the hotel product, there appears to be huge discrepancies between these platforms. In terms of sheer amount of review data, TripAdvisor and Expedia are comparable to each other, while Yelp is substantially smaller. In terms of topics contained in the review texts, each platform manifests certain “flavors”, which may reflect different trending themes, which, in turn, may reflect potentially different user bases on these platforms. In terms of overall sentiment, TripAdvisor and Expedia are similar while Yelp is quite unique with a more polarized distribution. Between TripAdvisor and Expedia, which are comparable in review volume, the former seems to have higher overall quality. This seems to explain why TripAdvisor has been widely perceived as a premier data source either based upon anecdotes or empirical evidence found elsewhere (e.g., Chua & Banerjee, 2013; Jeacle & Carter, 2011). Furthermore, in terms of review helpfulness TripAdvisor and Yelp appear to have more reviews that would potentially be seen as more helpful than Expedia. Besides, our analyses also revealed some important nuances that may reflect structural differences among these platforms. Specifically, the connections between rating and helpfulness and other review characteristics are not as strong in Expedia as the other two. In particular, review topics are not strong predictors for rating in Expedia data, which may suggest likely inconsistencies between what a reviewer writes and what satisfaction score he/she assigns to the product. Also, review length is not as strong a predictor of review helpfulness in Expedia than the other two, indicating the lack of quality due to the smaller variance in the amount of information contained in its reviews. Overall, Yelp seems to have the strongest performance in term of both rating and helpfulness, which may be attributed to its high variance in its review sentiment. Through these text analytics exercises and above findings, this study offers several important implications for both research and practice.

6.1. Implications for research

First and foremost, by showing the differences in major online review platforms this study contributes to the epistemology of social media analytics by suggesting that studies directly drawing data from online websites must, indeed, consider the inherent traits and potential biases in social media data (Ruths & Pfeffer, 2014; Tufekci, 2014). There are huge discrepancies in the representation of the hotel industry on these platforms. Online reviews could vary across different platforms in terms of linguistic characteristics, semantic features, sentiment as well as impact on users' of the websites. When sampling online review data, careful considerations must be given to the source of the data as well as representativeness of the data within a specific source. Our findings suggest that one of the platforms alone may not be a sufficient source for quality data because different platforms may possess fairly unique characteristics. For example, when the research relies upon a representative sample of review sentiment,

Model 8			Model 9				Model 10			
TPA	EXP	YLP	Overall	TPA	EXP	YLP	Overall	TPA	EXP	YLP
0.000	0.001	0.001	−0.001	−0.002	−0.001	0.000	−0.001	−0.002	−0.001	0.000
0.003	0.003	0.003	0.002	0.001	0.003	−0.000	0.002	0.001	0.003	−0.000
0.060	0.050	0.060	0.060	0.060	0.050	0.070	0.060	0.070	0.050	0.060
−0.000	−0.000	−0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
−0.050	−0.080	−0.050	0.010	0.010	0.006	0.030	0.030	0.030	0.040	0.030
0.010	0.004	0.020	0.005	0.008	−0.002	0.020	0.020	0.030	0.030	0.020
0.030	0.020	0.020	0.020	0.020	0.020	0.030	0.035	0.040	0.040	0.020
−0.030	−0.050	−0.050	−0.030	−0.030	−0.040	−0.030	−0.010	−0.007	−0.000	−0.030
−0.001	−0.030	−0.001	0.020	0.020	0.010	0.040	0.040	0.050	0.050	0.030
			0.090	0.090	0.090	0.080	0.100	0.100	0.090	0.090
							−0.080	−0.070	−0.110	−0.070
							−0.010	**0.003***	−0.020	−0.030
							0.006	0.010	0.020	−0.020
							−0.050	−0.050	−0.070	−0.070
							−0.060	−0.050	−0.070	−0.070
0.547	0.398	0.581	0.643	0.647	0.536	0.683	0.647	0.651	0.542	0.686
0.547	0.398	0.580	0.643	0.647	0.535	0.683	0.647	0.651	0.542	0.686

the distribution of sentiment must be closely examined due to the possible variations across platforms. Also, review contents may vary considerably by service level and other factors. As such, if the research focus is on identifying topics and opinions that reflect consumer experience and evaluation of the product, findings based upon single-source data should only be considered as candidate solutions (Ruths & Pfeffer, 2014). In this regard, this study offers insights into developing useful heuristics such as review length and rating to construct sampling rules in social media analytics research. While in a few recent studies review volume (e.g., Fang et al., 2016; Melián-González et al., 2013) and review length (e.g., Crotts et al., 2009) were used as "filters" for sampling purposes, the selection of data must be done in a more systematic way.

Second, there is a growing interest in understanding the communicative effects of online reviews within and outside hospitality and tourism (e.g., Filieri et al., 2015; Mudambi & Schuff, 2010; Racherla & Friske, 2012; Sparks et al., 2013). As an exploratory study, this paper sheds light on the theoretical development for assessing the utility and impact of eWOM in hospitality and tourism. On the one hand, the analytical framework used in our study shows considerably different explanatory power across platforms, which supports the claim by Ruths and Pfeffer (2014) and others that empirical models developed using big data may need to be confirmed with multiple data sources. On the other hand, compared to a number of recent analytics-based studies specifically measuring the consumer value of online reviews (e.g., Fang et al., 2016; Jeong et al., 2016; Liu & Park, 2015), this framework seems to be more effective in explaining why a review could be perceived as helpful (as indicated by the high overall R squares across different platforms). Obviously, this could be attributed to a number of reasons such as the comprehensiveness of the data set in terms of both size and variable inclusion as well as specific algorithms used in the text analytics exercises. Nonetheless, the analytical framework showed good explanatory power and, thus, can be expanded and further refined into a general theoretical framework for social media analytics based upon textual data. Furthermore, this framework seems to be able to capture inconsistencies between review content and the reviewer's satisfaction rating, suggesting possible directions for detecting inauthentic consumer evaluation of products, as demonstrated in Schuckert et al. (2015a) and others.

Third, this study reveals the intricacies between some of the variables associated with online reviews. For example, review length seems to be a strong indicator of helpfulness. More interestingly, the strong associations between review topics and both rating and helpfulness suggest that, within the product evaluation context, non-affective words themselves could have affective connotations similar to the effect of sentiment. This is consistent with the Xiang et al. (2015) study wherein topical "factors" generated from review texts were found highly correlated with rating. Also, this seems to reflect the findings of a recent study by Dodds et al. (2015) which shows that human language inherently possesses a universal positivity bias (i.e., language could be "colorful" in the emotive sense regardless whether it contains specific emotive words). However, whether this phenomenon is consistent within different product contexts and how it interplays with the overall sentiment of the review needs to be confirmed and could be an intriguing research question.

Finally, this study also sheds light on the structure of guest experience as revealed in online reviews and its connection with satisfaction. There is a growing literature on using online reviews to understand how various product and service aspects contribute to hotel guests' satisfaction (e.g., Crotts et al., 2009; Zhang & Cole, 2016). For example, in an exploratory study using Expedia review data Xiang et al. (2015) found that, in online hotel reviews, the mentions of "hygiene" factors such as room maintenance and cleanliness often take place in a negative context and act to prevent the guest from sharing any positive experience. In the present study, the association between topics related to basic service and core product (room and bathroom) and low ratings seems to confirm their findings across different platforms. This suggests that it is possible to identify, through user-generated content such as online reviews, meaningful structures among various aspects and attributes related to hospitality and tourism products. This is perhaps one of the promising research areas in social media analytics in hospitality and tourism.

6.2. Implications for practice

Although it was not the primary goal of this study to generate managerial insights, this study does offer plenty of implications for businesses. This study clearly shows that not all review websites are created equal; they vary considerably in quality and focus; they may represent different consumer segments; generally speaking, review websites come in different sizes and shapes. Therefore, these nuances in data quality must be taken into consideration when developing product and market intelligence for the firm using these data platforms. Hospitality and tourism businesses must make wise decisions when choosing these channels to engage with their existing and prospective customers, e.g., when providing feedback to positive/negative reviews (Sparks et al., 2013; Xie et al., 2014). Knowing that hygiene factors are usually the sources for complaints and dissatisfaction, businesses can more effectively monitor the social media space to look for clues for service failure and room for improvement. As persuasive communication tools, online reviews can be more effective when they are constructed with adequate amount of information, when they are rich in topics, and when they connote strong yet meaningful emotions consistent with rating and review content. Simply enforcing a house rule with the aim to filter out undesirable inputs (e.g., in the case of Expedia) might not be the best solution to ensure review quality. Therefore, businesses should develop effective means to encourage customers to share their experiences in rich and constructive ways.

7. Conclusions, limitations, and future research

Information technology creates new structures and dynamics in the market; therefore, it is imperative for us to gain a solid understanding of the changing reality, either from knowledge or business perspective (Werthner & Klein, 1999; Xiang, Wöber, & Fesenmaier, 2008). In this study we showed that online review data drawn from three dominant platforms on a specific industry sector and from a specific geographic region can be considerably different in both content and structure. By demonstrating the methodological challenges this study contributes to the emerging research areas related to big data and business analytics using social media data within and potentially beyond hospitality and tourism. Also, its findings reveal some of the intriguing interplays between information quality indicators which may well be considered important topics in the general research in the online product review and social media contexts.

While our overall research design is based upon a specific case (i.e., Manhattan) using only three representative platforms in the U.S., our approach is valid since the goal was to show incongruences within the general assumption that it does not matter from which website a researcher draws the data (and how). Nonetheless, our study has its limitations. For example, the data might be destination specific and thus the results, e.g., distribution of sentiment, rating and review helpfulness, must be interpreted with caution. Also, the text analytics tools we applied to the data have their own inherent limitations. For instance, the identified five topics in reviews reflect the most common themes in consumers' sharing of experience for the specific case. Apparently, the analysis did not capture the idiosyncrasies in hotel guest experiences, which could offer even more nuanced views of these data sources at a deeper level.

Nevertheless, this analysis helps us gain a better understanding of the big data nature of online reviews with the hospitality and tourism context. Building upon this study, future research can focus on developing sampling rules based upon heuristics such as review length and rating that can be used to establish research validity in social media analytics in more efficient ways. The analytical framework applied in this study can be expanded to include more theoretically meaningful variables (e.g., reviewer and user network information) and can be verified in different contexts in order to understand the structure of information quality in social media. Furthermore, future research can also explore the intriguing relationships between review content and sentiment especially in terms of how they may be used to understand the mental models of consumers sharing their experiences and product evaluation online. Therefore, this study points to several important research directions that will likely help develop the methodological and theoretical foundations for social media analytics in hospitality and tourism and beyond.

Acknowledgements

This study was partially supported by Natural Science Foundation of China Grant# 71531013.

References

Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management, 24*(6), 975–990.

Bailey, J. E., & Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science, 29*(5), 530–545.

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

Berezan, O., Raab, C., Tanford, S., & Kim, Y. S. (2015). Evaluating loyalty constructs among hotel reward program members using eWOM. *Journal of Hospitality & Tourism Research, 39*(2), 198–224.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679.

Brinol, P., & Petty, R. E. (2009). Persuasion: Insights from the self-validation hypothesis. In M. P. Zanna (Ed.), *Vol. 41. Advances in experimental social psychology* (p. 69e118). San Diego, CA: Academic Press.

Chua, A. Y., & Banerjee, S. (2013). Reliability of reviews on the Internet: The case of TripAdvisor. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1).

Chua, A. Y., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology, 66*(2), 354–362.

Crotts, J. C., Mason, P. R., & Davis, B. (2009). Measuring guest satisfaction and competitive position in the hospitality and tourism industry an application of stance-shift analysis to travel blog narratives. *Journal of Travel Research, 48*(2), 139–151.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60–95.

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., … Megerdoomian, K. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences, 112*(8), 2389–2394.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., … Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology, 66*(8), 1523–1545.

Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM, 57*(6), 74–81.

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

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

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221.

Frické, M. (2015). Big data and its epistemology. *Journal of the Association for Information Science and Technology, 66*(4), 651–661.

Gligorijevic, B. (2016). Review platforms in destinations and hospitality. In R. Egger, I. Gula, & D. Walcher (Eds.), *Open tourism* (pp. 215–228). Springer Berlin Heidelberg.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl. 1), 5228–5235.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems, 24*(2), 8–12.

Jeacle, I., & Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society* (Vol. 36,(4), 293–309.

Jeong, H., Mankad, S., Gavirneni, N., & Verma, R. (2016). What guests really think of your hotel: Text analytics of online customer reviews. *Cornell Hospitality Report, 16*(2), 3–17.

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–458.

Korfiatis, N., Garcia-Bariocanal, E., & Sanchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of reviews helpfulness vs. review content. *Electronic Commerce Research and Applications, 11*(3), 205–217.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., … Jebara, T. (2009). Life in the network: The coming age of computational social science. *Science (New York, NY, 323*(5915), 721.

Levy, S. E., Duan, W., & Boo, S. (2013). An analysis of one-star online reviews and responses in the Washington, DC, lodging market. *Cornell Hospitality Quarterly, 54*(1), 49–63.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research, 19*(4), 456–474.
Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management, 47*, 140–151.
Luca, M., & Zervas, G. (2015). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Harvard Business School NOM Unit Working Paper*, (14–006).
Lu, W., & Stepchenkova, S. (2012). Ecotourism experiences reported online: Classification of satisfaction attributes. *Tourism Management, 33*(3), 702–712.
Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management, 24*(2), 119–154.
Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
McKercher, B., Law, R., & Lam, T. (2006). Rating tourism and hospitality journals. *Tourism Management, 27*(6), 1235–1252.
Melián-González, S., Bulchand-Gidumal, J., & López-Valcárcel, B. G. (2013). Online customer reviews of hotels as participation increases, better evaluation is obtained. *Cornell Hospitality Quarterly, 54*(3), 274–283.
Mellinas, J. P., María-Dolores, S. M. M., & García, J. J. B. (2015). Booking. com: The unexpected scoring system. *Tourism Management, 49*, 72–74.
Mkono, M., & Tribe, J. (2016). Beyond reviewing: Uncovering the multiple roles of tourism social media users. *Journal of Travel Research*. http://dx.doi.org/0047287516636236. Article number: 0047287516636236.
Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS quarterly, 34*(1), 185–200.
Nieto, J., Hernández-Maestro, R. M., & Muñoz-Gallego, P. A. (2014). Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments. *Tourism Management, 45*, 115–123.
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.
Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly, 51*(4), 483–491.
Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research, 50*, 67–83.
Pearce, P. L., & Wu, M. Y. (2015). Entertaining international tourists: An empirical study of an iconic site in China. *Journal of Hospitality & Tourism Research*, 1096348015598202.
Phillips, P., Zigan, K., Silva, M. M. S., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of swiss hotel performance: A neural network analysis. *Tourism Management, 50*, 130–141.
Racherla, P., & Friske, W. (2012). Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications, 11*(6), 548–559.
Radev, D. R., Jing, H., Stys, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing & Management, 40*(6), 919–938.
Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063–1064.
Schuckert, M., Liu, X., & Law, R. (2015a). 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.
Schuckert, M., Liu, X., & Law, R. (2015b). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing, 32*(5), 608–621.
Scott, S. V., & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society, 37*(1), 26–40.
Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management, 39*, 1–9.
Tufekci, Z. (2014). *Big questions for social media big data: Representativeness, validity and other methodological pitfalls*. arXiv preprint arXiv:1403.7400.
Wang, G., Liu, X., & Fan, W. (2011). A knowledge adoption model based framework for finding helpful user-generated contents in online communities. ICIS 2011 Proceedings. Paper 15. Available at: http://aisel.aisnet.org/icis2011/proceedings/knowledge/15.
Werthner, H., & Klein, S. (1999). *Information technology and tourism: A challenging relationship*. Springer-Verlag Wien.
Xiang, Z., Schwartz, Z., Gerdes, 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*(1), 120–130.
Xiang, Z., Wöber, K., & Fesenmaier, D. R. (2008). Representation of the online tourism domain in search engines. *Journal of Travel Research, 47*(2), 137–150.
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.
Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management, 28*(1), 180–182.
Ye, Q., Li, H., Wang, Z., & Law, R. (2014). The influence of hotel price on perceived service quality and value in e-tourism an empirical investigation based on online traveler reviews. *Journal of Hospitality & Tourism Research, 38*(1), 23–39.
Yoo, K. H., Sigala, M., & Gretzel, U. (2016). Exploring TripAdvisor. In R. Egger, I. Gula, & D. Walcher (Eds.), *Open tourism* (pp. 239–255). Springer Berlin Heidelberg.
Zhang, Y., & Cole, S. T. (2016). Dimensions of lodging guest satisfaction among guests with mobility challenges: A mixed-method analysis of web-based texts. *Tourism Management, 53*, 13–27.
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.
Zhang, Z., Zhang, Z., & Yang, Y. (2016). The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior. *Tourism Management, 55*, 15–24.

Zheng Xiang, Ph.D., is Associate Professor in the Department of Hospitality and Tourism Management at Virginia Tech, USA. His research interests include travel information search, social media marketing, and business analytics for the tourism and hospitality industries. He is a recipient of Emerging Scholar of Distinction award by the International Academy for the Study of Tourism. He is a board member of International Federation for IT and Travel & Tourism (IFITT) and editorial board member of several international journals including Journal of Business Research, Journal of Travel Research, and Journal of Hospitality and Tourism Research.

Qianzhou Du is a Ph.D. student in the Department of Business Information Technology at Virginia Tech, USA. His research interests include business intelligence, text analytics, social media analytics, crowd sourcing, and business analytics for the tourism and hospitality industries. He is a recipient of ICTAS doctoral scholarship of Virginia Tech.

Yufeng Ma received the BE degree in Computer Science from Wuhan University, China in 2012. Currently he is a Ph.D. student in the Department of Computer Science at Virginia Tech, USA, where he is co-advised by Dr. Patrick Fan and Dr. Edward. A. Fox. His research interests include data mining, topic modeling in text mining, computer vision, deep learning and artificial intelligence in general.

Weiguo Fan is a full professor of the Department of Accounting and Information Systems and Director of the Center for Business Intelligence and Analytics in the Pamplin College of Business. He specializes in business intelligence, data and text analytics, social network analysis, social media analytics, business data mining using both internal and external data, complex system analysis and modeling, and consumer behavior modeling and analysis. He has applied analytics in financial statement risk analysis, search engine optimization, targeted and precision marketing, sentiment and opinion mining for product quality control, competitive intelligence, and deception detection in large scale business transactions.