International Journal of Information Management 48 (2019) 263–279

Contents lists available at ScienceDirect

International Journal of Information Management

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

Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor

Yung-Chun Chang[a,b], Chih-Hao Ku[c,*], Chun-Hung Chen[d]

[a] Graduate Institute of Data Science, Taipei Medical University, Taipei, Taiwan
[b] Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan
[c] College of Management, Lawrence Technological University, Southfield, USA
[d] Institute of Information Science, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Keywords:
Sentiment analysis
Hospitality
Natural language processing
Social media analytics
Visual analytics
Google trends
TripAdvisor

ABSTRACT

Analyzing and extracting insights from user-generated data has become a topic of interest among businesses and research groups because such data contains valuable information, e.g., consumers' opinions, ratings, and recommendations of products and services. However, the true value of social media data is rarely discovered due to overloaded information. Existing literature in analyzing online hotel reviews mainly focuses on a single data resource, lexicon, and analysis method and rarely provides marketing insights and decision-making information to improve business' service and quality of products. We propose an integrated framework which includes a data crawler, data preprocessing, sentiment-sensitive tree construction, convolution tree kernel classification, aspect extraction and category detection, and visual analytics to gain insights into hotel ratings and reviews. The empirical findings show that our proposed approach outperforms baseline algorithms as well as well-known sentiment classification methods, and achieves high precision (0.95) and recall (0.96). The visual analytics results reveal that Business travelers tend to give lower ratings, while Couples tend to give higher ratings. In general, users tend to rate lowest in July and highest in December. The Business travelers more frequently use negative keywords, such as "rude," "terrible," "horrible," "broken," and "dirty," to express their dissatisfied emotions toward their hotel stays in July.

1. Introduction

Nowadays, the ubiquitous Internet serves as a dominant channel for information diffusion. Popular social media, such as Twitter and Facebook, as well as online commerce company Amazon, enable users to express their opinions toward a variety of products and services. This form of word-of-mouth (WOM) communication introduces a new and important source of information for business intelligence and marketing (Zhang, Li, & Chen, 2012). Specifically, not only are online user-generated contents (UGCs), such as hotel reviews, essential to other potential customers, retailers, and product manufacturers, but also to businesses owners as they reveal customers' opinions towards their products and services (Giachanou & Crestani, 2016). Such information may have a significant impact on product sales such as online hotel bookings (Zhao, Ye, & Zhu, 2016). Further, owners can employ such information to enhance the quality of their services and products, and even devise new marketing strategies (Rhee, Yang, & Kim, 2016). Unfortunately, the rapid growth of the amount of UGCs impedes us from obtaining a comprehensive view of these opinions. That is, carrying out a manual analysis or survey is costly, and as a result, only data from limited customers can be reached and leveraged. Therefore, an automated analysis of reviews can be fruitful for both business providers and their customers.

Recent studies related to hotel review analysis mainly focused on an econometric approach (Jin, Karen, Ali, & Yong, 2017; Xie, Zhang, & Zhang, 2014), a survey approach (Filieri, Alguezaui, & McLeay, 2015), a statistical and modeling approach (Ya-Han & Kuanchin, 2016), and an opinion mining approach (Eivind, Burnett, & Nørvåg, 2012; Liu, 2012). However, most of these studies either use overall ratings, hotel stars, or the number of hotel reviews, as well as responses (Jin et al., 2017). Another approach is to choose a few, selected cities (Ya-Han & Kuanchin, 2016) for hotel review analysis due to the significant number of reviews available. A fine-grained text analysis such as an aspect-level sentiment analysis together with rating analysis and information visualizations can provide additional insights to the hotel review analysis.

Geetha, Singha, and Sinha (2017) point out that sentiment analysis has been mostly performed on Twitter and Amazon reviews to measure customer sentiments, while a research study specialized in sentiment

* Corresponding author at: College of Management, Lawrence Technological University, 21000 W 10 Mile Rd, Southfield, MI 48075, USA.
E-mail address: cku@ltu.edu (C.-H. Ku).

https://doi.org/10.1016/j.ijinfomgt.2017.11.001
Received 29 April 2017; Received in revised form 22 September 2017; Accepted 4 November 2017
Available online 14 November 2017

analysis of hotel reviews and its relations with customer ratings is still lacking. Furthermore, most studies use sentiment analysis and search data separately (Hutchin, 2016). To the best of our knowledge, limited research pays attention to developing a heuristic framework which includes data collection, data preprocessing, aspect extraction, and data analysis, and then ultimately makes sense of the data with visualizations for marketing. To bridge this knowledge gap, we developed an integrated framework for conducting sentiment and aspect analysis, along with an experimental study. We then used this framework to conduct an empirical study to collect and analyze hotel ratings and reviews from TripAdvisor data and analyze Google Trends data with novel sentiment analysis and visual analytics.

This study contributes to both research on and practice in multidisciplinary domains such as tourism, information systems, and computer science. We present a novel aspect-sentiment analysis method which concentrates on aspect detection for sentiment analysis. To detect sentiment with aspects from online reviews effectively, we cast aspect-sentiment analysis into a two-stage classification task. First, raw text reviews are transformed into a sentiment-sensitive tree structure where syntactic, semantic, and contextual information are preserved. The convolution tree kernel is then used to identify sentiment in online reviews. Once the sentiment of reviews is recognized, the aspects of sentiment are further detected. We propose an aspect detection method of sentiment in online reviews that employs six features covering syntactic and contextual clues.

Our experimental results validate that the proposed two-stage method can successfully recognize sentiment and detecting aspects. In addition, the proposed sentiment sensitive tree structure effectively employs syntactic structures, sentiment semantics, and the content of these text segments. Meanwhile, the developed features for aspect detection can capture the relation between sentiment and aspects. As a result, the proposed method outperforms baseline models as well as LibShortText (Yu, Ho, Juan, & Lin, 2013) and LDA-SVM (Blei, Ng, & Jordan, 2003), which are widely used to identify aspect-sentiment analysis for online reviews. Finally, hidden patterns and relations from the collected data can be discovered through visual analytics techniques, such as timeline analysis and word clouds.

The rest of this paper is organized as follows. First, we provide a review of contemporary studies in social media analytics, internet search results, hospitality and tourism, and sentiment, as well as aspect-based sentiment analysis. Next, we develop a scraping program to collect Hilton hotel reviews and ratings from TripAdvisor. We then propose a framework to conduct sentiment and aspects analysis with empirical experiments. Visual analytics techniques are applied to visualize TripAdvisor and Google Trends data to explore hidden information and relationships, which can be used for marketing or decision-making purposes. Finally, we draw conclusions and posit future research streams.

2. Literature review

2.1. Social media analytics

Social media websites enable users to generate and distribute information and experiences by interactive Web 2.0 technologies and applications, e.g., Facebook, Twitter, TripAdvisor, etc. This has led to an explosion of online UGCs, e.g., tweets, opinions, and reviews. However, such massive amounts of unstructured and semi-structured data also give rise to challenges and opportunities for data analysis.

Social media analytics involves techniques to collect, extract, analyze, and present user-generated data to support decision making, insight discovery, or other business-related operations (Holsapple, Hsiao, & Pakath, 2014). Similarly, Fan and Gordon (2014) specify three important stages: capture, understand, and present for social media analytics. During the capture stage, data are collected from various sources through news feeds, application program interfaces (APIs), or data crawling tools. Opinion mining, sentiment analysis, topic modeling, and trend analysis are often used to uncover hidden gems in these data. Different techniques such as network analysis, text and data mining, as well as natural language processing (NLP) are frequently used in this stage (Fan, Wallace, Rich, & Zhang, 2006). To present or make sense of large volumes of information, visualizations and visual analytics can support data exploration, discovery, and complex reasoning (Ribarsky, Xiaoyu Wang, & Dou, 2014). Statistical analysis, data mining, and machine learning are often used together with interactive visualizations (Ribarsky et al., 2014). Recently, there is an increasing interest in social media analytics using crawlers, NLP, and machine learning models to track trending topics for conducting sentiment analysis as well as opinion mining about products and services (Fan & Gordon, 2014; Xiang et al., 2017).

2.2. Internet search results (Google trends)

Search engines enable users to retrieve relevant information by typing keywords. Such search queries not only lead to a huge amount of UGC but also provide valuable information about users' interests and intentions. For example, the search volume and frequency data have been studied to measure perceived pollution and its relation to inbound tourism in China (Xu & Reed, 2017), to measure investors' attention to predict stock prices (Da, Engelberg, & Gao, 2011), to predict housing market trends and outperform experts' predictions from the National Association of Realtors for future U.S. home sales (Wu & Brynjolfsson, 2015), and to measure public concern over swine flu (Fenichel, Kuminoff, & Chowell, 2013). Among these studies, Google Trends data is the most frequently used because of the popularity of the Google search engine.

Google Trends[1] enables users to explore 'trending stories,' a collection of searching queries detected by Google's algorithms, by categories and locations. Google Trends are based on Google search results and present the search index of a particular keyword relative to the total search volume over a period by different categories and locations (Bangwayo-Skeete & Skeete, 2015). Consumers have been increasingly using Google for travel plans such as searching destination hotels and flight information (Bangwayo-Skeete & Skeete, 2015). Recently, Google Trends data has been studied to conduct tourism research. Li, Pan, Law, and Huang (2017) proposed a framework with search trend data to forecast tourism demand, and it achieved better performance than econometric models. Similarly, Bangwayo-Skeete and Skeete (2015) and Yang, Pan, and Song (2014) use Google Trends data to forecast tourism demand. To better understand consumer purchasing decisions for hotel bookings, Zhao et al. (2016) investigated the relationship between consumers' hotel searching behavior and hotel booking data on Expedia.com. They found that the higher volume of search queries for a hotel was associated with a higher room booking rate.

Fig. 1 shows a sample of query results for a recent five-year period, 2012–2017, across three popular hotel brands, Hilton, Marriott, and IHG, on Google Trends. Among them, the brand Hilton has been consistently searched most frequently during this period, and the Hilton and Marriott brands demonstrate similar user search patterns, such as peaks and downtime. The numbers of search interests are relative to the highest value in the U.S. during the sampled five-year period. For example, a value of 100 represents the peak in June 2012.

2.3. Hospitality, tourism and hotel reviews

The hospitality and tourism industry is an ideal application for social media analytics (Lu & Stepchenkova, 2015) because consumers' opinions, recommendations, ratings, and behavioral data contain rich information for data analysis. Hotel review websites such as

[1] Google Trends, https://trends.google.com/trends/.

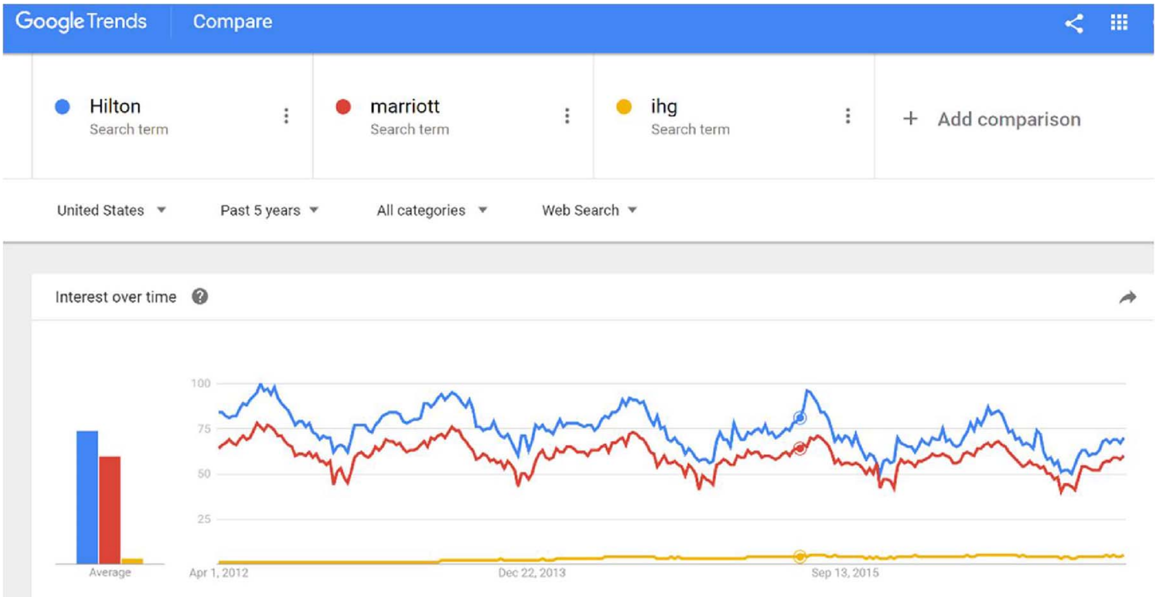

Fig. 1. A sample of query results for Hilton, Marriott, and IHG in Google Trends.

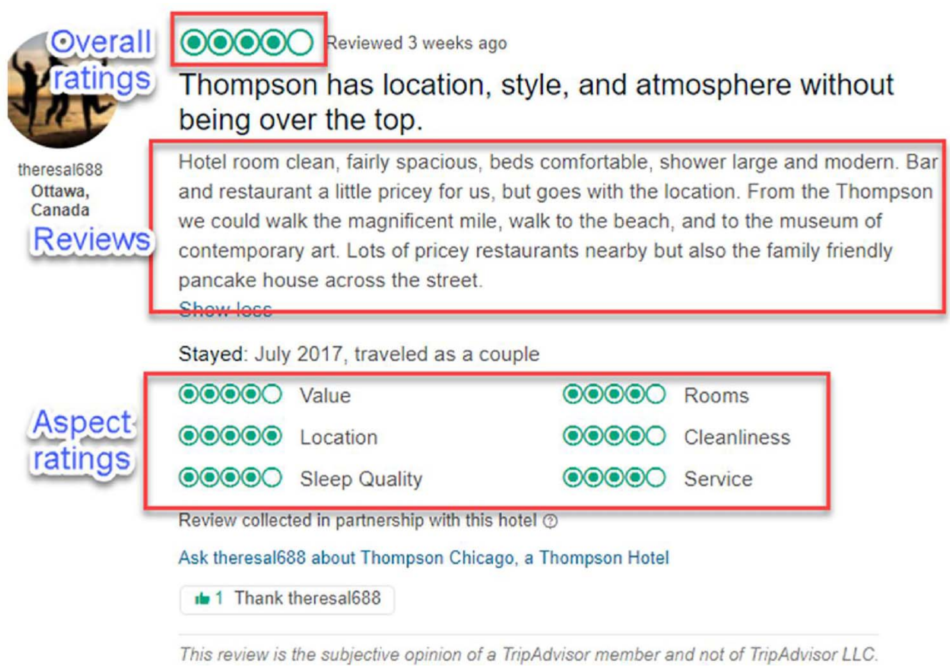

Fig. 2. A Sample of Hotel Review and Ratings.

TripAdvisor include rich WOM information which may affect other consumers' behaviors (Herrero, San Martín, & Hernández, 2015). For example, Herrero et al. (2015) examine users' perceptions regarding hotel reviews and show that users' decision-making process is based on the perceived value of the given information, credibility of the source, and similarity between users and review generators. Additionally, consumers' intention to book a room could be affected by Wi-Fi price and reviews (Eriksson & Fagerstrøm, 2017). Customer views and ratings provide great business values to hotels (Xie et al., 2014) and positively influence hotel bookings (Torres, Singh, & Robertson-Ring, 2015). The hotel reviews include quantitative values (e.g., numerical or rating values) and qualitative evaluations (e.g., users' opinions). Fig. 2 shows three major components of a hotel review from TripAdvisor: overall ratings, reviews, and aspect ratings. A traveler's experience at an accommodation is measured by ratings on a scale of 1 (terrible) to 5 (excellent), represented by green circles in Fig. 2. An overall rating involves different aspects of this facility. For example, TripAdvisor currently provides six aspects for users to review: sleep quality, location, room, service, value, and cleanliness. This type of online review can be used by marketing researchers for gathering the public perception of this location through sentiment analysis and topic categorization (Huang, Peng, Li, & Lee, 2013).

2.4. Sentiment analysis

Sentiment analysis involves the process of extracting and classifying opinions expressed in digital documents or Web content, and then the resulting data are analyzed against well-developed lexicons, a list of words or phrases classified into positive, negative, and neutral (Aydoğan & Akcayol, 2016; Geetha et al., 2017). Recently, sentiment analysis has attracted increasing attention in academia (Zhang, Guo, & Goes, 2013). For example, companies can analyze online comments

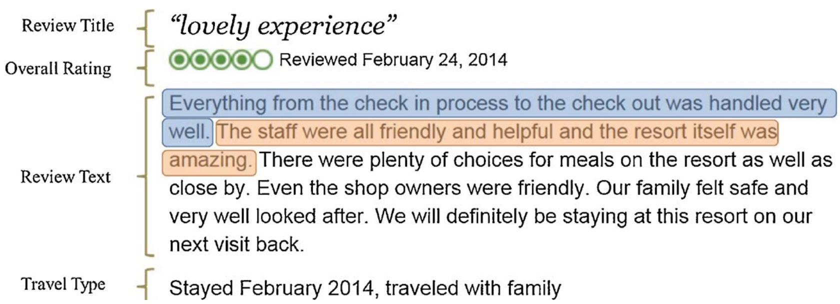

Fig. 3. A Sample Hotel Review.

which reflect people's sentiments or perceptions on movie reviews and foretell public interests (Singh, Piryani, Uddin, & Waila, 2013), consumer reviews on products (Lipizzi, Iandoli, & Ramirez Marquez, 2015; Shelke, Deshpande, & Thakare, 2017), and purchase decision on marketing (Lak & Turetken, 2014; Popescu & Etzioni, 2005). For tourism research, Hutchin (2016) predicted future visitors to the UK and Canada through sentiment analysis of collected hotel reviews and comments from TripAdvisor and Facebook, respectively. To explore customer opinions on hotel services, Wu et al. (2010) developed an interactive visualization system that enables users to compare opinions of different groups.

In general, sentiment analysis contains three fundamental challenges: aspect detection (Bhattacharjee & Petzold, 2016), opinion word detection, and sentiment orientation identification (Paltoglou & Thelwall, 2012). An aspect (also called an opinion target) is an important indicator for clarifying the topic of an opinion. Without this information, determining whether the opinion is expressed on a specific service/amenity or an entire accommodation review lacks clarity. Fig. 3 displays a typical hotel review that contains the overall rating of a variety of facets of the hotel, including check-in/out, location, and service. However, it would be difficult for decision makers to determine the reviewer's assessment of individual aspects. Depending on the nature of the traveler, the emphasis may be different. For instance, a family may focus on the safety of the public area while a solo traveler may prefer a high-quality bed. Knowing the distinctive aspect of each review is imperative given that the overall score may be similar across consumers' reviews. Consequently, an increasing amount of work is aimed at aspect-based sentiment analysis.

2.5. Aspect-based sentiment analysis

Various degrees of granularities of sentiment analysis can be conducted depending on the scope of the data being a document, sentence, entity or at aspect-level (Liu, 2012). Unfortunately, in a document- or sentence-level analysis, fine-grained opinions cannot be easily obtained. However, an aspect-level analysis can be used to resolve this issue. For example, the sentence "*…very convenient to O'Hare and decent service, but outside of that it's way overpriced for the rooms and level of comfort…*" is positive about the *service*, but negative about the *room*. That is, an opinion consists of a *sentiment* (positive or negative) and a *target* of opinion. Both aspect extraction and sentiment analysis are required to improve the quality of services or products. NLP techniques are often used together with aspect-based sentiment analysis.

Different types of aspect-based sentiment analysis have been proposed, including rule-based, topic model-based, and supervised learning. For instance, Hu and Liu (2004) develop an association rule-based method to extract aspects from product reviews. To mine the associated rules, they extract frequent itemsets from noun phrases of product reviews based on the Apriori algorithm (Han, Kamber, & Pei, 2011). However, this method extracts product features and opinion words, but neglects to connect product features at an aspect-level, which renders it difficult to group the extracted terms into distinct aspect categories.

In recent years, some research has been conducted based on Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to develop topic modelling methods for aspect-based sentiment analysis tasks. These approaches utilize variations of LDA to uncover latent topics in a dataset, expecting that these topics will correspond to ratable aspects of the entity under review. For instance, Brody and Elhadad (2010) develop a local topic model for extracting aspects and determining sentiment in the review text. This method works at the sentence level and uses a small number of topics to detect the aspects automatically. Zhao, Jiang, Yan, and Li (2010) present a maximum entropy hybrid model that can collaboratively recognize aspect and opinion words. The hybrid model incorporates a discriminative maximum entropy model into a generative topic model, and it leverages syntactic features to enhance the performance for recognizing aspects and opinion words.

Existing methods primarily treat aspect-based sentiment analysis as supervised learning problems. Given a training dataset containing a collection of manually-tagged instances of aspect-sentiments, computer scientists often use a supervised learning method to train a classifier for sentiment classification and further recognize aspects. Jakob and Gurevych (2010) view aspect extractions as a sequence labeling problem and employ conditional random fields (CRFs) (Lafferty, McCallum, & Pereira, 2001) to recognize aspect expressions. They combine dependency information of a text expression with syntactic features, such as part-of-speech (POS) tags, to detect and extract opinion aspects. To identify the important aspects, Yu, Zha, Wang, and Chua (2011) develop an algorithm for aspect ranking based on the aspect frequency simultaneously with the influence of consumers' opinions towards each aspect of their overall opinions. The experiment results indicate that the performance of sentiment classification tasks can be improved through this approach on aspects recognition, and it can be further applied to the aspect ranking results. Liu, Xu, and Zhao (2015) formulate the opinion aspect extraction as a graph-based co-ranking task. Their proposed method extracts candidates of the opinion aspect and word through constructing a heterogeneous graph to model semantic relations and opinion relations. Finally, each candidate is given a confidence score which is then ranked by a co-ranking algorithm. A threshold is used to filter out those candidates with lower scores, and the remaining candidates are considered to be the results.

Our method differs from existing aspect-sentiment analysis approaches for two reasons. First, the proposed method can represent syntactic, content, and semantic information in text segments of the review through the proposed sentiment sensitive tree structure. Therefore, the method can effectively utilize the structured information and constituent dependencies knowledge derived from sentences for sentiment classification tasks. Second, most approaches analyze the text of reviews for aspect detection. Our method further considers the properties of reviews (e.g., a customer's profile and travel type) to enhance the aspect detection performance.

3. Methodology

We propose a method that automatically recognizes the sentiment of online hotel reviews. Aspect categories are detected from the review

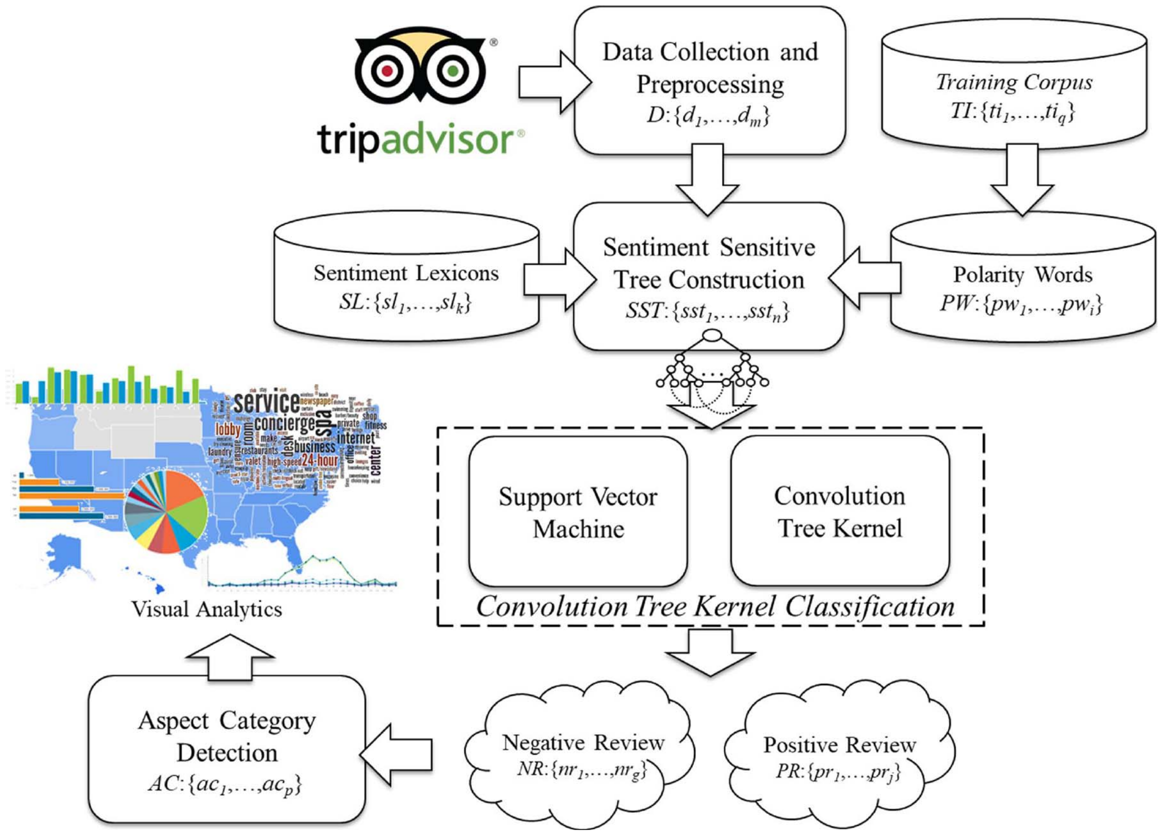

Fig. 4. The system architecture.

texts to help readers comprehend these reviews. Fig. 4 shows the system architecture of the proposed aspect-based sentiment analysis method, which is comprised of five key components that are described in the following corresponding sections: *data collection and processing* (section 3.1), *sentiment sensitive tree (SST) construction* (section 3.2), *convolution tree kernel classification* (section 3.3), *aspect category detection* (section 3.4), and *data visualization* (section 3.5). We first gather hotel reviews from TripAdvisor and further process data cleaning to ensure data quality. Text syntactic information (e.g., parse tree) is useful to resolve the relationships between entities; therefore, we propose the *SST* structure, which depicts a path along the syntactic parse tree that consists of sentiment expressions. Also, the semantics in the content is assessed to decorate the SST with polarity. We adopt the convolution tree kernel (Collins & Duffy, 2002) to measure the similarity between review texts regarding their SSTs. The tree kernel is incorporated into a support vector machine (SVM) (Joachims, 1998) to learn a classifier for each structural type, which detects sentiment in hotel reviews. The aspects behind the sentiment are then detected based on the properties (e.g., travel type, customer's profile, and location) of reviews. Next, the data visualization provides an integrated, multi-feature analysis service for data exploration and decision making. Specific implementation of sub-components in this model is introduced in the following sections.

3.1. Data collection and preprocessing

The Hilton hotel was selected in this study for two reasons. First, it is the most well-known and valuable hotel brand, with a value of US $7.8 billion in 2016, based on Brand Finance data.[2] Secondly, we used Google Trends to compare a variety of hotel brands, e.g., Hilton, Marriott, Hyatt, Sheraton, Holiday Inn, etc., and found that "Hilton" is the most frequently searched keyword among all hotel brands. TripAdvisor was selected because it is a well-known travel company and provides an enormous amount of user-generated content. By selecting a highly trafficked company like TripAdvisor, the likelihood of collecting rich and sufficient data for our analysis greatly increases.

In addition to the overall rating, reviewers also provided eight aspect ratings in each review: *Value, Location, Sleep Quality, Rooms, Cleanliness, Service, Check-in/Front Desk, and Business Service.* The ratings ranged from 1 to 5 stars, which can serve as ground-truth for quantitative evaluations of our method. Customer profiles and properties of the hotels, such as Location, Highlight, and Amenities, were also collected. Simple preprocessing was then performed on the review data: 1) define sentiment of reviews (i.e. a 1–3 rating star for negative and 4–5 rating star for positive); 2) remove reviews without any aspect label; 3) convert words into lower cases; 4) remove punctuations, stop words[3] and words appearing less than 5 times in the corpus; and 5) stem each word to its root with Porter Stemmer (Porter, 1980). After

Table 1
The statistics of data corpus.

# of positive reviews	500,316
# of negative reviews	133,961
# of aspects on *Value*	538,898
# of aspects on *Location*	524,583
# of aspects on *Sleep Quality*	483,969
# of aspects on *Rooms*	526,889
# of aspects on *Cleanliness*	540,187
# of aspects on *Service*	629,871
# of aspects on *Check-in/Front Desk*	15,188
# of aspects on *Business Service*	10,289

[2] The World's Most Valuable Hotel Brands, http://ftnnews.com/accommodation/30048-the-world-s-most-valuable-hotel-brands.html.

[3] http://jmlr.csail.mit.edu/papers/volume5/lewis04a/a11-smart-stop-list/english.stop.

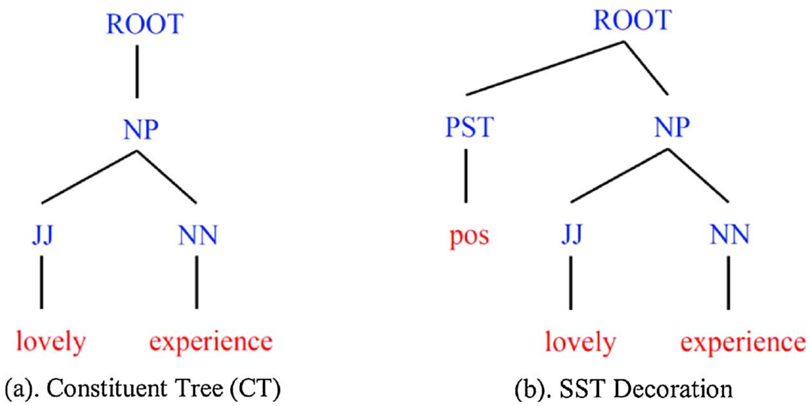

Fig. 5. The CT and SST decoration operation of the example segment.

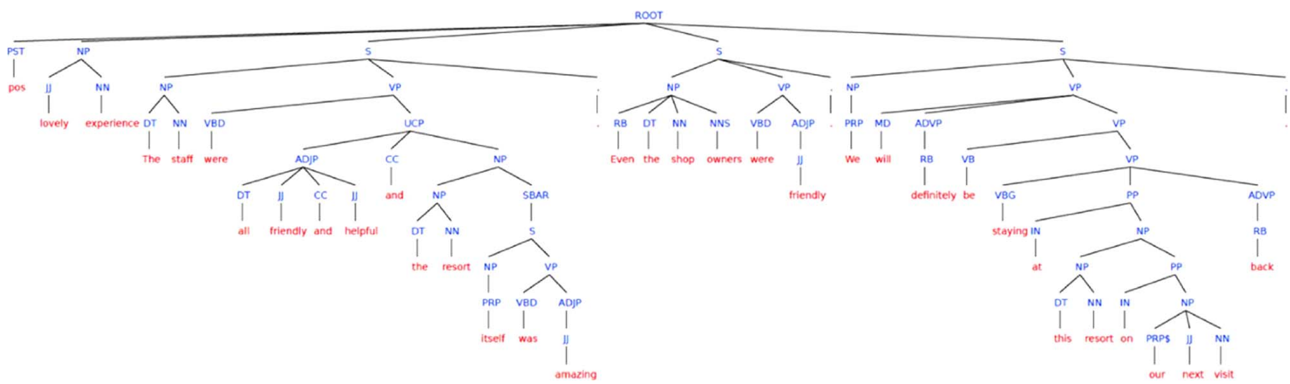

Fig. 6. The extension operation of SST of the example hotel review.

these preprocessing steps, the dataset consists of 634,277 reviews and 5,456,307 sentences from 749 Hilton hotels in the U.S.; the details are illustrated in Table 1. The number of aspects on Check-in/Front Desk and Business Service is typically low. We will explore the reason in the visual analytics section.

3.2. SST construction

Next, we represent hotel reviews with the SST structure, which is a constituent tree (CT) of review titles enhanced with three operations: *decoration*, *extension*, and *pruning*. Jiang, Zhang, Fu, Niu, and Yang (2010) and Gamon (2004) had shown that CT of review titles is effective in sentiment classification because it can model the syntactic structure of text that affects the sentiment classification performance. We exemplify the operators through the examples in Fig. 5 to further facilitate comprehension. Fig. 5(a) presents the syntactic parse tree of the example review title (c.f. Fig. 3), while Fig. 5(b) demonstrates how SST polishes CT with three operators that will be explained in the following paragraphs.

The first operation we propose is SST Decoration. Intuitive indicators of sentiment may include verbs and adjectives, yet not all deliver clear sentimental expression. Therefore, we intend to find only those that are strongly associated with a polarity of sentiment in an SST to improve the sentiment classification performance (denoted as *polarity words* hereafter.) To achieve this, a log-likelihood ratio (LLR) (Manning, Raghavan, & Schütze, 2008) is utilized to compile a list of polarity words representing each sentiment. The adopted definition of LLR is given in Eq. (1), which calculates the log-likelihood of the hypothesis that the presence of w in set S is beyond chance. Let w denote a word and S a type of sentiment. $N(S)$ and $N(\neg S)$ are the number of reviews that do and do not contain sentiment S, respectively. $N(w\hat{}S)$, which is shortened as k, is the number of reviews containing w and S simultaneously, while $N(w\hat{}\neg S)$ is the number of reviews including w but not S, denoted as l. To further simplify the formula, we also define $m = N(S) - k$ as the number of reviews containing S without word w, and $n = N(\neg S) - l$ which means those with neither S nor w. A maximum likelihood estimation is performed to obtain probabilities $p(w)$, $p(w|S)$, and $p(w|\hat{}S)$. Finally, Eq. (1) can be applied to determine LLR values of every word in the corpus, where those with a higher LLR value are believed to be strongly related to a certain sentiment. We then sort LLR values of all words in the training data, and the top 300 candidates are kept as a polarity keyword list. For each SST that holds a polarity word, we attach a positive or negative sentiment tag (denoted as PST and NST) as a child under the root to integrate sentimental semantic information into the SST structure (as shown in Fig. 5(b)).

$$LLR(w, S) = 2\log\left[\frac{p(w|S)^k(1-p(w|S))^m p(w|\neg S)^l(1-p(w|\neg S))^n}{p(w)^{k+l}(1-p(w))^{m+n}}\right] \tag{1}$$

To the best of our knowledge, the majority of existing tree kernel-based approaches working on the sentence level verified the usefulness of constituent tree structure of review title on sentiment classification; however, the information within CT may not be sufficient for the current task. The review title “Business trip,” for instance, is too short to signify any sentiment information. To include indicative sentiment context, the extension operator extracts sentiment expressions from the review content. The constituent tree structure of the extracted sentence is concatenated into the SST if it contains a sentiment lexicon.[4] As shown in Fig. 6, the extended SST includes richer context information than the original after incorporating the context from three sentences.

In this study, we adopt SVM for sentiment classification. SVM performs classification on a vector space by estimating a boundary between data of different classes with maximum margin (Manning et al., 2008). The model itself is error-tolerant to a certain degree;

[4] Here, we adopt a well-known sentiment lexicon compiled by Wilson, Wiebe, and Hoffmann (2005).

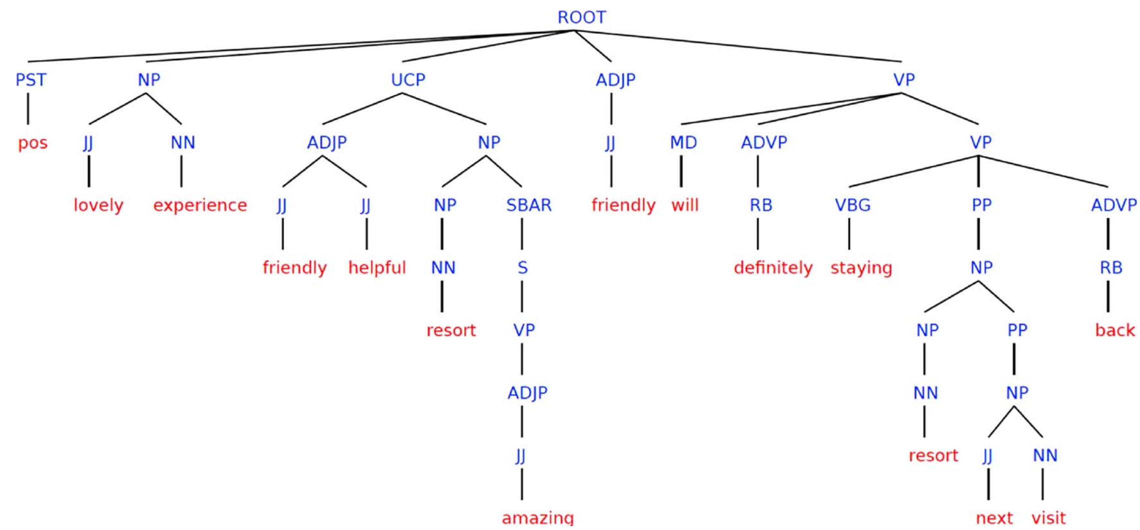

Fig. 7. The pruning operation of SST of the example segment.

nevertheless, we would like to improve the quality of the input data as much as possible. We noticed that SSTs may be comprised of redundant elements that can potentially undermine the effectiveness of sentiment classification. Therefore, we performed pruning on the tree. The pruning operator condensed SSTs through the following procedures.

First, even though the problem of insufficient sentiment information can be alleviated by including extra sentences from review content, many irrelevant contexts are likely to be included in sentiment analysis. To isolate contexts associated with the sentiment of review, we present sentiment contexts for each extended sentence as a sub-tree attached to the lowest common parent of the sentiment lexicon. For instance, the sentiment context "resort amazing" is extracted from the extended sentence "The staff was all friendly and helpful and the resort itself was amazing," by the sub-tree rooted at "NP" node, which is the lowest common parent of nodes "resort" and "amazing." Next, since frequent words are generally too common to be useful for sentiment discrimination, the pruning operator removes punctuations and stop words[5]; any word with less than five times of occurrence in the corpus is also removed along with its corresponding elements in SST.

Second, an SST might contain duplicate nodes without siblings and carry the same tag as their parents. These nodes may obstruct the computation of similarity in the tree kernel, since the similarity is determined by the part where two trees overlap. To overcome this issue, duplicate nodes are removed by the pruning operator to produce a tidier SST, as shown in Fig. 7.

3.3. Convolution tree kernel classification

SVM can employ various kernels as the internal mechanism for computing the similarity (i.e., dot product) among pairs of samples projected onto a different vector space. It is straightforward to adopt the convolution tree kernel (CTK) (Collins & Duffy, 2002) to resolve the similarity between sentences under our representation scheme. CTK can model multi-level structural information and is suitable for types of data such as parse trees. A syntactic parse tree T can be denoted by a vector of integers:

$$\phi(T) = (\#subtree_1(T),\ ...,\#subtree_n(T),\ ...,\#subtree_m(T)), \tag{2}$$

where $\#subtree_n(T)$ is the number of occurrences of the n^{th} type of sub-trees ($subtree_n$) in T. However, the number of different sub-trees in T is exponential to the size of T. Thus, a direct calculation of the feature vector $\phi(T)$ is not viable. To circumvent this problem, the CTK evaluates the structural similarity as follows:

$$\begin{aligned} K_{CTK}(T_1|T_2) &= \langle \varphi(T_1)|\varphi(T_2)\rangle \\ &= \sum_k \#subtree_k(T_1)\cdot\#subtree_k(T_2) \\ &= \sum_k \left(\sum_{n_1\in N_1} I_{subtree_k}(n_1)\right)\left(\sum_{n_2\in N_2} I_{subtree_k}(n_2)\right) \\ &= \sum_{n_1\in N_1, n_2\in N_2} \Delta(n_1,\ n_2) \end{aligned} \tag{3}$$

where N_1 and N_2 are the sets of nodes in trees T_1 and T_2, respectively, and $I_{subtree_k}(n)$is a function of node n whose output is 1 if a $subtree_k$ exists under n, and zero otherwise. Consequently, the kernel K_{CTK} assesses structural similarity between two sentiment sensitive trees SST_1 and SST_2 by computing the amount of shared sub-trees, as in the following formula:

$$K_{CTK}(SST_1,\ SST_2) = \sum_{n_1\in N_1, n_2\in N_2} \Delta(n_1,\ n_2), \tag{4}$$

where N_1 and N_2 are the sets of nodes in SST_1 and SST_2, respectively. Furthermore, Δ (n_1, n_2) denotes the common sub-trees under nodes n_1 and n_2. The recursive algorithm is:

(1) if n_1 and n_2 have different direct children nodes, Δ (n_1, n_2) = 0;
(2) if both n_1 and n_2 are pre-terminals (POS tags), Δ (n_1, n_2) = 1 × λ;
(3) otherwise, evaluate Δ (n_1, n_2) recursively as:

$$\Delta(n_1,\ n_2) = \lambda \prod_{i=1}^{\#ch(n_1)} (1 + \Delta(ch(n_1,\ i),ch(n_2,\ i))) \tag{5}$$

where $\#ch(n_1)$ is the number of children of node n_1, $ch(n,i)$ is the i^{th} child of node n, and λ a decaying parameter in the range of [0,1] for the purpose of smoothing the output value among trees of variable sizes. Collins & Duffy (2002) stated that the time complexity of this kernel is $O(|N_1|\cdot|N_2|)$, which is dependents on the number of nodes in the two input trees.

3.4. Aspect category detection

Aspect category detection is a task that aims to detect aspect categories mentioned in a piece of text. Typically, this can be considered as a multi-label classification problem. Compared with the specific terms, aspect categories are more generic and sometimes can only be implicated by the surface terms. For instance, the review in Fig. 3 contains three latent aspect categories, Check-in/out, Value, and Service. To resolve the differences between surface and latent aspects in text, we

[5] http://jmlr.csail.mit.edu/papers/volume5/lewis04a/a11-smart-stop-list/english.stop.

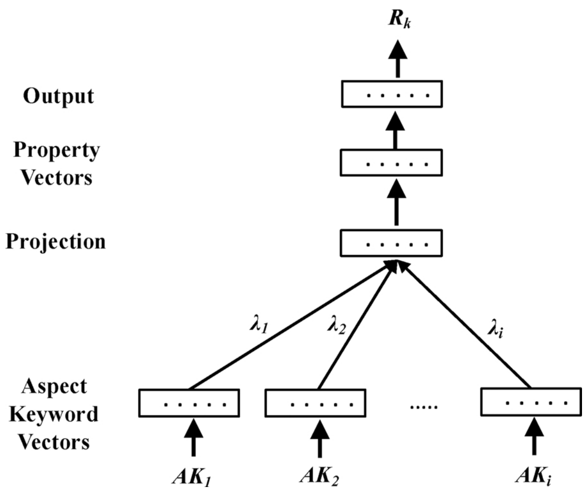

Fig. 8. The DAV model represents each hotel review *Rk* as aspect keyword vectors *AKi* that are present in the review, weighted by scaling LLR scores λ_i.

present the Distributed Aspect Vectors (DAV) based on embeddings to model the aspect-sentiment of hotel reviews. Compared with the Bag-Of-Words (BOW) model, this distributed model of hotel reviews can incorporate wider context information that covers the entire review into the representation. Moreover, semantic relations of various surface words can also be captured due to the nature of a vector space model. These characteristics cannot be easily accomplished in a BOW-based approach, since it would consume a significant amount of storage for a sizable *n*-gram dictionary. In addition, Chang, Chen, Hsieh, Chen, and Hsu (2015) experimented with using keywords to classify emotions and demonstrated their effectiveness, which encouraged us to employ this approach in the aspect category detection task.

At the outset, we use the Continuous Bag-of-Words (CBOW) method (Le & Mikolov, 2014) to acquire vector representations of words in the corpus. We then utilize LLR to extract keywords associated with the aspect categories. Subsequently; a hotel review is jointly represented by aspect keyword embeddings. More specifically; weight λ_i for an aspect keyword; AK_i; is determined by its LLR value; and a hotel review; R_k; is represented by the weighted average of those embeddings; as depicted in Fig. 8. As illustrated in Fig. 9; if an aspect keyword is not present in a piece of review; we first calculate the mean vector of all words and cosine similarity between the mean and every keyword vector. Then; the keywords 1with a higher similarity are selected to represent this review. In essence; each hotel review is projected onto a point in the keyword vector space where various types of classifiers can be employed. In addition to aspect keyword vectors; the properties of hotel reviews are also represented through feature vectors which are appended to the aspect keyword vectors; including customer's profile (i.e. TripCollective Progress and Badges); travel type; and time of year.

3.5. Visual analytics

Visual analytics provides an interactive visual interface to reveal hidden structures and details (Fan & Gordon, 2014). It combines machine and human strength to process and explore large data and provides decision-support information (Sacha et al., 2014). The power of visual analytics comes from the machine's capability to store and analyze big data efficiently together with humans' perceptive skills, cognitive reasoning, and domain knowledge. This paper aims to use visual analytics techniques to explore the extracted data, e.g., hotel ratings, aspect-sentiment, and types of travelers using Tableau. Tableau[6] is a representative of applications from the information visualization domain (Sacha et al., 2014). It provides varying visual analyses such as timeline analysis, location-based analysis, and dashboard analysis. Additionally, it provides trendline, forecast, and cluster models for data analytics. However, the collected raw data cannot be visualized directly. We first process the raw data to exclude data records with a wrong year and incomplete locations. The raw data includes multi-values such as "|*Value-2*|*Location-5*|*Sleep Quality-4*|*Rooms-3*|*Cleanliness-1*|*Service-3*|." Therefore, we split and classify each aspect and extract the numerical value for each aspect. Finally, we also transform each hotel address into the corresponding geographical location. The data visualization and exploration are then reported in Section 5: Visual Analytics, Results and Discussion.

[6] Tableau, https://www.tableau.com/.

4. Experiments

4.1. Experimental settings

We conducted an extensive evaluation of the performance of the proposed system and other baseline methods. First, we implemented a support vector regression (SVR) model that uses averaged word embeddings (with 300 dimensions) to represent each hotel review, denoted as SVR_{WE}. Next, we verified the effect of our keyword extraction approach by including a sentiment keyword-based model learned by SVM (denoted as SVM_{KW}). Moreover, a probabilistic graphical model that exploits the LDA (Blei et al., 2003) as review representation was employed to train an SVM to classify the hotel reviews, denoted as SVM_{LDA}. Included next was a logistic regression model from LibShortText (Yu et al., 2013) which was trained by review title, denoted as *LST*. We also included the Naïve Bayes method, denoted as *NB* (McCallum & Nigam, 1998), as the baseline. The implementation details of these methods are provided as follows. We employed the Stanford parser[7] to generate the output of parse trees and POS tagging. The dictionary required by NB and SVM_{LDA} was created by excluding words in the English stop word list mentioned in Section 3.2.1. As for unobserved words, we used Laplace smoothing in NB, and an LDA toolkit[8] was employed in SVM_{LDA}. The convolution kernel of an SST was formed by the Moschitti's tree kernel toolkit (Moschitti, 2004). The distributed aspect keyword vector was integrated with vectors of review properties for LibSVM[9] to detect aspect categories. We adopted a 10-fold cross-validation scheme. Evaluation metrics include the precision, recall, F_1-measure, and accuracy. Precision and recall are used to measure the positive predictive values and sensitivity, respectively, and F_1-measure is a balanced mixture of precision and recall values.

4.2. Performance evaluation and comparison

The performances of sentiment classification systems are listed in Table 2. First, our approach outperforms all of the compared methods. This is not surprising because the compared methods and feature sets simply use syntactic features and cannot successfully uncover deep semantics of reviewers' sentiments in hotel reviews. In contrast, our method incorporates semantic and context-dependent features, thus achieving the best performance. As for other methods, the NB, which is a statistical method with keywords as features, performs less than ideal. It shows that inter-word relations cannot be modelled by this method, since it only considers surface word weightings. The F_1-measure of the NB classifier is 0.77. SVR_{WE}, on the other hand, overcomes the above weakness by utilizing a vector space regression model, which results in a 5% increase in F_1-measure. Moreover, SVM_{KW} greatly outperforms the methods above and achieves substantial proficiency in classifying the sentiments with a 0.848 F_1-measure. This indicates that keyword

[7] http://nlp.stanford.edu/software/lex-parser.shtml.
[8] http://nlp.stanford.edu/software/tmt/tmt-0.4/.
[9] https://www.csie.ntu.edu.tw/~cjlin/libsvm/.

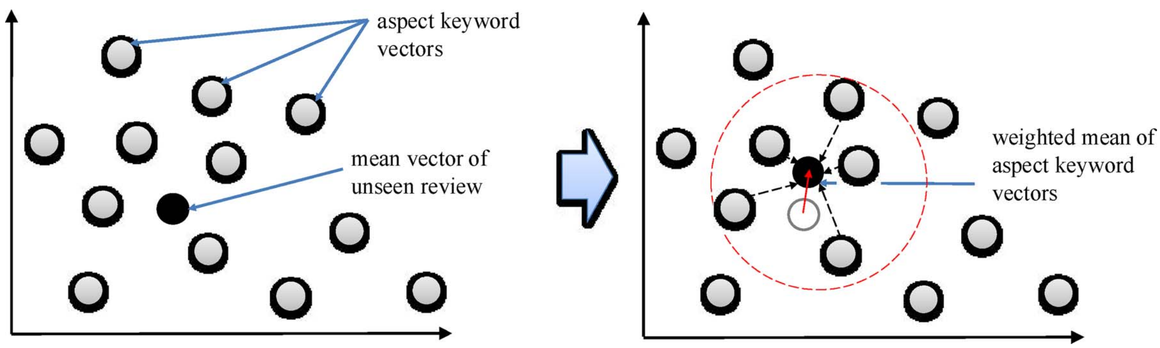

Fig. 9. DAV calculates the weighted average of nearest aspect keyword vectors when a review contains no keywords.

Table 2
The sentiment classification results of the compared methods.

System	*Precision*	*Recall*	*F_1-measure*	*Accuracy*
NB	0.846	0.711	0.772	0.711
SVR_{WE}	0.843	0.768	0.788	0.768
SVM_{KW}	0.852	0.844	0.848	0.844
SVM_{LDA}	0.893	0.897	0.895	0.879
LST	0.895	0.899	0.897	0.889
Our method	**0.954**	**0.967**	**0.935**	**0.893**

information ranked by LLR alone can be sufficiently effective in deciding the sentiment of a review. Notably, LST obtained an outstanding 0.9 F_1-measure since it further examines the content of the title through an n-gram model. Hence, the LST method outperforms other baselines.

To detect aspect of hotel reviews, word vectors are learned using CBOW (Le & Mikolov, 2014) with a hierarchical soft-max algorithm at the outset of the training stage. The dimensionality is set to 300. In the meantime, for every word in the corpus, we calculate its LLR values. According to Chang et al. (2015), 100 keywords can be useful for categorization of texts. Thus, we retain only this amount of words after sorting by LLR as keywords for each sentiment. Next, we transform each hotel review into the weighted sum of keyword vectors, where the weights are determined by its scaled LLR values. We adopt LibSVM (Chang & Lin, 2011)with the radial basis function (RBF) kernel and multi-label output format for the classification of aspects. More specifically, the multi-label mode operates by building a binary classification model for each label of an instance, and each instance may carry multiple labels. Table 3 shows the precision, recall, F_1-measure, and accuracy for detecting the aspects mentioned in over six hundred thousand hotel reviews.

The results in Table 3 show that DAV can successfully detect aspects contained in hotel reviews. It indicates that aspect keywords and properties in reviews are valuable clues when identifying aspects therein. It also demonstrates that dense vector representation, such as embeddings, can capture latent associations between aspects and keywords. Furthermore, using LLR as a weighting mechanism can positively affect the discriminating capability of the classifier to correctly identify aspects. These representative features can be optimally utilized by SVM or other vector-based classifiers. In addition, the proposed method requires no feature engineering or human supervision, as the weights of keywords are calculated automatically. We further conducted deeper analysis of the detection results and found that a large number of reviews convey aspects on *Value*, *Location*, *Sleep Quality*, *Rooms*, *Cleanliness*, and *Service*. We, therefore, can achieve high performances on all of these aspects. However, a relatively lower number of hotel reviews on certain aspects, *Check-in/Front Desk* and *Business Service,* led to a lower precision and recall rate and less-than-ideal system performance.

Table 3
The performance of DAV on aspect classification.

Aspect	*Precision*	*Recall*	*F_1-measure*	*Accuracy*
Value	0.850	0.997	0.918	0.848
Location	0.830	0.996	0.904	0.828
Sleep Quality	0.772	0.978	0.863	0.762
Rooms	0.832	0.996	0.907	0.829
Cleanliness	0.851	0.998	0.919	0.849
Service	0.993	0.921	0.956	0.993
Check-in/Front Desk	0.490	0.465	0.478	0.976
Business Service	0.469	0.432	0.450	0.983

In sum, the proposed model effectively extracts the syntactic, contextual, and semantic information in hotel reviews. Moreover, irrelevant syntactic elements are eliminated, essentially reducing noise, so that the informative context of those sentences that convey sentiments can be efficiently learned. Therefore, our method outperforms the compared approaches and achieves an encouraging sentiment classification performance. Moreover, the distributed aspect vector is effective for aspect determination of hotel reviews since it incorporates aspect keywords vectors with property vectors. Hence, our method obtains high performance in detecting aspects of hotel reviews.

5. Visual analytics, results and discussion

5.1. Google trends exploration

The data collected from TripAdvisor included overall ratings, aspect ratings, and Hilton hotel reviews in the U.S. We first explore the search interest in Hilton and TripAdvisor by interest over time, interest by subregion (states, metros, and cities), and related topics as well as queries. The query results are then exported to .csv files for further analysis in Tableau. The query results are accumulated from January 2004 to March 2017.

A timeline analysis and dashboard (see Fig. 10) are used to compare the average ratings (top panel), the number of reviews (middle panel), and the relative value of "Hilton" in Google Trends (bottom panel) from January 2004 to March 2017. A red-blue diverging color and date annotations are used to facilitate data exploration and comparison. A red and blue color denotes lower and higher scores, respectively. The rating scores on TripAdvisor range from 1 (a lower rating) to 5 (a higher rating), and the maximum search interest (index) is 100 on Google Trends. The average ratings and the number of hotel reviews have increased, while there is a decreasing trend of searching "Hilton" since January 2004. The peak time of searching "Hilton" usually occurs in June or July, and the valley of the curve occurs in December of each year. Additionally, there is an increasing trend of hotel reviews generated in July and a decreasing trend in December and January of each year. A couple of interesting observations are that the average rating score in July is usually the lowest and the highest rating score often occurs in December and January. This indicates the more users search hotel information using Google, the more hotel reviews will be

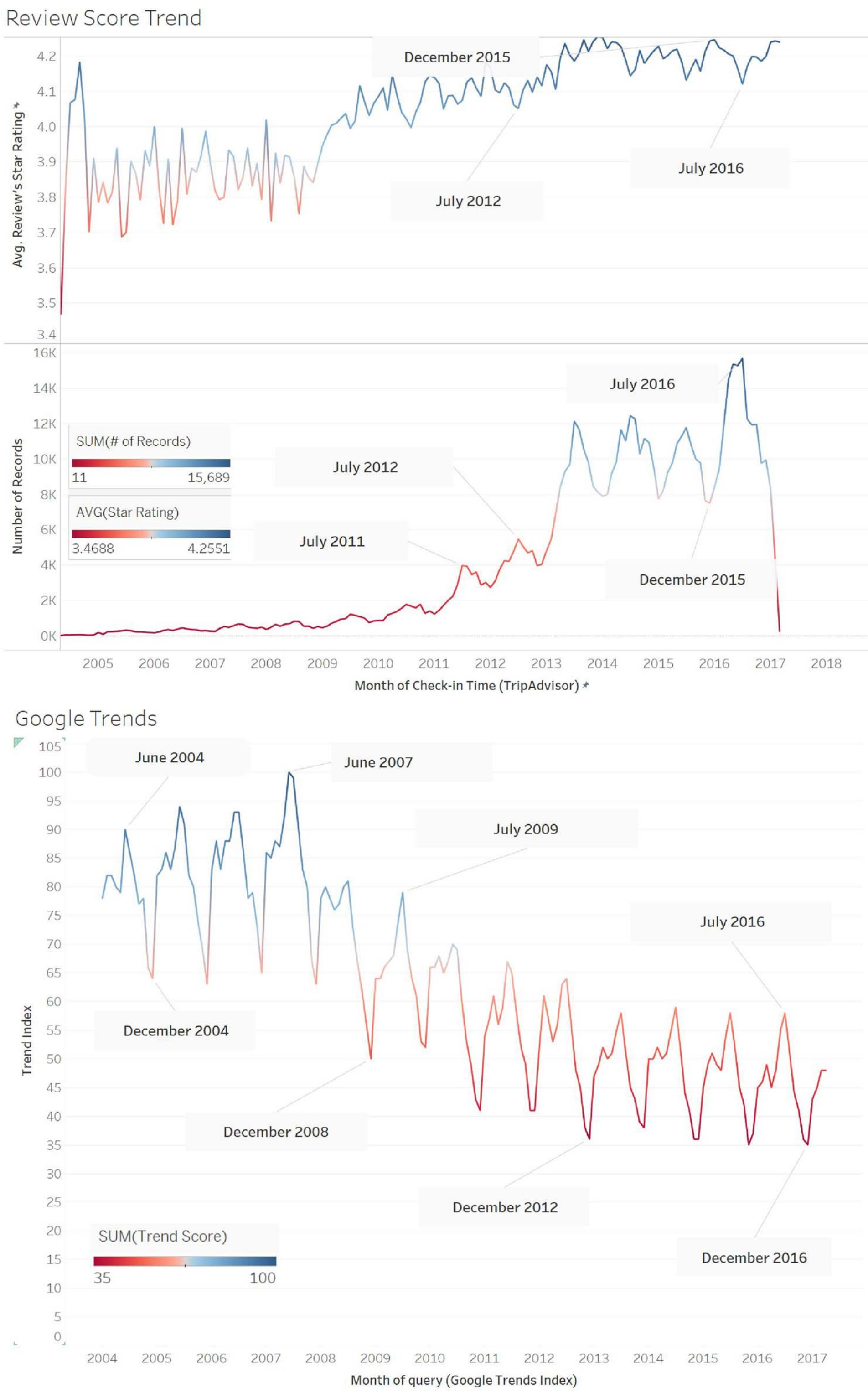

Fig. 10. Timeline analysis on average review ratings (top), the number of reviews (middle), and Google Trends indexes (bottom).

generated in TripAdvisor later; however, a lower rating score can be expected.

To better understand the search interest of a "Hilton hotel" by various locations during the sample period in Google Trends, we examine the results by sub-region (state) and city level. The searching results reveal that Hawaii was the most popular location to search "Hilton hotels," followed by the District of Columbia, Maryland, and Florida. For a city-level result (see Fig. 11), Honolulu was the most popular city to search "Hilton hotel," followed by Washington, Orlando, Nashville, and San Diego.

We examine an equally interesting question; *what keywords (topics) do users usually search together with a "Hilton hotel" query*? The results

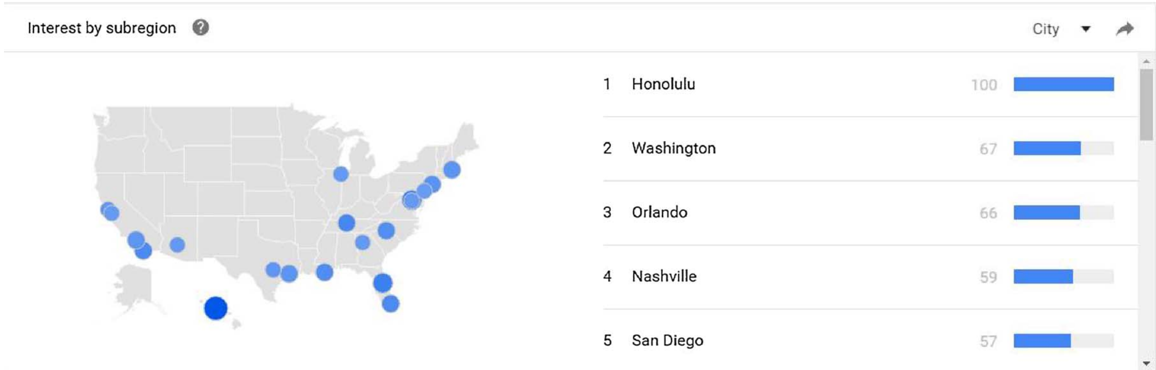

Fig. 11. Search interest of “Hilton Hotel” by cities in Google Trends.

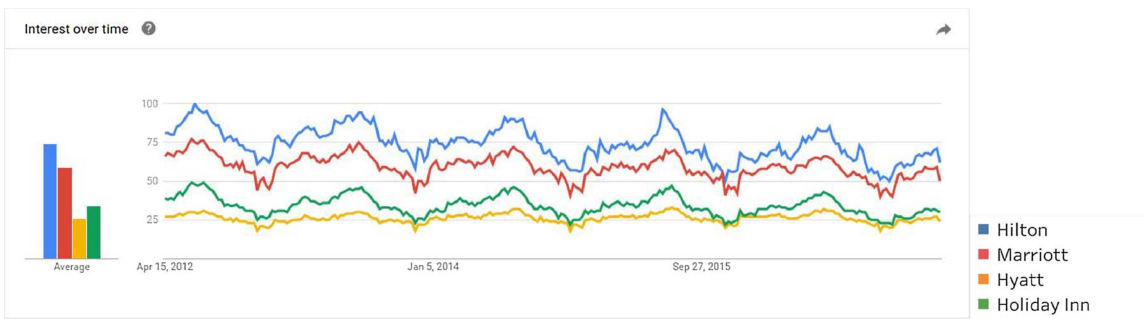

Fig. 12. Search interest of Hilton, Marriott, Hyatt, and Holiday Inn in Google Trends.

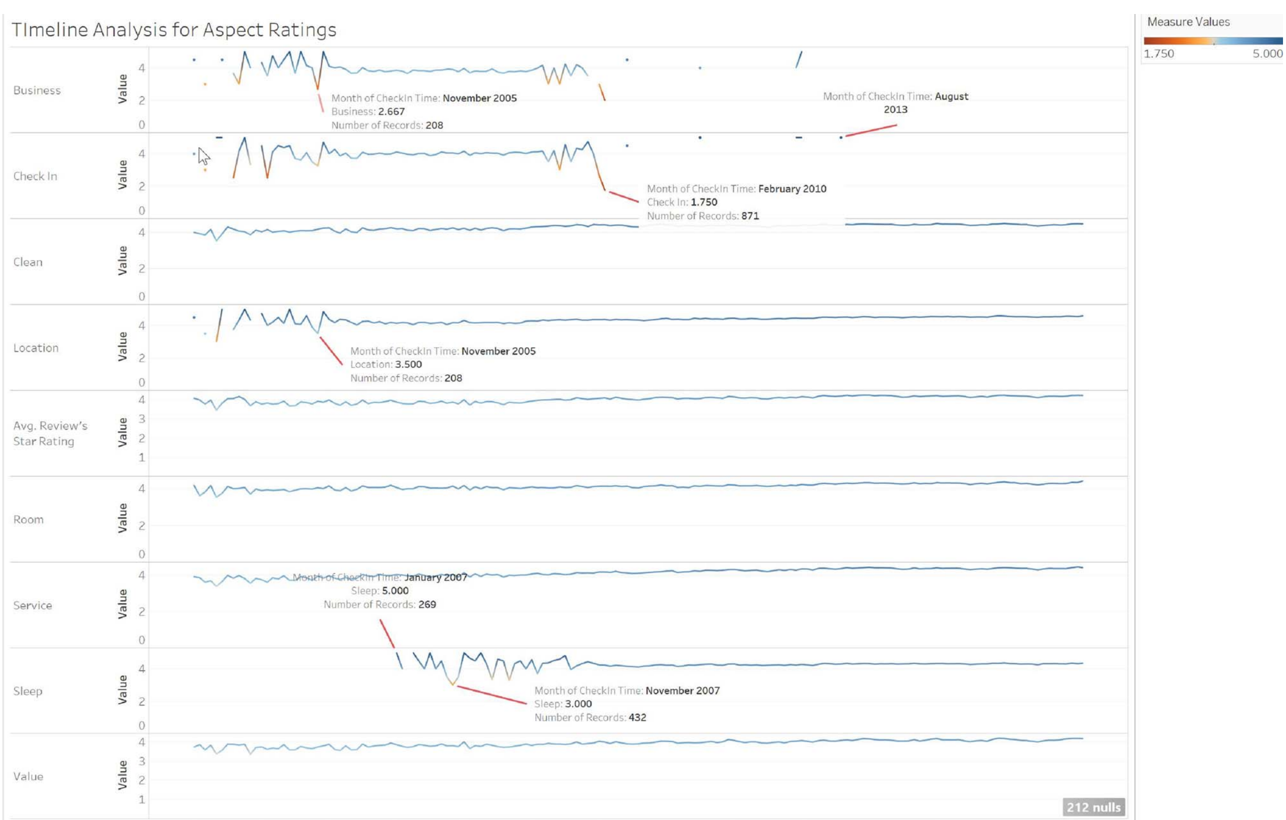

Fig. 13. Timeline analysis on aspects of Hilton hotels.

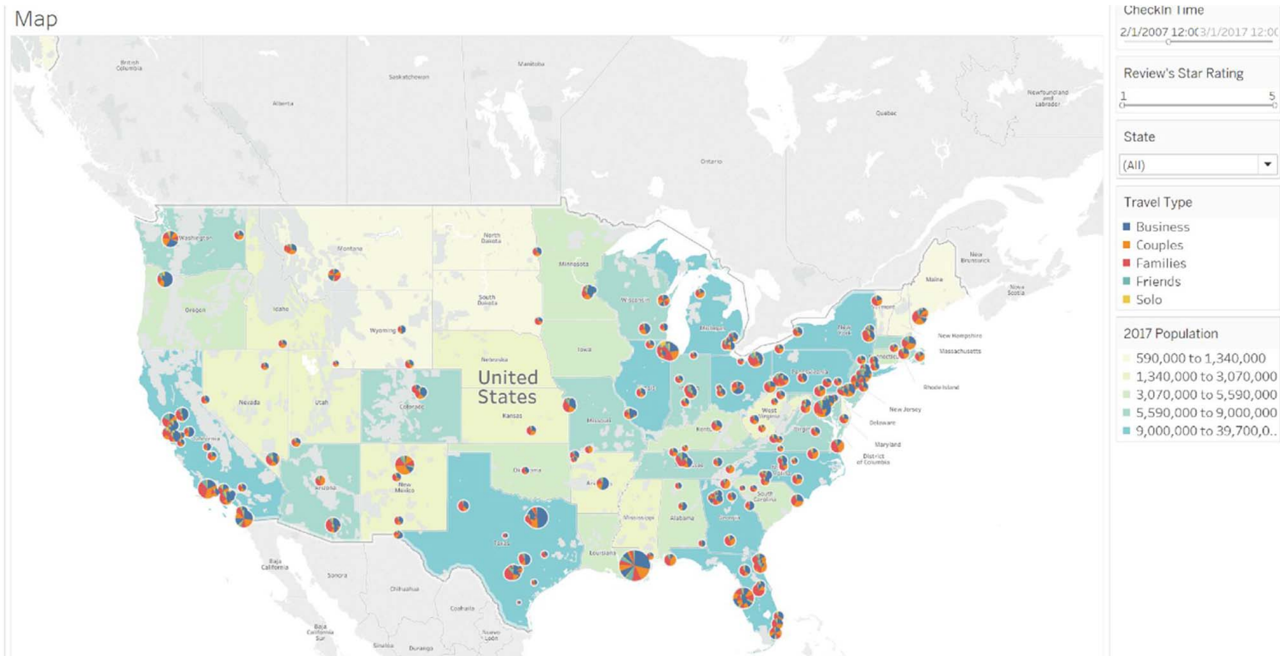

Fig. 14. Geographical analysis of travel types and rating scores.

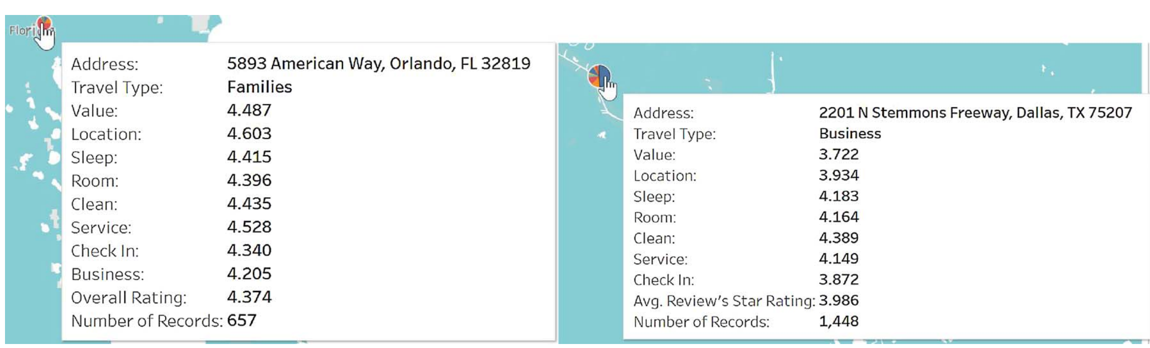

Fig. 15. Geographical analysis of specific hotels on aspect ratings.

show that users also search for Hilton's major competitors, such as Marriott, Hyatt, and Holiday Inn. A comparative analysis reveals that Hilton is still the most popular brand during the sample period (see Fig. 12), which is consistent with the brand value report conducted by Brand Finance.[10] Other related topics include Airbnb, Expedia, TripAdvisor, Trivago, Google Flights, Yelp, etc. This real-time trend may be useful for exploring future opportunities and generating marketing value. For example, if many users are inclined to search TripAdvisor and Yelp together, a hotel manager may consider establishing a partnership with Yelp. Today, social media monitoring is important for brand management (Rout & Bontcheva, 2015). Further, the search volume of TripAdvisor on Google Trends can be a good indicator of hotel managers' response strategies (Lee, Xie, & Besharat, 2016).

5.2. Visual analytics using tableau and word cloud

A timeline analysis is conducted for all average ratings of eight aspects and review star ratings. We filtered the noise, such as check-in time earlier than 2014, because of limited ratings (less than ten ratings in a month) and wrong review years, such as 1899. After data filtering, the total number of data records is 633,866. Fig. 13 shows the overall results. Referring back to Table 1, we notice that the number of aspects on Check-in/Front Desk and Business Service are especially low, but the reason is unknown. Several observations from this timeline analysis are identified. First, the aspect ratings Business Service and Check-in/Front Desk were removed in 2013. More precisely, only sparse ratings were identified on both ratings after ten years. Secondly, the Sleep Quality rating has been introduced since 2007, and we notice a few low ratings in 2007, 2008, and 2009. After 2009, no notable change in rating scores occurred.

To gain deeper insights from the data, we developed an interactive geographical visualization (see Fig. 14). Each pie chart in the map refers to one or more hotel locations. If two Hilton hotels are very close, the result will be combined in a single pie chart. A larger pie chart indicates a higher number of hotel reviews, while the five distinct colors represent five different travel types: Business, Couples, Families, Friends, and Solo travelers. We then filter time and review ratings to focus on the important locations. Overall, Orlando has the most hotel views, followed by California, New York, and Texas. When a single city is selected, the Hilton hotel in Oahu, HI, received the most reviews (9087) between January 2004 and March 2017, followed by the Hilton hotel in Chicago, IL (6458 reviews), New York City (6210 reviews), and Waikoloa, HI (5288 reviews).

The pie charts on the map enable users to explore types of travelers that frequent a hotel and provides an overview of the average ratings

[10] The World's Most Valuable Hotel Brands, http://ftnnews.com/accommodation/30048-the-world-s-most-valuable-hotel-brands.html.

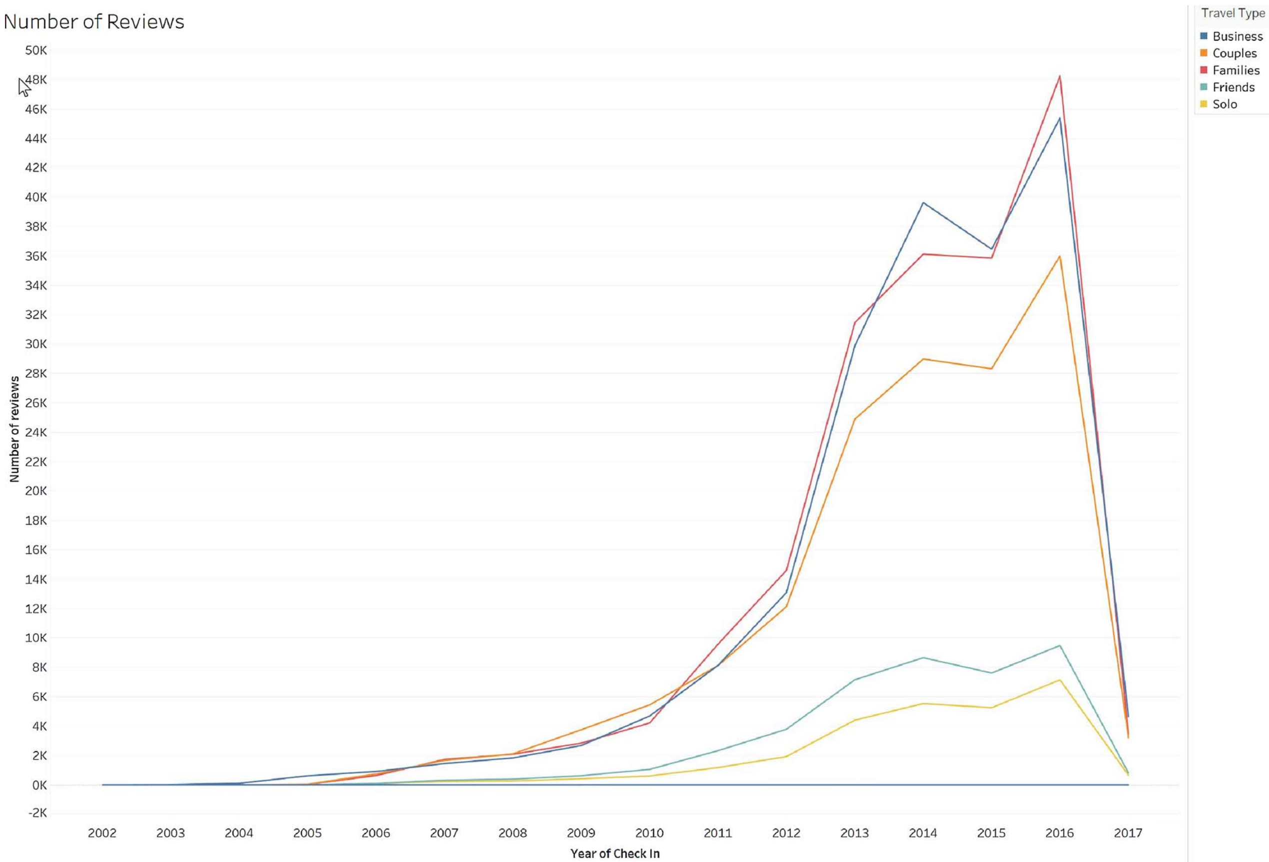

Fig. 16. Timeline analysis of the number of reviews for each travel type.

Travel Type	Clean	Location	Room	Service	Sleep	Value	Overall Rating
Business	4.3502	4.4185	4.1722	4.2841	4.2423	3.9745	4.0818
No filling in	4.2957	4.4105	4.0988	4.1040	4.3436	3.9365	4.0577
Solo	4.4920	4.5166	4.3513	4.3893	4.3901	4.1436	4.2462
Couples	4.4673	4.5284	4.2904	4.3577	4.3364	4.0835	4.2185
Families	4.4095	4.5218	4.2764	4.3376	4.2955	4.0876	4.2045
Friends	4.4827	4.5424	4.3210	4.3613	4.4021	4.1275	4.2476

Measure Values
3.9365 4.5424

Fig. 17. Heatmap analysis of type of travelers and aspects.

for each aspect, e.g., Location, Sleep Quality, Room, Service, Cleanliness, Check-in/Front Desk, Business Service (Internet), Average Star Rating, and Value. For example, Fig. 15 (left) reveals that Families and Couples are two major types of travelers to Honolulu, and the overall rating value is relatively lower compared to some of the aspect ratings. Timothy and Sung-Byung (2015) investigated and compared six hotel aspects: value, location, sleep quality, rooms, cleanliness, and service and found that guests place different importance on each aspect. This further indicates that overall rating cannot be directly used to enhance the quality of service. Therefore, it is important to know the ratings on each aspect. Fig. 15 (right) indicates that more than half of travelers are Business travelers at a Hilton hotel in Texas and Family travelers at a Hilton hotel in Orlando. This is not surprising because Orlando is a well-known city for theme parks, which is a popular destination for family trips (Ya-Han & Kuanchin, 2016). Compared to hotels in Orlando, the hotels in Texas received a relatively lower rating (less than 4 out of 5) on the Average Star rating, Value, and Location. By performing similar analyses, hotel managers have the opportunity to identify the types of travelers in each hotel and examine the ratings for each aspect in order to enhance services and explore marketing opportunities for each hotel.

To know the total number of hotel reviews for each travel type every year, we generated a line chart (see Fig. 16). The majority of reviews came from Business, Families, and Couples. The total number of reviews has increased dramatically since 2010. More specifically, Business and Families leave hotel reviews most frequently. For example, the total number of reviews generated by Families was 48,254 in 2016, followed by Business (45,386) and Couples (36,000). Therefore, the overall rating scores may be strongly influenced by these three types of travelers. To positively influence overall rating scores, knowing the aspect rating for each travel type is also very important.

The relationship between each type of travelers and aspects is explored next. Since the aspects of Check-in/Front Desk and Business Service are no longer available, so we exclude both aspects from this analysis to avoid extreme values. A table with the red-blue-white diverging palette (see Fig. 17) was used to indicate higher (blue) and lower (red) values. Overall, travelers perceive a relatively lower value in Value and a higher value in Location. For each type of traveler, for example, business people pay more attention to Room and Sleep Quality. Similar findings were reported by Timothy and Sung-Byung

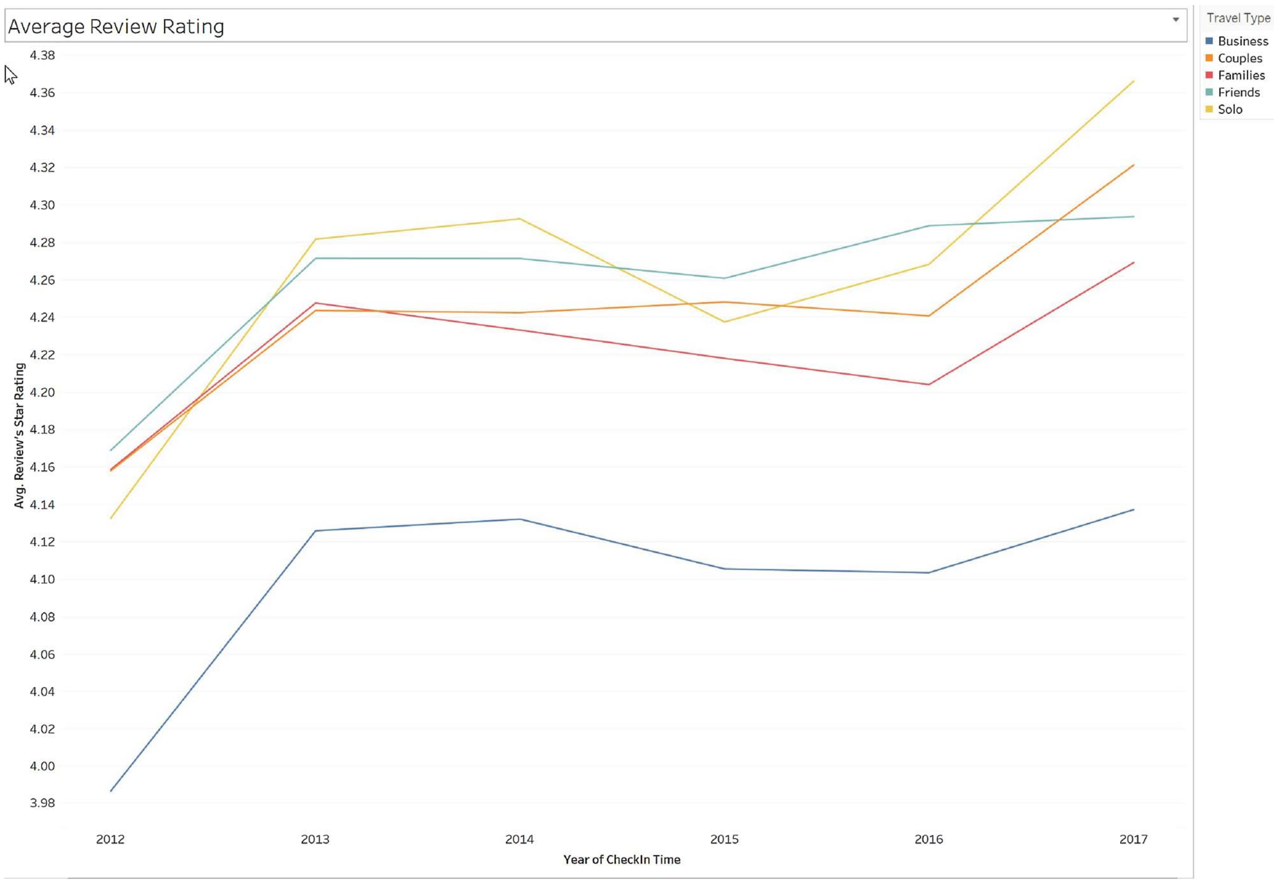

Fig. 18. Timeline analysis of the average review star ratings and each travel type.

(2015); they found that the Room was the most important aspect when hotels received lower ratings from guests. Also, Torres et al. (2015) found that TripAdvisor's overall ratings and the number of reviews are related to the average value of hotel transactions and the popularity of a hotel. However, this study did not take different travel types into account. When we look at the average review ratings for the sample period (see Fig. 18), Business travelers are inclined to rate lowest, followed by Families and Couples, while Solo travelers generally tend to rate relatively higher.

Finally, a time series analysis (see Fig. 19) using line charts was conducted to explore high and low rating values based on travelers' check-in month. We found a similar pattern for each aspect. A higher rating was identified in January, May, September, and December, while a lower rating value was found in July and October. Travelers usually leave a higher rating on Location and Cleanliness, followed by Room, Service, and Sleep Quality. Value was clearly the lowest rated aspect, which may surprise some marketers.

A more comprehensive analysis of Fig. 19 reveals that a greater portion of positive reviews was posted by couples on the Location aspect in December. In contrast, most negative reviews were made by Business travelers on the Room aspect in July. To gain a deeper understanding of reviews, we use word clouds (shown in Fig. 20) to visualize the aspect-sentiment keywords, in which colors correspond to aspect-sentiment categories and sizes are scaled by LLR. In doing so, specific aspect-sentiment can be clearly associated with their descriptions. Results indicate that green (positive) words are in accordance with features, e.g., "shopping," "breakfast," and "restaurants." This indicates that couples pay much more attention to the convenience of nearby shopping locations during their stay. Moreover, couples inclined to give positive ratings for restaurants in a hotel, especially those that serve breakfast. On the contrary, negative-related keywords (in red) are mostly associated with the facilities of a room. For instance, two evident words "A/C" and "carpet" may relate to negative sentiments, e.g., "broken" and "dirty" towards the air conditioner and carpet. A deeper text analysis indicates a considerable number of business travelers' complaints are related to the noise of air conditioner and dirty carpets. Levy, Duan, and Boo (2013) developed a complaint framework and found that front desk service, room cleanliness, bathroom issues, and guestroom noise were frequent complaints. Their findings are similar to our word cloud visualization. Our word cloud further reveals the frequency of those complaints. It is worth noting that the overcrowding frequently caused by tourists in July may lead to longer check-in lines, which may explain the diminished front-desk effectiveness and responsiveness that often causes business travelers' discontent. Consequently, the co-occurrence of terms, e.g. desk and rude, may indicate the negative sentiment towards a service. Fig. 20 also reveals a high correlation between the overall polarity of a hotel review and important aspect-specific keywords. It has been suggested that hotel managers should embrace social networks, monitor UGC, and manage online reputation actively (Levy et al., 2013; O'Connor, 2010). Decision makers can refer to this analysis to support and augment their marketing strategies and activities. For instance, a hotel can establish a partnership with stores to provide an exclusive discount to attract more couple travelers. Moreover, hotel managers should strive to resolve noise problems stemming from air conditioners to minimize business customers' complaints.

6. Conclusion and future research

This paper sheds light on a more inclusive analysis and visualization of reviews from TripAdvisor and data on Google Trends. This study reveals the intricacies of users' ratings, sentiments, and types of

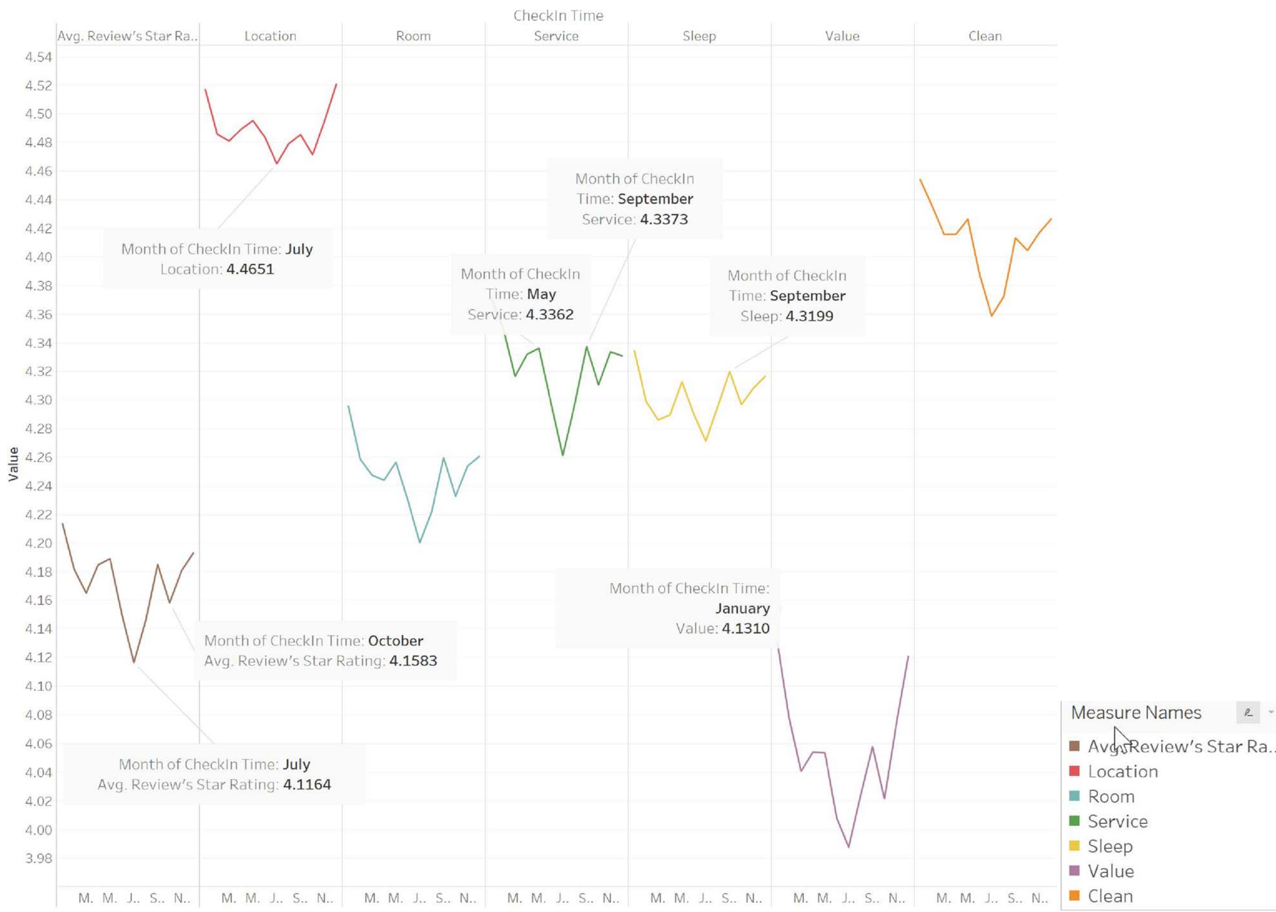

Fig. 19. Timeline analysis of high and low ratings for each aspect.

travelers at different times and locations. It also indicates that a sole source of data, lexicon, and method is not sufficient to discover the genuine value of hotel ratings and reviews. Furthermore, existing research is mostly limited to document-level and sentence-level analysis.

Our major contribution to this study includes a proposed, integrated framework to collect and process data, extract and classify aspect-level information, and finally visualize TripAdvisor and Google Trends data. The proposed aspect-level approach, together with text syntactic information and the CTK, can be justified by its high precision and recall. Our visual analytics provides multiple perspectives by using timeline

Fig. 20. An aspect-sentiment keyword cloud. Colors and their corresponding sentiment: Green—Positive sentiment of Couple travel type on the aspect of Location in December. Red—Negative sentiment of Business travel type on the aspect of Room in July.

and location-based analyses. The experimental results demonstrate that our proposed method outperforms the baseline models. The analyses results can be used to improve the service of each hotel and discover marketing opportunities.

This study is not without limitations, and additional research streams are identified. First, multiple data resources such as Expedia.com, Hotels.com, and Google Reviews can be used in the future, even though no clear, direct evidence suggests that differences of reviews exist across those websites. Secondly, the use of mobile devices for leaving reviews can be incorporated into studies to determine whether device type affects reviews and whether "in the moment" marketing campaigns can overcome reviews and ratings. Mobile devices are increasingly being used for marketing.[11] Red Roof Inn, for instance, turns flight cancellations into marketing campaigns for new customers[12] using a mobile search strategy. Finally, growing concerns have surfaced about fake and paid online reviews (Filieri et al., 2015); therefore, the ratings and reviews may have been manipulated. Future research may benefit from investigating further reviewer profiles, previous reviews, and writing style.

[11] Mobile marketing strategies for hotels, http://www.traveltripper.com/blog/four-important-mobile-marketing-strategies-for-hotels/.

[12] Red Roof Inn Case, http://www.mmaglobal.com/case-study-hub/case_studies/view/31739.

References

Aydoğan, E., & Akcayol, M. A. (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques. *2016 international symposium on innovations in intelligent systems and applications (INISTA),* 1–7.

F.Bangwayo-Skeete, P., & W.Skeete, R. (2015). Can google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management, 46*, 454–464.

Bhattacharjee, K., & Petzold, L. (2016). What drives consumer choices? Mining aspects and opinions on large scale review data using distributed representation of words. *2016 IEEE 16th international conference on data mining workshops (ICDMW),* 908–915.

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

Brody, S., & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. *HLT '10 human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics* (pp. 804–812).

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology,* 2 27:21-27:27.

Chang, Y. C., Chen, C. C., Hsieh, Y.-L., Chen, C. C., & Hsu, W. L. (2015). Linguistic template extraction for recognizing reader-Emotion and emotional resonance writing assistance. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing* (pp. 775–780).

Collins, M., & Duffy, N. (2002). Convolution kernels for natural language. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Vol. Eds.), *Advances in neural information processing systems: vol. 14*, (pp. 625–632). MIT Press.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance, 66*, 1461–1499.

Eivind, B., Burnett, T. H., & Nørvåg, K. (2012). *A study of opinion mining and visualization of hotel reviews.* 229–238: ACM.

Eriksson, N., & Fagerstrøm, A. (2017). The relative impact of wi-fi service on young consumers' hotel booking online. *Journal of Hospitality & Tourism Research* 1096348017696844.

Fan, W., & Gordon, M. D. (2014). *The power of social media analytics. Communications of the ACM*, *Vol. 57*, New York, NY: ACM74–81.

Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). *Tapping the power of text mining. Communications of the ACM*, *Vol. 49*, New York, NY: ACM76–82.

Fenichel, E. P., Kuminoff, N. V., & Chowell, G. (2013). Skip the trip: Air travelers' behavioral responses to pandemic influenza. *PUBLIC LIBRARY OF SCIENCE, 8*, e58249.

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.

Gamon, M. (2004). Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis. *COLING '04 proceedings of the 20th international conference on computational linguistics*.

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels − an empirical analysis. *Tourism Management, 61*, 43–54.

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys, 49* 28:21-28:41.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Herrero, Á., San Martín, H., & Hernández, J. M. (2015). How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. *International Journal of Contemporary Hospitality Management, 27*, 1573–1597.

Holsapple, C., Hsiao, S.-H., & Pakath, R. (2014). Business social media analytics: Definition, benefits, and challenges. *AMCIS 2014 proceedings*.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *KDD '04 proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177).

Huang, S., Peng, W., Li, J., & Lee, D. (2013). Sentiment and topic analysis on social media: A multi-task multi-label classification approach. *WebSci '13 proceedings of the 5th annual ACM web science conference* (pp. 172–181).

Hutchin, E. (2016). *How Can Google Trends and Sentiment Analysis of Tripadvisor and Facebook Predict Visitor Numbers to the United Kingdom and Canada from 2010 to 2015.* Dublin: National College of Ireland [Unpublished Masters].

Jakob, N., & Gurevych, I. (2010). Extracting opinion targets in a single- and cross-domain setting with conditional random fields. *EMNLP '10 proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 1035–1045).

Jiang, P., Zhang, C., Fu, F., Niu, Z., & Yang, Q. (2010). An approach based on tree kernels for opinion mining of online product reviews. *2010 IEEE international conference on data mining* (pp. 256–265).

Jin, L. Y., Karen, X., Ali, B., & Yong, T. (2017). Management responses to online WOM: Helpful or detrimental? *SSRN Electronic Journal,* 44.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *ECML '98 proceedings of the 10th european conference on machine learning* (pp. 137–142).

Lafferty, J., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML '01 proceedings of the eighteenth international conference on machine learning* (pp. 282–289).

Lak, P., & Turetken, O. (2014). Star ratings versus sentiment analysis ?A comparison of explicit and implicit measures of opinions. *2014 47th hawaii international conference on system sciences* (pp. 796–805).

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *ICML'14 proceedings of the 31 st international conference on international conference on machine learning: 32,* (pp. 1188–1196).

Lee, Y., Xie, K., & Besharat, A. (2016). Management response to online WOM: Helpful or detrimental? *Amcis 2016* (pp. 1–10).

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

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57–66.

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

Liu, K., Xu, L., & Zhao, J. (2015). Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. *IEEE Transactions on Knowledge and Data Engineering, 27*, 636–650.

Liu, B. (2012). *Sentiment analysis and opinion mining.* San Rafael: Morgan & Claypool Publishers.

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*, 119–154.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (1st ed.). New York: Cambridge University Press.

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *Learning for text categorization: Papers from the 1998 AAAI workshop* (pp. 41–48).

Moschitti, A. (2004). A study on convolution kernels for shallow semantic parsing. *ACL '04 proceedings of the 42nd annual meeting on association for computational linguistics* (pp. 8). .

O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management, 19*, 754–772.

Paltoglou, G., & Thelwall, M. (2012). Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology, 3* 66:61-66:19.

Popescu, A.-M., & Etzioni, O. (2005). Extracting product features and opinions from reviews. *HLT '05 proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 339–346).

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*, 130–137.

Rhee, H. T., Yang, S.-B., & Kim, K. (2016). Exploring the comparative salience of restaurant attributes: A conjoint analysis approach. *International Journal of Information Management, 36*, 1360–1370.

Ribarsky, W., Xiaoyu Wang, D., & Dou, W. (2014). Social media analytics for competitive advantage. *Computers & Graphics, 38*, 328–331.

Rout, D., & Bontcheva, K. (2015). A human-annotated dataset for evaluating tweet ranking algorithms. *HT '15 proceedings of the 26th ACM conference on hypertext & social media* (pp. 95–99).

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B. C., Ellis, G., & Keim, D. A. (2014). Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics, 20*, 1604–1613.

Shelke, N., Deshpande, S., & Thakare, V. (2017). Domain independent approach for aspect oriented sentiment analysis for product reviews. *Proceedings of the 5th international conference on frontiers in intelligent computing: Theory and applications* (pp. 651–659).

Singh, V. K., Piryani, R., Uddin, A., & Waila, P. (2013). Sentiment analysis of movie reviews: A new feature-Based heuristic for aspect-Level sentiment classification. *2013 international mutli-conference on automation, computing, communication, control and compressed sensing (iMac4s)* (pp. 712–717).

Timothy, R. H., & Sung-Byung, Y. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior, 50*, 576–587.

Torres, E. N., Singh, D., & Robertson-Ring, A. (2015). Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. *International Journal of Hospitality Management, 50*, 77–83.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *HLT '05 proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347–354).

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How google searches foreshadow housing prices and sales. *SSRN Electronic Journal,* 89–118.

Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., et al. (2010). OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics, 16*, 1109–1118.

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

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.

Xu, X., & Reed, M. (2017). Perceived pollution and inbound tourism in China. *Tourism Management Perspectives, 21*, 109–112.

Ya-Han, H., & Kuanchin, C. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management, 36*, 929–944.

Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research, 53*, 433–447.

Yu, J., Zha, Z.-J., Wang, M., & Chua, T.-S. (2011). Aspect ranking: Identifying important product aspects from online consumer reviews. *HLT '11 proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies: 1,* (pp. 1496–1505).

Yu, H. F., Ho, C. H., Juan, Y. C., & Lin, C. J. (2013). LibShortText: A Library for Short-text Classification and Analysis LibShortText: A Library for Short-text Classification and Analysis. In 1–5.

Zhang, Z., Li, X., & Chen, Y. (2012). Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews. *ACM Transactions on Management Information Systems (TMIS), 3* 5:1-5:23.

Zhang, Z., Guo, C., & Goes, P. (2013). Product comparison networks for competitive analysis of online word-of-Mouth. *ACM Transactions on Management Information Systems (TMIS), 3* 20:21-20:22.

Zhao, W. X., Jiang, J., Yan, H., & Li, X. (2010). Jointly modeling aspects and opinions with a maxEnt-LDA hybrid. *EMNLP '10 proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 56–65).

Zhao, D., Ye, Q., & Zhu, K. (2016). The influence of consumers' information search behavior on purchase decisions for experience goods: Empirical evidence from hotel industry. *WHICEB 2016 proceedings* (pp. 297–302).