G Model
INFMAN 2967 No. of Pages 17

ARTICLE IN PRESS

Information & Management xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Information & Management

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

The effect of user-controllable filters on the prediction of online hotel reviews

Ya-Han Hu[a], Kuanchin Chen[b], Pei-Ju Lee[c,*]

[a] Department of Information Management, National Chung Cheng University, Chiayi, 62102, Taiwan, ROC
[b] Department of Business Information Systems, Western Michigan University, 3344 Schneider Hall, Kalamazoo, MI 49008-5412, United States
[c] Department of Information Management, National Chung Cheng University, Chiayi, 62102, Taiwan, ROC

ARTICLE INFO

Article history:
Received 1 June 2016
Received in revised form 16 November 2016
Accepted 24 December 2016
Available online xxx

Keywords:
Online hotel review
Review helpfulness
Sentiment analysis
Review extremity
Prediction
Electronic word of mouth (eWOM)

ABSTRACT

Product reviews have gained much popularity in recent years. This study examines the theoretical foundation of review helpfulness and reports how the interactions among three user-controllable filters together with three groups of predictors affect review helpfulness. Reviews from TripAdvisor.com were analyzed against three analytical models. The results show that these groups of variables have a varying effect on different user-controllable filters. Review rating and number of words are key predictors of helpfulness across all three filters. The recency, frequency, and monetary (RFM) model has received a consistent support across all filters as well. Managerial implications are provided.

1. Introduction

With the rapid development of the Internet, online social media has become a popular platform for users to share their personal experiences. The shared information, termed user-generated content (UGC) in the academic literature, is the first source of information for many people to make their decisions [1,2]. Of all types of UGC, online consumer reviews represent the majority for purchase decisions. Online review websites are a major channel of communication that provides valuable information to consumers [3,4]. These websites also tap into online reviews for the opportunity of promotions, customer service, and other revenue-generating activities [5]. Studies have shown that travel websites greatly influence the tourism industry; 62% of travelers search the Internet for their upcoming travel activities and 43% of visitors read online reviews written by other travelers [6–9].

Although the growth in the volume of online hotel reviews is a welcomed trend for consumers, it also likely causes information overload for those who wish to meaningfully use it. Travelers must manually filter helpful reviews on travel websites, which considerably increases the search cost to locate hotel reviews helpful to meet their goals. This is the reason many product websites offer review helpfulness to help readers sort through a sea of reviews. Hotel reviews are not an exception. Most hotel review websites also provide some form of helpfulness indicators as well. Assessment of helpful reviews is, therefore, an important and essential task for consumers [10–12]. Review helpfulness typically refers to the total number or percentage of positive votes a product review has received; it represents a consumer's analysis of how the review matches the expectations for the trip in mind [13].

Online tourism websites with more helpful reviews can provide more valuable information to potential customers. Therefore, the development of an automatic review evaluation system to identify high-quality reviews on tourism websites can both reduce the search time for a consumer to locate the desired information, and facilitate the creation of diversified services compared with those of the existing websites. Therefore, online review helpfulness has become a key variable of interest in the product review literature that spans across multiple disciplines.

Mudambi and Schuff [12] were among the first to provide a theoretically grounded explanation of review helpfulness. They concluded with a model where the construct relationships vary between search goods and experience goods. Therefore, the requirements for a review to be considered helpful are not quite the same between the two types of goods. This is consistent with the literature where experience goods are defined as goods that require consumers to sample or "experience" the product before

* Corresponding author.
E-mail address: pjlee@mis.ccu.edu.tw (P.-J. Lee).

http://dx.doi.org/10.1016/j.im.2016.12.009
0378-7206/© 2016 Elsevier B.V. All rights reserved.

formulating their own quality assessment, but quality assessment could be conducted for search goods before they are even purchased [14]. Similarly, readers of hotel reviews expect to learn from other people's experience with the hotel, which makes hotels an experience good. Although Mudambi and Schuff's model that predicts helpfulness with review rating, word count, and total votes are a great foundation model, it was not specifically designed for experience goods or more specifically for hotel reviews.

In fact, their theoretical basis of review helpfulness was information diagnosticity from Jiang and Benbasat [15] and others, which suggest that diagnosticity is highly desired when the salient product attributes are better assessed through experiences. Additionally, the accessibility–diagnosticity model indicates that "accessible information is not used as an input for judgement and choice when more diagnostic or probative information is available" ([16] Herr et al., 1991; p. 457). Therefore, factors in addition to review rating, word count, and total votes are also cues to enrich Mudambi and Schuff's model for hotel reviews.

Moreover, studies of product reviews have traditionally focused on searching for an optimal set of predictors of review helpfulness, but neglecting the fact that even the predictors may interact with each other. For example, climate and seasonal shifts may affect tourism demand [17–21], which points to a possible interaction between travel season and geographic location of hotels. This is the reason several travel websites have offered filters of hotel reviews based on these characteristics. Because of the availability of these filters, the visibility of a review may be altered through the selection of a filter. In the end, it affects a review's opportunity to be voted on for review helpfulness [11]. If interactions among predictor variables are not accommodated in a theoretical model for hotel reviews, the predictive power or even the accuracy of the model may be hampered.

Based on the above assessment, the present study is designed with the following objectives:

1. To enrich the theoretical model of Mudambi and Schuff with additional predictor variables from the relevant literature.
2. To provide empirical evidence of interaction effects for the common filters of hotel reviews (i.e., travel regions, travel seasons, and travel types on review helpfulness).
3. To improve the performance of review helpfulness prediction models by considering the above two objectives.

2. Related work

2.1. Factors affecting review helpfulness

Table 1 summarizes the predictors of review helpfulness from the literature. The predictors used in these studies can be divided into the following three categories: review quality (i.e., review content and review readability), review polarity (i.e., review sentiment and review subjectivity), and reviewer (i.e., reviewer characteristics and RFM (recency, frequency, and monetary) features). Such a classification is rooted in both the diagnosticity and electronic word-of-mouth (eWOM) literatures. For example, Wang et al.'s [49] finding of informant credibility supports that the characteristics of the information provider (i.e., product reviewer) are related to acceptance of a product. Similarly, Li et al. [50] also

Table 1
Previous studies on review helpfulness.

Work	Data source	Search (S)/Experience (E) goods	Review quality		Review polarity		Reviewer	
			Review content	Readability	Sentiment	Subjectivity	Reviewer characteristics	RFM
Kim et al. [22]	Amazon	S/E	✓		✓			
Liu et al. [23]	Amazon	S	✓	✓	✓			
Forman et al. [24]	Amazon	E	✓	✓		✓		
Zhang [25]	Amazon	S/E	✓			✓		
Liu et al. [26]	IMDB	E	✓				✓	
Otterbacher [27]	Amazon	S/E	✓				✓	
O'Mahony and Smyth [28]	TripAdvisor	E	✓				✓	
Mudambi and Schuff [12]	Amazon	S/E	✓					
Chen and Tseng [29]	Amazon	S	✓		✓		✓	
Ghose and Ipeirotis [30]	Amazon	S/E	✓	✓		✓	✓	
Yu et al. [31]	IMDB	E	✓		✓			
Ngo-Ye and Sinha [32]	Amazon	E	✓					
Liu et al. [33]	Amazon	S	✓		✓	✓		
Dong et al. [34]	Amazon	S	✓	✓	✓			
Ngo-Ye and Sinha [35]	Amazon/Yelp	E E	✓					✓
Hu et al. [36]	Amazon	E	✓		✓			
Hwang et al. [37]	TripAdvisor	E	✓		✓			
Yin et al. [38]	Yelp	E	✓				✓	
Lee and Choeh [39]	Amazon	S	✓				✓	
Martin and Pu [40]	Amazon/Yelp/TripAdvisor	S E E	✓	✓	✓			
Zhu et al. [41]	Yelp	E	✓	✓			✓	
Liu and Park [42]	Yelp	E		✓			✓	
Weathers et al. [43]	Amazon	S/E	✓				✓	
Huang et al. [44]	Amazon	S	✓				✓	
Ahmad and Laroche [45]	Amazon	S	✓		✓			
Chua and Banerjee [46]	Amazon	S/E	✓	✓				
Fang et al. [47]	TripAdvisor	E	✓	✓	✓		✓	
Hu and Chen [11]	TripAdvisor	E	✓	✓	✓	✓	✓	
Qazi et al. [48]	TripAdvisor	E	✓		✓	✓	✓	

indicate that source credibility, product review content, and authorship are key predictors for product review helpfulness.

Review quality directly affects review helpfulness; previous studies have explored the influence of review content on review helpfulness from different aspects, including the length of the review, review structure, readability, and writing style. Moreover, sentiment analysis has become a popular research topic and is widely used in online review analysis. Sentiment analysis determines consumers' sentiment polarity (i.e., positive, negative, or neutral sentiment) according to consumers' reviews of products or service experiences. Many studies have also explored the relationship between sentiment strength and review helpfulness [12,51]. In addition, evidence suggests that reviewer characteristics (such as reviewer reputation) affect review helpfulness [27,48,52,53]. Details of individual categories of these predictors are presented in the following sections.

2.1.1. Review quality

A review with adequate readability is considered to be more helpful to users than a review that is difficult to read and contains numerous typographical errors. However, although many reviewers are not concerned about spelling accuracy when writing online reviews, their spelling errors may cause difficulties when reading. Kim et al. [22] considered three types of features, namely term frequency–inverse document frequency (TF–IDF) scores, review length, and review readability, for explaining review helpfulness. Forman et al. [24] revealed that review readability has a positive effect on review helpfulness and that spelling errors have a negative effect on review helpfulness.

2.1.2. Review polarity

Sentiment analysis has become a popular research topic in recent years; many studies have been investigated on the relationship between sentiment strength and review helpfulness using different review data sets [51,54,55,56]. Hu et al. [36] explored the relationship between evaluation, sentiment, and sales amount through 4405 book reviews on Amazon.com; the results revealed that sentiment features affect the overall sales significantly. Mudambi and Schuff [12] analyzed 1587 reviews of consumer electronics on Amazon.com and reported that sentiment strength strongly affects review helpfulness, particularly for search goods. Ghose and Ipeirotis [30] learned from product reviews on Amazon.com that subjectivity, informativeness, readability, and linguistic correctness of reviews can significantly affect review helpfulness. They also reported that reviews containing a mix of subjective and objective statements were considered more helpful to users. The information-quality framework was used by Chen and Tseng [29] to evaluate the quality of product reviews. They concluded that high-quality reviews tend to be extremely subjective. Furthermore, Cao et al. [53] suggested that the stronger the sentiment polarity of reviews, the higher is the chance of them being considered helpful for consumers. In addition, semantic features are more critical than the basic stylistic characteristics. In semantic element selection, Korfiatis et al. [57] demonstrated that the sentiment feature has a stronger influence than the length of reviews. Chua and Banerjee [46] considered review helpfulness as a ratio of the number of helpful votes over the total number of votes; they defined review sentiment as favorable or unfavorable and influenced by review star ratings, which affect review helpfulness according to product type. Review star rating is positively related to more favorable reviews. Ahmad and Laroche [45] further divided the review sentiment into positive and negative; they confirmed that reviews using stronger sentiment words are considered to more greatly affect review helpfulness (i.e., positive or negative words have a positive or negative effect on review helpfulness) than reviews with vague or neutral words.

2.1.3. Reviewer

On most social media websites, reviewers' personal information is available to users for reference. Previous studies have noted that consumers tend to believe the helpfulness of a review when travel websites incorporate the information of reviewer identity and personal characteristics in the review [42,52,58,13]. Accurately identifying reviewers with greater reputations is essential for consumers when deciding which review to believe. [13] Hu et al. suggested that consumers consider evaluation scores, reviewer reputation, and the number of past reviews of reviewers. Reviews written by reputed reviewers or by reviewers who have written extensive reviews on the social media websites are more prominent. Otterbacher [27] studied product reviews on Amazon.com and used the total votes, badge type, and website rankings of the reviewer to evaluate reviewer reputation.

Other studies have also revealed that the past average votes of the reviewer can be used to predict review helpfulness [47]. Qazi et al. [48] also addressed that the qualitative factors of reviewers, such as review type and cumulative helpfulness, are as equally critical as the quantitative factors, such as the number of concepts. Although the qualitative factor such as the length of the review has a diminishing effect on review helpfulness when the review length exceeds 144 characters, other qualitative factors remain influential [44]. The combination of reviews and reviewers provides a more complete explanation of review helpfulness to consumers [42]. Weathers et al. [43] considered three factors—reviewer credibility, review diagnosticity, and product type—and concluded that review credibility refers to reviewer characteristics (e.g., trustworthiness and expertise) and review diagnosticity refers to review quality (e.g., uncertainty and equivocality).

Past studies have shown that gender and age affect how consumers react to online reviews. Females and younger people had higher intention to engage in eWOM. For example, the e-generation is immersed in technology growing up. The pervasiveness of personal technology has made them more open and comfortable to electronic communications. Moreover, there is also a gender difference in their interest in review content and preference over stylistic choices [58,59]. Therefore, the metadata such as gender and age of reviewers are also collected; these factors provide the background knowledge of reviewers that helps us to profile these reviewers, and address the impact on population [27,60,61].

The RFM model is a valuable tool for measuring the values of customers, and it aims to predict consumers' future purchasing according to their past transactions [62]. This model functions by combining the three indicators (i.e., recency, frequency, and monetary values) of consumers; these indicators describe reviewers who have recently made a transaction, purchase frequently, and, on average, spend more money. The RFM model combines the three indicators, groups consumers according to their behaviors, and provides different marketing strategies for distinct consumer groups [63,64]. This model has been widely used in direct marketing, such as in sending direct messages and e-mails.

Ngo-Ye and Sinha [32] were the first to apply the RFM model to evaluate the effect of reviewer engagement on review helpfulness prediction. The authors collected reviews from Yelp.com and Amazon.com and considered both textual (i.e., bag-of-words method) and RFM features of the reviews. Their results confirmed that the variables in the RFM model provide a strong explanation of review helpfulness.

2.2. Previous studies on hotel review helpfulness

Although many studies have addressed online review helpfulness, few have focused on online hotel reviews. O'Mahony and

Smyth [28] collected reviews of Las Vegas and Chicago hotels on TripAdvisor.com. They defined helpful reviews as those that received more than 75% positive feedback votes of all votes. They considered reviewer reputation, content, social, and sentiment features as independent variables (IVs). Their experimental results indicated that reviewer reputation is the predictor for classifiers to achieve favorable predictive ability.

Ghoses et al. [10] developed a hotel recommendation system that considered consumer heterogeneity, hotel characteristics, purchasing history, and online hotel reviews for hotel evaluation. Their results revealed that the readability and subjective features of hotel reviews can be used to predict review helpfulness effectively.

Hwang et al. [37] collected 3124 reviews of Taiwan hotels from TripAdvisor.com and divided the reviews into two categories (i.e., helpful/not helpful) according to two hotel managers. They considered three types of features—content, sentiment, and review quality—and applied three types of methods (i.e., TF–IDF, topic model-based latent Dirichlet allocation (LDA), and semantic-based LDA) in the selection of IVs. The results confirmed that the three content features are major predictors and the topic model-based LDA provided the most favorable performance in the predictive model.

Yin et al. [38] collected 16,269 reviews of 307 San Francisco hotels from Yelp.com. They defined review helpfulness as the total number of helpful votes received by a review and determined the sentiment features by matching review contents with the sentiment lexicon (i.e., using the revised Dictionary of Affect in Language). In addition, their study considered other features, such as the review ratings, review length, reviewer characteristics, and hotel information in experimental evaluation. The results indicated that review sentiment, review length, and reviewer characteristic features can be critical factors in predicting review helpfulness.

Zhu et al. [41] also accumulated hotel reviews from Yelp.com and provided a definition of review helpfulness identical to that of Yin et al. [38]. The main objective of Zhu et al. [41] was to investigate the relationship between reviewer credibility and review helpfulness and to examine the moderation effects of review extremity and hotel price. A total of 16,265 hotel reviews were collected. Two independent variables (i.e., reviewer expertise and online attractiveness) were selected, and several review and hotel features were used as control variables in their study. The results indicated that opinion leaders do not necessarily receive higher helpful votes; the effects of both reviewer expertise and online attractiveness were moderated by hotel price.

Hu and Chen [11] examined the interaction effects of star-class hotels and review ratings on review helpfulness prediction. They confirmed that users are likely to read the reviews when their review rating does not correspond with users' expectations on star-class hotels and, therefore, the likelihood of these reviews being voted is increased. Hu and Chen [11] collected 349,582 reviews of 450 hotels in Orlando or Las Vegas from TripAdvisor.com and considered four types of features: review content, review sentiment, review author, and review visibility. The results revealed that review visibility has a strong effect on review helpfulness.

In summary, the abovementioned studies have shown that the review quality, review polarity, and reviewer characteristics are important factors to predict hotel review helpfulness, which also provide some evidences that review diagnosticity, review quality, and reviewer credibility significantly affect review helpfulness of experience goods. However, most of these studies overly focus on only locating the best set of predictors, neglecting a possibility for interaction among the predictors. Consequently, the predictive power of the resulting model may be inaccurate or biased. In addition, user-controllable filters (i.e., travel region, travel season, and travel type) may artificially alter the visibility of reviews, thereby affecting their opportunities to be voted on for helpfulness. Such dynamics have received very little attention in the literature.

3. Research method

Fig. 1 illustrates the research process. The hotel reviews were collected from TripAdvisor.com. The preprocess procedures, including review length calculations, word and sentence segmentation, and part-of-speech (POS) tagging, were performed. A total of 39 features, including one dependent variable (DV) and 38 IVs,

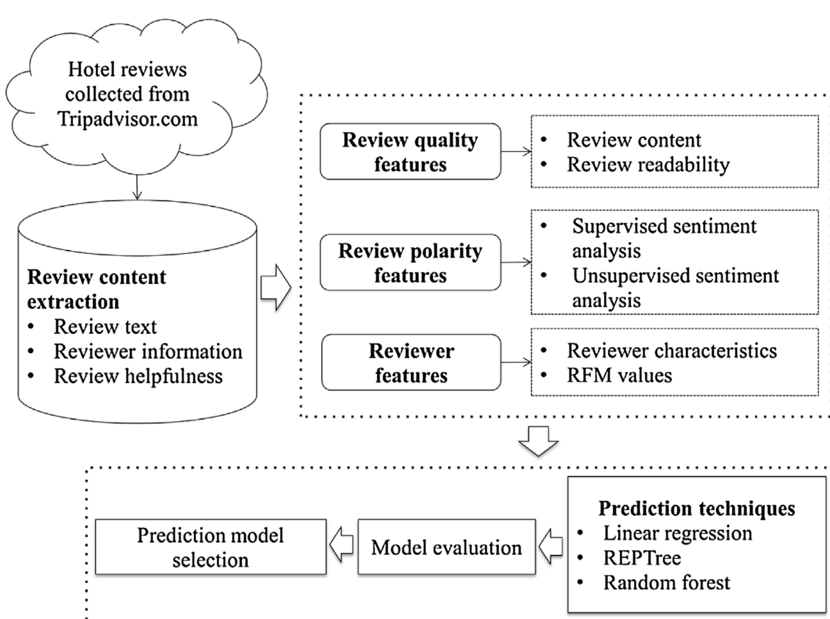

Fig. 1. Research process.

were considered. The DV was review helpfulness, which is defined as follows:

$$\text{Reviewhelpfulness}_i = \frac{\text{HelpfulVotes}_i}{\text{Elapsed}_{\text{Month}_i}} \quad (1)$$

where HelpfulVotes_i is the number of votes the review i has received for its helpfulness, $\text{Elapsed}_{\text{Month}_i}$ is the number of months since review i has been posted, that is, the difference between the date of review posted and the date the review was crawled. Usually, the longer a review has been posted, the more likely it will receive helpfulness votes. Dividing helpfulness votes by elapsed time will reduce the effect of systematic bias due to this above-review longevity issue. Previous research [41] has also shown that elapsed time is related to review helpfulness.

The IVs were divided into the following categories: review quality, review polarity, and reviewer features. Three supervised learning techniques were used for developing prediction models, namely linear regression (LR), reduced error-pruning tree (REPtree), and random forest (RF).

3.1. Review collection and preprocessing

The present study collected complete sets of hotel reviews of five famous travel destinations in the United States—New York City, Las Vegas, Chicago, Orlando, and Miami—from TripAdvisor.com. According to the data on the Travelers' Choice of 2015 provided by TripAdvisor.com, the selected five cities were voted by millions of travelers as the top five most popular travel destinations in the United States; they were also listed as the top five cities in the 2016U.S. Place Equity Index Resonance Report. The complete sets of hotel reviews of the five cities were collected between June 1, 2012, and May 31, 2015. Before retrieving the features from the reviews, several data-filtering tasks were performed. First, to simplify the analysis, we collected reviews written only in English. Second, because this study investigated the interaction effects of review region, review season, and travel type on review helpfulness, the reviews without such information (i.e., missing data) were removed; the selected five cities were also grouped according to three travel regions: north (New York City/Chicago), south (Miami/Orlando), and west (Las Vegas) regions. Consequently, the descriptive statistics of the selected hotel reviews of the five cities are presented in Table 2.

The content of each review comprised review rating, review title, review content, reviewer information, number of helpful votes, and the date the review was posted. Before extracting review features from the review content, several preprocessing tasks were performed. First, we used Google's spell-check function to rectify spelling errors in the collected reviews. We used Stanford CoreNLP for other preprocessing tasks such as word and sentence segmentation as well as POS tagging [65]. After segmentation, the POS tagging was assigned to each word according to the meaning of adjacent words in the same sentence; the tags included those for nouns (N), adjectives (JJ), or adverbs (RB).

Table 2
Hotel review data sets from TripAdvisor.com.

Region	City	# of Hotels	# of Reviewers	# of Reviews
North	New York City	401	200,313	227,931
	Chicago	147	71,116	79,436
South	Miami	96	25,942	28,197
	Orlando	301	125,783	147,179
West	Las Vegas	213	198,740	234,259
	Total	1158	621,894	717,002

3.2. Selected review features

Table 3 details the investigated research variables and their definitions. On the basis of previous studies on review helpfulness, this study considered three types of IVs: review quality, review polarity, and reviewer.

3.2.1. Review quality

The information contained in each review includes review rating, the date the review was posted, review title, reviewer information, and the number of helpful votes. We collected the overall ratings of hotels from every review (RATING), ranging from 1 (*extremely negative*) to 5 (*extremely positive*). To estimate the required education level for the reader to comprehend the review (i.e., content readability), we analyzed the total length and the average length of every review with different units in characters, syllables, words, and sentences. On the basis of the collected hotel reviews, variables associated with the review length were considered, which include the number of characters (LENGTH_CHAR), the number of syllables (LENGTH_SYLL), the number of words (LENGTH_WORD), the number of sentences (LENGTH_SENT), the average number of syllables per word (SYLL_PER_WORD), and the average number of words per sentence (WORD_PER_SENT).

Some studies have indicated that readers' reading abilities and speeds increase with an increase in review readability [30,40]. Content readability is estimated through various measurements; in most cases, researchers calculate content readability according to the corresponding education level. To prevent the bias of using a single measurement, we calculated a set of readability indices for each review, including the Automated Readability Index (ARI) [66], Coleman–Liau Index (CLI) [67], Flesch Reading Ease Scale (FRES) [68], Flesch–Kincaid Grade Level (FGL) [68], Gunning Fog Index (FOG) [69], and Simple Measure of Gobbledygook (SMOG) [70]. For more information on these readability indices, please refer to [11].

In addition to the aforementioned content features, the subjectivity of each review was also analyzed. This study adopted Opinionfinder [71,72], one of the most widely used text-mining tools, to identify subjective sentences from texts automatically. Opinionfinder uses a classifier trained with machine-learning methods. On the basis of the extracted sentences in reviews, subjective analysis can be performed. The results identify more subjects when the number of subjective sentences in the review is higher. The degree of subjectivity (SUBJECTIVITY) can be defined as the ratio of the number of subjective sentences to the total number of sentences in the review (Eq. (2))

$$\text{SUBJECTIVITY} = \frac{\text{Sub}}{\text{Sentence}} \quad (2)$$

where Sub denotes the number of subjective sentences in the review and Sentence denotes the total number of sentences after punctuation.

3.2.2. Review sentiment

In addition to the subjectivity analysis, Opinionfinder can also perform unsupervised sentiment analysis. The review texts were separated into sentences according to punctuation marks automatically. Each word was assigned a POS tag, such as nouns (N), adjectives (JJ), or adverbs (RB). Thereafter, polarity classifiers were used to identify the word polarity for each word. The types of review polarity were strong positive (STR_POS), strong negative (STR_NEG), strong sentiment (STR_SENTI), weak sentiment (WEAK_SENTI), weak positive (WEAK_POS), and weak negative (WEAK_NEG). This study calculated the scores of each type of word

Table 3
List of IVs.

Variable category	Variable Name	Description	Type (value/range)
Review Quality	RATING	Review rating	Numeric (1–5)
	LENGTH_CHAR	Number of characters	Numeric
	LENGTH_SYLL	Number of syllables	Numeric
	LENGTH_WORD	Number of words	Numeric
	LENGTH_SENT	Number of sentences	Numeric
	SYLL_PER_WORD	Average number of syllables per word	Numeric
	WORD_PER_SENT	Average number of words per sentence	Numeric
	ARI	Automated Readability Index	Numeric
	CLI	Coleman–Liau Index	Numeric
	FRES	Flesch Reading Ease Scale	Numeric
	FGL	Flesch–Kincaid Grade Level	Numeric
	FOG	Gunning Fog Index	Numeric
	SMOG	Simple Measure of Gobbledygook	Numeric
Review polarity	SUBJECTIVITY	Number of subjective sentences/Number of sentences	Numeric
	STR_POS	Strong positive score	Numeric
	STR_NEG	Strong negative score	Numeric
	WEAK_POS	Weak positive score	Numeric
	WEAK_NEG	Weak negative score	Numeric
	STR_SENTI	Strong sentiment score	Numeric
	WEAK_SENTI	Weak sentiment score	Numeric
	N_SENTI_SCORE	Overall sentiment score	Numeric
	N_SENTI_CLASS	Unsupervised sentiment class	Nominal (Negative/Neutral/Positive)
	S_SENTI_CLASS	Supervised sentiment class	Nominal (Negative/Neutral/Positive)
	STANFORD_VERYNEG	Strong negative score calculated by Stanford CoreNLP	Numeric
	STANFORD_NEG	Negative score calculated by Stanford CoreNLP	Numeric
	STANFORD_NEUTRAL	Neutral score calculated by Stanford CoreNLP	Numeric
	STANFORD_POS	Positive score calculated by Stanford CoreNLP	Numeric
	STANFORD_VERYPOS	Strong positive score calculated by Stanford CoreNLP	Numeric
Reviewer	REVIEWER_LEVEL	Reviewer level	Nominal (NA/NewReviewer/Reviewer/Senior Reviewer/Contributor/Senior Contributor/TopContributor)
	REVIEWER_SEX	Reviewer gender	Nominal (NA/Male/Female)
	REVIEWER_AGE	Reviewer age	Nominal (NA/13–17 years old/18–24 years old/25–34 years old/35–49 years old/50–64 years old/above 65 years old)
	JOIN_MONTHS	Number of months the reviewer has joined TripAdvisor.com	Numeric
	NUM_PAST_REVIEW	Total number of past reviews	Numeric
	NUM_PAST_HOTEL	Total number of hotels the reviewer has reviewed	Numeric
	NUM_PAST_VOTE	Total number of past votes	Numeric
	RECENCY	Recency score	Numeric
	FREQUENCY	Frequency score	Numeric
	MONETARY	Monetary score	Numeric

in every review and the IVs related to review sentiment, as follows:

$$\text{Strongsubjpostivesentimentscore}(\text{STR}_{\text{POS}}) = \frac{\text{str}_{\text{pos}_i}}{\text{senti}_{\text{tot}_i}} \tag{3}$$

$$\text{Strongsubjnegativesentimentscore}(\text{STR}_{\text{NEG}}) = \frac{\text{str}_{\text{neg}_i}}{\text{senti}_{\text{tot}_i}} \tag{4}$$

$$\text{Weaksubjpostivesentimentscore}(\text{WEAK}_{\text{POS}}) = \frac{\text{weak}_{\text{pos}_i}}{\text{senti}_{\text{tot}_i}} \tag{5}$$

$$\text{Weaksubjnegativesentimentscore}(\text{WEAK}_{\text{NEG}}) = \frac{\text{weak}_{\text{neg}_i}}{\text{senti}_{\text{tot}\,i}} \tag{6}$$

$$\text{Strongsubjsentimentscore}(\text{STR}_{\text{SENTI}}) = \frac{(\text{str}_{\text{pos}_i} + \text{str}_{\text{neg}_i})}{\text{senti}_{\text{tot}\,i}} \tag{7}$$

$$\text{Weaksubjsentimentscore}(\text{WEAK}_{\text{SENTI}}) = \frac{(\text{weak}_{\text{pos}_i} + \text{weak}_{\text{neg}_i})}{\text{senti}_{\text{tot}\,i}} \tag{8}$$

$$\text{Overallsentimentscore}(N_{\text{SENTI}}\text{SCORE}) = (\text{str}_{\text{pos}_i} * 2 + \text{weak}_{\text{pos}_i}) - (\text{str}_{\text{neg}_i} * 2 + \text{weak}_{\text{neg}_i}) \tag{9}$$

where $\text{str}_{\text{pos}_i}$ denotes the number of times strong positive words appeared in review i, $\text{str}_{\text{neg}_i}$ denotes the number of times strong negative words appeared, $weak_{pos_i}$ denotes the number of times weak positive words appeared, $weak_{neg_i}$ denotes the number of times weak negative words appeared, and $n_{senti}score_{tot\,i}$ denotes the summation of str_pos$_i$, str_neg$_i$, weak_pos$_i$, and weak_neg$_i$.

This study also employed Stanford CoreNLP for supervised learning-based sentiment analysis. Previous studies have demonstrated that Stanford CoreNLP has high sentiment classification accuracy. Unlike the traditional bag-of-words model, this algorithm first identifies the relationship between words through syntax analysis automatically and then obtains the results of the syntax tree. We used the Stanford Sentiment Treebank corpus, which contains 11,855 sentences and 215,154 short sentences.

The supervised sentiment variable of review *i*, denoted as S_SENTI_CLASS$_i$, was determined through the following process: the algorithm generated the five sentiment categories through the aforementioned method and then calculated individual sentiment score according to the ratio of these five categories of review *i*. The review was classified as positive if the result was greater than 0 (STANFORD_VERYPOS and STANFORD_POS), neutral if the result equaled 0 (STANFORD_NEUTRAL), and negative if the result was less than 0 (STANFORD_VERYNEG and STANFORD_NEG).

3.2.3. Reviewer characteristics

This study considered the following reviewer features to evaluate review helpfulness, in addition to using text characteristics: the contribution of a reviewer to TripAdvisor.com (REVIEWER_LEVEL), the gender of the reviewer (REVIEWER_SEX), the age of the reviewer (REVIEWER_AGE), and the number of months between the date a review was posted by the reviewer and the date his or her account was registered (JOIN_MONTHS). Reviews that the reviewer had posted in the past had a higher probability of obtaining more comments over time; furthermore, the reviewer who joined the social group earlier also had a higher potential of posting more reviews and receive more attention. Therefore, we traced reviewers' characteristics at the moment when they posted the review according to reviews' historical records. An example of a reviewer's historical records is presented in Fig. 2. We collected reviewers' previous review information, including the total number of past reviews (NUM_PAST_REVIEW), the total number of hotels the reviewer has reviewed (NUM_PAST_HOTEL), and the total number of past votes (NUM_PAST_VOTE).

Ngo-Ye and Sinha [32] applied RFM theory to online review research and described the overall contribution of a reviewer to the entire online social network. For the target reviewer, the following

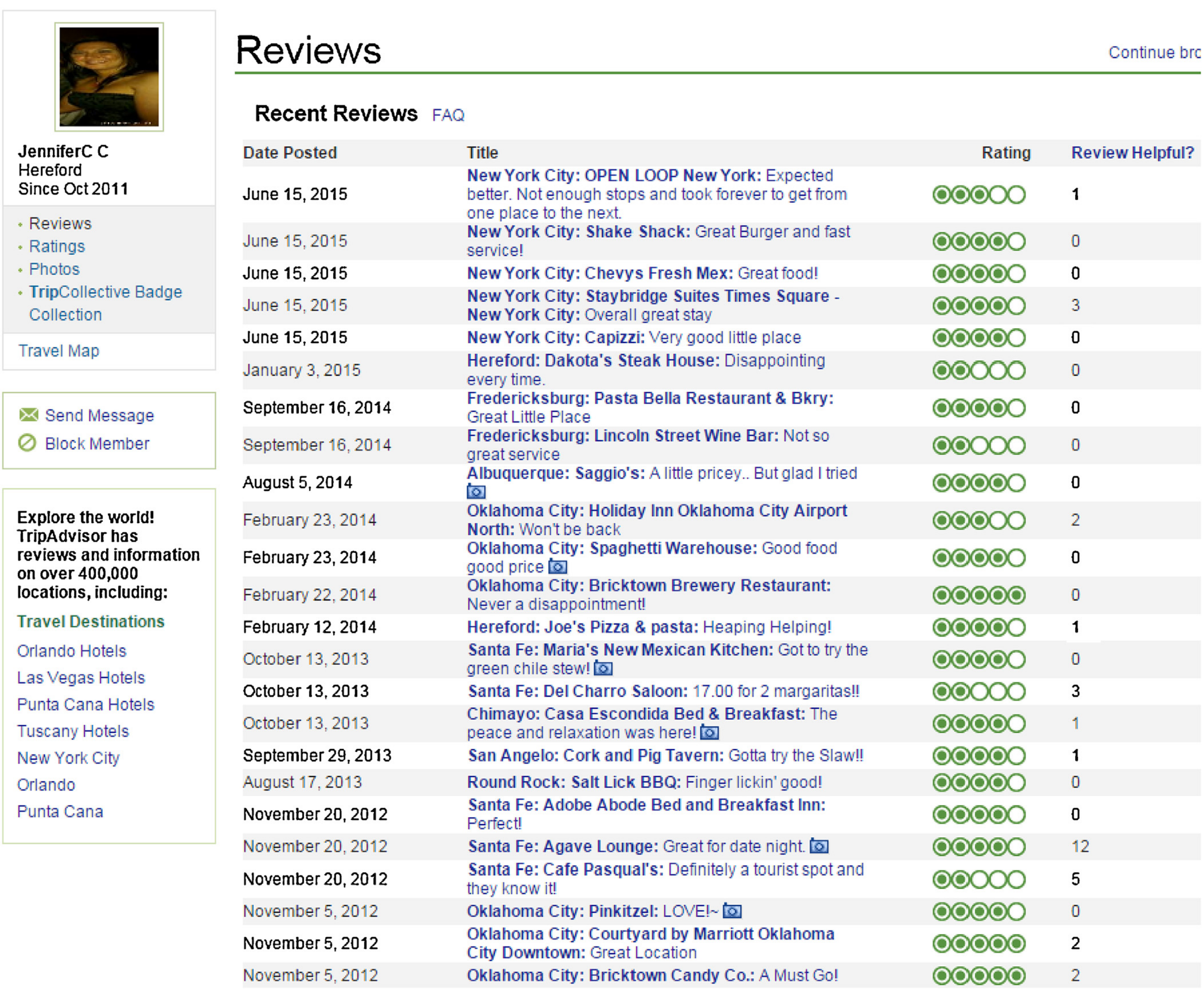

Fig. 2. An example of a reviewer's historical record.

Please cite this article in press as: Y.-H. Hu, et al., The effect of user-controllable filters on the prediction of online hotel reviews, Inf. Manage. (2017), http://dx.doi.org/10.1016/j.im.2016.12.009

three RFM-related variables were considered: the date difference between the latest review and the previous review (RECENCY), the total number of reviews before the current review (FREQUENCY), and the total number of votes the reviewer had received in the past (MONETARY).

3.3. Investigated prediction techniques

We used Weka 3.6.14 data-mining software to construct the models [73]. Three prediction techniques were applied because of the characteristics of analysis—LR, REPTree, and RF. The LR is a widely used statistical method for modeling a dependent variable

Table 4
Descriptive statistics for the reviews of the three regions.

Variable category	Variable Name	New York City/Chicago (NC) (n = 307,367) mean (std. dev.)	Miami/Orlando (MO) (n = 175,376) mean (std. dev.)	Las Vegas (LV) (n = 234,259) mean (std. dev.)
Review Quality	RATING	4.181 (1.012)	4.067 (1.113)	4.097 (1.078)
	LENGTH_CHAR	589.813 (494.822)	686.307 (659.918)	611.765 (599.693)
	LENGTH_SYLL	199.023 (165.939)	231.055 (220.144)	206.696 (200.744)
	LENGTH_WORD	137.4 (117.86)	161.441 (157.642)	145.936 (144.516)
	LENGTH_SENT	10.115 (8.463)	11.8 (11.067)	11.193 (11.256)
	SYLL_PER_WORD	1.419 (0.111)	1.404 (0.111)	1.391 (0.114)
	WORD_PER_SENT	15.255 (7.895)	15.704 (10.148)	15.254 (9.025)
	ARI	6.316 (2.519)	6.191 (2.6)	5.734 (2.657)
	CLI	9.499 (1.561)	9.266 (1.584)	8.929 (1.627)
	FRES	71.322 (10.797)	72.207 (11.53)	73.737 (11.484)
	FGL	6.932 (2.121)	6.824 (2.206)	6.531 (2.239)
	FOG	9.555 (2.528)	9.466 (2.676)	9.294 (2.706)
	SMOG	6.943 (1.616)	6.863 (1.684)	6.669 (1.682)
Review Polarity	SUBJECTIVITY	0.368 (0.254)	0.346 (0.252)	0.334 (0.256)
	STR_POS	0.47 (0.187)	0.454 (0.192)	0.452 (0.197)
	STR_NEG	0.109 (0.132)	0.13 (0.148)	0.13 (0.146)
	WEAK_POS	0.334 (0.169)	0.335 (0.173)	0.335 (0.181)
	WEAK_NEG	0.087 (0.098)	0.081 (0.097)	0.082 (0.1)
	STR_SENTI	0.579 (0.172)	0.583 (0.175)	0.582 (0.183)
	WEAK_SENTI	0.421 (0.172)	0.416 (0.175)	0.417 (0.183)
	N_SENTI_SCORE	11.496 (9.561)	11.363 (10.758)	10.245 (9.643)
	N_SENTI_CLASS	Negative: 16,125 Neutral: 5084 Positive: 286,158	Negative: 12,420 Neutral: 3514 Positive: 159,442	Negative: 16,412 Neutral: 5130 Positive: 212,717
	S_SENTI_CLASS	Negative: 147,429 Neutral: 33,122 Positive: 126,816	Negative: 97,725 Neutral: 17,227 Positive: 60,424	Negative: 123,927 Neutral: 24,015 Positive: 86,317
	STANFORD_VERYNEG	0.016 (0.064)	0.018 (0.073)	0.018 (0.072)
	STANFORD_NEG	0.439 (0.238)	0.473 (0.24)	0.455 (0.247)
	STANFORD_NEUTRAL	0.139 (0.145)	0.143 (0.145)	0.157 (0.153)
	STANFORD_POS	0.361 (0.231)	0.329 (0.23)	0.331 (0.232)
	STANFORD_VERYPOS	0.045 (0.1)	0.037 (0.091)	0.04 (0.097)
Reviewer	REVIEWER_LEVEL	N/A: 67,793 New Reviewer: 22,715 Reviewer: 34,046 Senior Reviewer: 58,648 Contributor: 48,789 Senior Contributor: 40,290 Top Contributor: 35,086	N/A: 40,746 New Reviewer: 12,768 Reviewer: 18,804 Senior Reviewer: 32,352 Contributor: 27,398 Senior Contributor: 22,671 Top Contributor: 20,637	N/A: 60,332 New Reviewer: 16,293 Reviewer: 23,898 Senior Reviewer: 41,952 Contributor: 35,793 Senior Contributor: 29,741 Top Contributor: 26,250
	REVIEWER_SEX	N/A: 192,903 Male: 62,942 Female: 51,522	N/A: 110,780 Male: 33,719 Female: 30,877	N/A: 149,618 Male: 46,245 Female: 38,396
	REVIEWER_AGE	N/A: 202,448 13–17 yrs: 70 18–24 yrs: 2878 25–34 yrs: 21,877 35–49 yrs: 42,260 50–64 yrs: 32,059 >65 yrs: 5775	N/A: 115,777 13–17 yrs: 79 18–24 yrs: 1,689 25–34 yrs: 11,257 35–49 yrs: 27,036 50–64 yrs: 16,346 >65 yrs: 3192	N/A: 155,724 13–17 yrs: 55 18–24 yrs: 2007 25–34 yrs: 17,418 35–49 yrs: 30,516 50–64 yrs: 24,132 >65 yrs: 4407
	JOIN_MONTHS	48.448 (33.635)	47.015 (33.819)	45.032 (33.859)
	NUM_PAST_REVIEW	81.566 (433.494)	83.925 (387.312)	88.149 (453.818)
	NUM_PAST_HOTEL	2.883 (2.305)	2.843 (2.304)	2.763 (2.31)
	NUM_PAST_VOTE	22.958 (65.185)	23.988 (56.34)	22.341 (53.549)
	RECENCY	487.615 (312.174)	481.384 (312.777)	476.956 (312.974)
	FREQUENCY	0.339 (0.990)	0.394 (1.086)	0.367 (0.985)
	MONETARY	0.313 (1.318)	0.525 (2.062)	0.454 (1.853)
DV	HELPFULNESS	0.147 (0.596)	0.258 (0.813)	0.213 (0.904)

according to a linear combination of one or more IVs. Considering the experimental results, researchers are recommended to use statistical modeling such as LR to strengthen the predictive capabilities of the model. REPTree is an extension of regression tree (RT) techniques with a hierarchical structure comprising branches and nodes. The internal node represents one of the selected IVs and its branches represent a subset of predictor values. The leaf node represents a set of instances satisfying a specific set of decision rules in the tree. REPTree adopts the RT tree logic that creates multiple trees at each iteration and then selects the most favorable iteration as the representative. Therefore, REPTree is a fast decision tree (DT) learner, which constructs an RT by using information gain as the splitting criterion and shapes the tree using REP. Finally, the RF is an ensemble learning method developed by constructing multiple DTs [74]. In the training process, RF applies the bagging technique to bootstrap instances; a set of DTs is then constructed on the basis of each set of bootstrap instances with a subset of features. Thereafter, a random subset of features is selected. After the set of trees is constructed, a prediction regarding unseen samples can be generated by selecting the majority class of individual trees.

4. Analysis and results

4.1. Descriptive statistics of the collected reviews

The descriptive statistics of the hotel reviews of the three regions are displayed in Table 4, revealing that the average number of reviews per hotel was significantly different. Among the three travel regions, the south region received the most helpfulness votes per a review on average (0.258), followed by the west (0.213) and north (0.147) regions. The average review ratings in all three regions were relatively similar. With regard to the average review length, the south region ranked first (11.8 sentences per review), followed by the west (11.193) and north (10.115) regions.

4.2. Baseline replication

The purpose of a baseline model is to replicate existing research for the possible refinement of existing theories. In Mudambi and Schuff [12], product reviews were analyzed for two categories of products—search goods and experience goods. Because hotels are experience goods, their results for experience goods are particularly relevant to the present study. Table 5 confirms that our results from a 10-fold cross-validation are relatively similar to those of Mudambi and Schuff, with rating being the strongest predictor of review helpfulness, followed by rating squared and then by word count.

The major difference is that the model fit degrades significantly in our model ($R^2 = 0.015$) versus that in Mudambi and Schuff [12] ($R^2 = 0.361$), indicating that the existing list of predictors from Mudambi and Schuff [12] may not be sufficient to explain the amount of variance for review helpfulness in the context of hotel reviews. This is consistent with the conclusion of other studies (e.g., [11], in which readers of hotel reviews tend to rely on

Table 5
Baseline replication.

	Coefficient	Std. Error	Standardized Coefficient	*t*-value	Sig.
(Constant)	0.671	0.008		84.766	0.000
Rating	−0.292	0.005	−0.406	−62.770	0.000
Rating^2	0.037	0.001	0.359	55.480	0.000
Word count	0.000	0.000	0.079	66.489	0.000

$R^2 = 0.015$, correlation coefficient (CC) = 0.1216, mean absolute error (MAE) = 0.2574, root- mean-squared error (RMSE) = 0.7578.

additional cues to assess the quality or helpfulness of a hotel review. Therefore, the following sections are designed to reveal insights regarding predictors supported in the relevant literature as well as interactions among the major predictors.

4.3. Extensions to the existing model

4.3.1. Interaction effect

As travel websites offer options to filter reviews according to certain attributes, the probability of a review being available on the top (termed review visibility by Hu and Chen [11] in the search result is partially controlled by the filter the reader has chosen. Eventually, the reviews on the top have a greater chance of being voted as helpful. For TripAdvisor.com, filters are available for city, traveler type, language, and time of year. Because we focused on reviews written only in English, the remaining three filters were relevant to the present study.

In this section of the study, we examine the interactions among the aforementioned filters for their effects on review helpfulness. This area has not been explored in previous product review studies. The majority of related studies have focused on identifying the most favorable combination of predictors, thereby neglecting that predictors also interact with each other. Such interactions would affect the linear relationship between the predictors and the DV. Table 6 presents the interaction effects among travel region (TravelRegion), travel type (TravelType), and time of year (TimeOfYear) for travel. In the present study, we sampled the five top cities into three geographical regions, namely north region (New York City/Chicago, NC), south region (Miami/Orlando, MO), and west region (Las Vegas, LV). To simplify the analysis, we categorized the reviews by two traveler types (i.e., TravelType = Business/Non-business, B/NB) and two periods (TimeOfYear = Summer–Fall/Winter–Spring, SF/WS). Consequently, we obtained a $3 \times 2 \times 2$ design.

Table 6 shows that all the three main effects were significant. The four interaction effects are also statistically significant, meaning that helpfulness varies among travel destination, types of travel, and the time of the year. As Fig. 3 shows, helpfulness of business travels peaked when the destination is in the southern geographic region. The slope of the two lines in Fig. 4 varies more between the northern and southern regions than between the southern and western regions. Helpfulness shown in Fig. 5 is more similar for Summer–Fall travels between business and non-business travels than that for the Winter–Spring time.

4.3.2. Model building

A 10-fold cross-validation was applied to all of the experimental evaluations. To evaluate the model performance, the metrics of correlation coefficient (CC), mean absolute error (MAE), and root-mean-squared error (RMSE) were considered. CC describes the degree of the linear relationship between observed and simulated data. The CC ranged from −1 to 1 for a perfectly negative or positive interrelationship. The MAE measures the average value of the sample of the observed and simulated data; an MAE of zero suggests a perfect fit or high accuracy. The RMSE measures the squared and averaged value of the sample of the observed and simulated data. Similar to the MAE, an RMSE of zero indicates a perfect fit for the data; however, the RMSE value will be higher for greater sampling error because the values are squared. Both error measurements describe the variation in the observed and simulated data.

On the basis of the results in Section 4.3.1, the complete set of hotel reviews in each travel region was further divided into four subsets: business trip during Summer–Fall (SF-B), non-business trip during Summer–Fall (SF-NB), business trip during Winter–

Table 6
Interaction effects among city, traveler type, and time of year for travel Dependent variable: Review Helpfulness.

Source	Type III Sum of Squares	df	Mean Square	*F*	Sig.
Corrected Model	9436.162	11	857.833	1505.870	0.000
Intercept	12354.334	1	12354.334	21687.224	0.000
TravelRegion	421.718	2	210.859	370.149	0.000
TravelType	379.132	1	379.132	665.542	0.000
TimeOfYear	3293.759	1	3293.759	5781.979	0.000
TravelRegion * TravelType	171.979	2	85.989	150.949	0.000
TravelRegion * TimeOfYear	83.178	2	41.589	73.007	0.000
TravelType * TimeOfYear	86.861	1	86.861	152.478	0.000
TravelRegion * TravelType * TimeOfYear	32.892	2	16.446	28.870	0.000
Error	408440.195	716990	0.570		
Total	445300.151	717002			
Corrected Total	417876.357	717001			

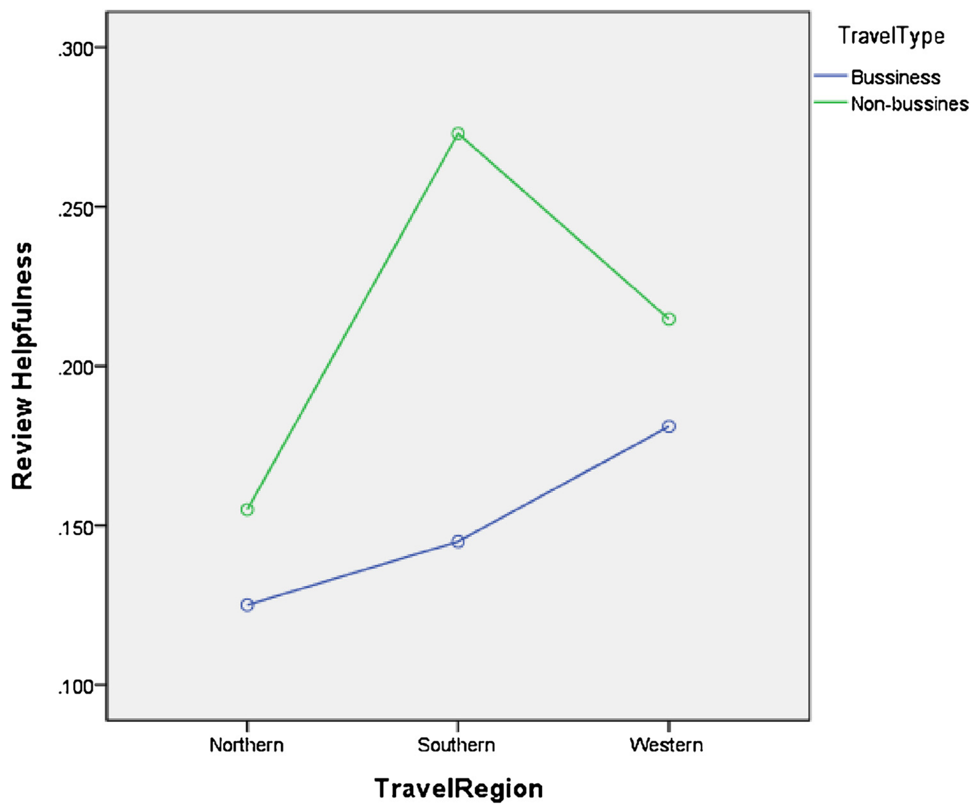

Fig. 3. Interaction plot (TravelType by TravelRegion).

Spring (WS-B), and non-business trip during Winter–Spring (WS-NB).

The CC, MAE, and RMSE values for the 12 data sets for the classification of travel types were compared and are presented in Table 7. The CC values revealed that all three prediction techniques had weak-to-moderate linear relationships between observed and simulated data, and the samples were directly related. On average, the CC and RMSE performance indicators show that RF was the best model to predict helpfulness, followed by REPTree and LR. REPtree had the lowest MAE, followed by RF and LR, but there is no significant statistical difference between REPTree and RF on MAE. Therefore, judging by all three performance indicators, RF should be recommended as the most effective model among the three. It is worth noting that all the developed models in this study significantly outperform the baseline model (Table 5). This provides empirical evidence that additional variables extracted from review diagnosticity, reviewer credibility, and other aspects help to improve the predictive power of all three models.

Among the different regions, NC and LV tend to have a lower CC; however, NC-WS-B and LV-WS-B had a higher CC. A higher CC indicates that the predictors were adequately related to the outcome variables. The MAE and RMSE were extremely similar for all data sets. The MAE for B dominated the lower 50%; the MAEs for NC were relatively lower than those of MO and LV across all data sets. These results indicate adequate classification and low prediction errors for all data sets. In addition, the selected features can be used to increase the performance of review helpfulness prediction. The RMSE for B was also in the lower 50% range. In general, NC had lower RMSE of the prediction techniques. In addition to the NC being the region with a lower RMSE, NC-SF had the lowest MAE and RMSE. In addition to the lower end, LV had the highest MAE and RMSE. As evident from this evaluation, LV-WS-NB had a significantly higher RMSE among all data sets. Because the

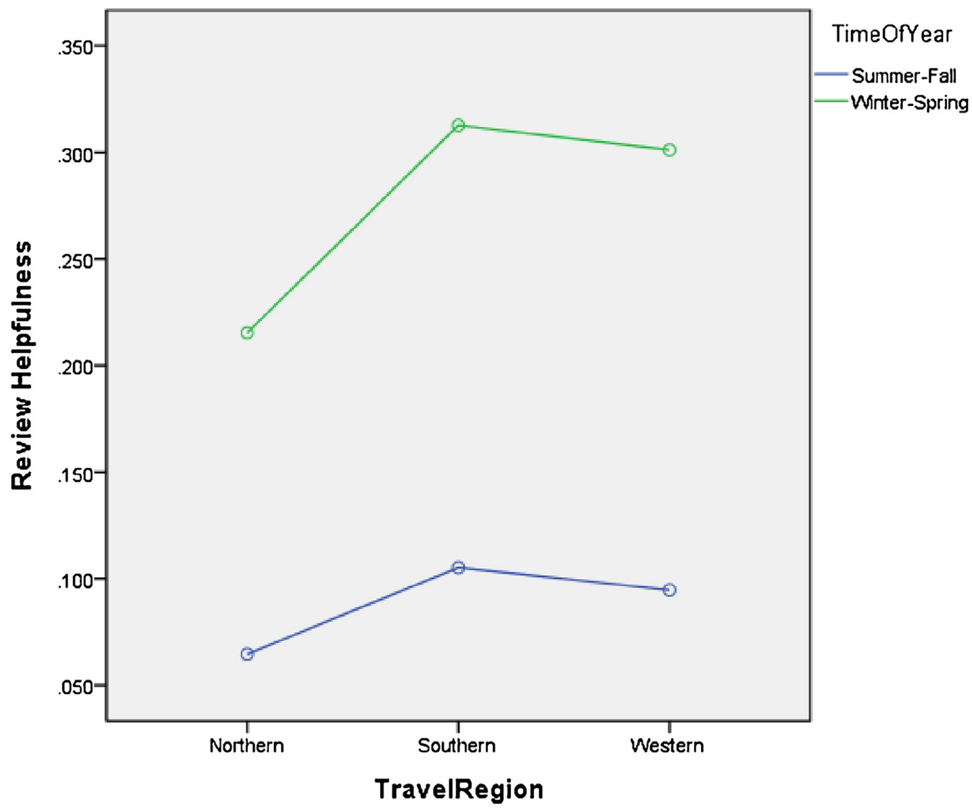

Fig. 4. Interaction plot (TimeOfYear by TravelRegion).

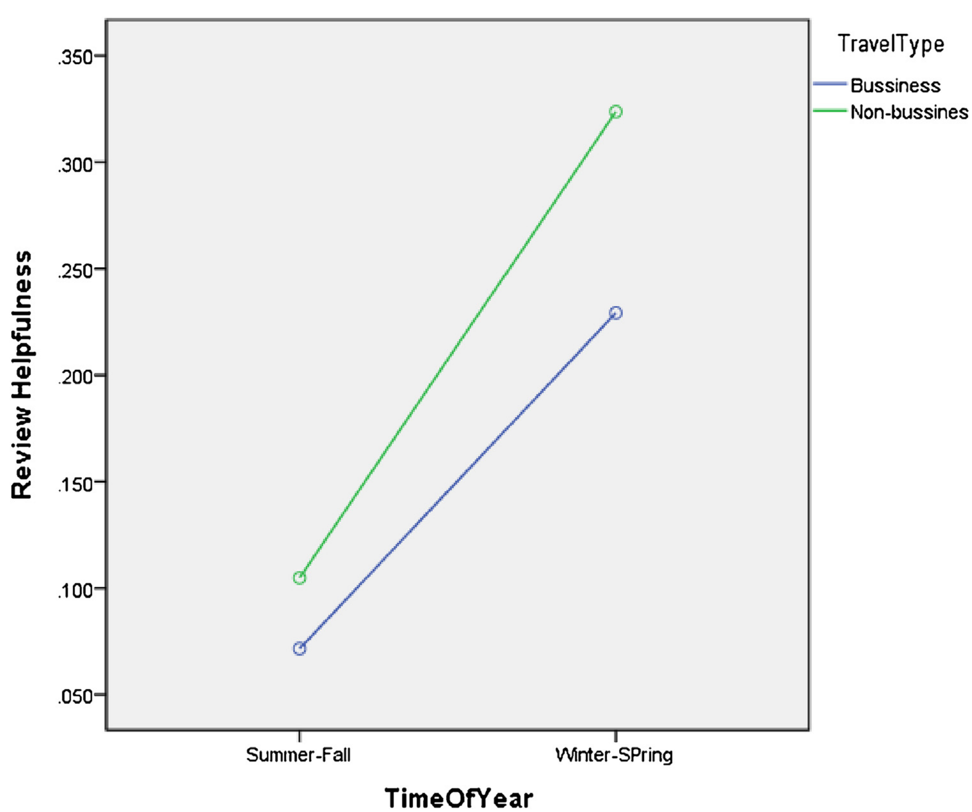

Fig. 5. Interaction plot (TravelType by TimeOfYear).

Table 7
Experimental results of different prediction models using all features[a].

Data set	TravelRegion	TimeOfYear	TravelType	Classifiers	CC	MAE	RMSE
NC-SF-B	NC	SF	B	LR	0.248	0.069	0.166
				REPtree	0.369	0.059	0.160
				RF	0.386	0.062	0.158
NC-SF-NB	NC	SF	NB	LR	0.240	0.085	0.237
				REPtree	0.452	0.073	0.218
				RF	0.482	0.074	0.214
NC-WS-B	NC	WS	B	LR	0.298	0.284	0.670
				REPtree	0.405	0.226	0.646
				RF	0.503	0.214	0.607
NC-WS-NB	NC	WS	NB	LR	0.297	0.329	0.799
				REPtree	0.537	0.231	0.707
				RF	0.550	0.238	0.700
MO-SF-B	MO	SF	B	LR	0.280	0.078	0.176
				REPtree	0.356	0.070	0.172
				RF	0.458	0.068	0.163
MO-SF-NB	MO	SF	NB	LR	0.297	0.143	0.386
				REPtree	0.573	0.107	0.332
				RF	0.629	0.108	0.316
MO-WS-B	MO	WS	B	LR	0.334	0.286	0.607
				REPtree	0.520	0.203	0.553
				RF	0.576	0.213	0.529
MO-WS-NB	MO	WS	NB	LR	0.391	0.493	1.024
				REPtree	0.662	0.307	0.836
				RF	0.683	0.318	0.816
LV-SF-B	LV	SF	B	LR	0.270	0.098	0.282
				REPtree	0.480	0.084	0.262
				RF	0.565	0.081	0.244
LV-SF-NB	LV	SF	NB	LR	0.201	0.106	0.399
				REPtree	0.239	0.093	0.412
				RF	0.527	0.084	0.348
LV-WS-B	LV	WS	B	LR	0.285	0.369	0.961
				REPtree	0.438	0.255	0.928
				RF	0.593	0.263	0.813
LV-WS-NB	LV	WS	NB	LR	0.301	0.456	1.166
				REPtree	0.612	0.271	0.971
				RF	0.627	0.289	0.956
Average				LR	0.287	0.233	0.573
				REPtree	0.470	**0.165**	0.516
				RF	**0.548**	0.167	**0.489**

[a] NC: north region (New York and Chicago), LV: west region (Las Vegas), MO: south region (Miami and Orland), SF: Summer–Fall, WS: Winter–Spring, B: business travel, NB: non-business travel.

RMSE had unequal weight toward errors, greater errors gained higher RMSEs, and the LV had greater variations across the samples.

4.3.3. Reduction of independent variables (IVs)

To further simplify the models, we conducted the correlation-based feature subset selection (CFS) technique to reduce the dimensionality of the data sets [75]. Specifically, the CfsSubsetEval module with BestFirst search method in WEKA was used in our study. This method chooses a set of IVs that maximize their correlations with the DV, but minimize the correlation among them.

After applying the CFS method, the average number of IVs was reduced to 15.42, which is approximately two-fifths of the original IVs. As shown in Table 8, after feature selection, RF has the highest average CC, MAE, and RMSE, followed by REPTree and LR. Compared with the results in Table 7, the total number of IVs is greatly reduced with the CFS method at the expense of a small reduction of performance measured as the average CC, MAE, and RMSE.

5. Discussion

The present study extends Mudambi and Schuff's diagnosticity-based theory to include predictors of product reviews from diagnosticity, eWOM, and UGC literatures. Mudambi and Schuff's model was first checked for its validity in the context of hotel reviews—a form of experience goods that is distinctively different from search goods. The predictors were also checked together with the interactions among user-controllable filters for their collective

Table 8
Experimental results of different prediction models using feature selection techniques[a].

Data set	Travel Region	Time Of Year	Travel Type	Classifiers	CC	MAE	RMSE
NC-SF-B	NC	SF	B	LR	0.245	0.069	0.166
				REPtree	0.334	0.066	0.162
				RF	0.349	0.067	0.161
NC-SF-NB	NC	SF	NB	LR	0.238	0.085	0.238
				REPtree	0.455	0.079	0.218
				RF	0.447	0.081	0.220
NC-WS-B	NC	WS	B	LR	0.296	0.283	0.670
				REPtree	0.390	0.228	0.649
				RF	0.461	0.214	0.625
NC-WS-NB	NC	WS	NB	LR	0.294	0.328	0.800
				REPtree	0.459	0.265	0.745
				RF	0.473	0.268	0.739
MO-SF-B	MO	SF	B	LR	0.276	0.078	0.176
				REPtree	0.287	0.071	0.184
				RF	0.461	0.067	0.163
MO-SF-NB	MO	SF	NB	LR	0.290	0.143	0.387
				REPtree	0.542	0.119	0.340
				RF	0.581	0.121	0.330
MO-WS-B	MO	WS	B	LR	0.334	0.286	0.608
				REPtree	0.439	0.239	0.582
				RF	0.466	0.240	0.573
MO-WS-NB	MO	WS	NB	LR	0.388	0.493	1.025
				REPtree	0.587	0.353	0.903
				RF	0.597	0.357	0.898
LV-SF-B	LV	SF	B	LR	0.272	0.097	0.282
				REPtree	0.451	0.089	0.266
				RF	0.534	0.087	0.248
LV-SF-NB	LV	SF	NB	LR	0.200	0.106	0.399
				REPtree	0.246	0.100	0.412
				RF	0.562	0.092	0.338
LV-WS-B	LV	WS	B	LR	0.298	0.368	0.957
				REPtree	0.433	0.284	0.920
				RF	0.555	0.276	0.834
LV-WS-NB	LV	WS	NB	LR	0.300	0.455	1.167
				REPtree	0.551	0.322	1.024
				RF	0.564	0.325	1.012
Average				LR	0.286	0.233	0.573
				REPtree	0.431	0.185	0.534
				RF	**0.504**	**0.183**	**0.512**

[a] NC: north region (New York and Chicago), LV: west region (Las Vegas), MO: south region (Miami and Orland), SF: Summer–Fall, WS: Winter–Spring, B: business travel, NB: non-business travel.

relationship with hotel review helpfulness. The extended formulation of predictors was tested in three models, namely LR, RF, and REPTree, to examine their predictability.

Although the resulting performance metrics (CC, MAE, and RMSE) indicate that all three models outperform the baseline model of Mudambi and Schuff's, we went a step further to also report the extent to which the set of predictors are applicable when user-controllable filters are in action. TripAdvisor.com provides four user-controllable filters, namely geographic location, type of travel, season of travel (i.e., TimeOfYear), and language. Since we are concerned with the English hotel reviews, our work focuses on only the first three filters. As a result, we had three geographic regions, two travel types (B versus NB), and two travel seasons (SF versus WS). The combination divides the sample into 12 subsamples ($3 \times 2 \times 2$).

Table 9 summarizes critical IVs identified through the CFS method for the 12 subsamples individually. For review content features, our findings show that the review rating (RATING) and the number of words in a review (LENGTH_WORD) have strong effects on review helpfulness, which is consistent with existing studies [12,47]. In addition, although review readability has been exploited in past studies relating to review helpfulness [24,29,30,57], our work shows that review readability is not always valued by all travelers.

For review polarity features, both positive and negative sentiment score features displayed strong effects on review helpfulness in all subsamples. Research [29,76] has identified that sentiment adjectives can be widely identified in sentences expressing opinions. Positive opinions can provide consumers with favorable factors that they want to learn about regarding

Table 9
Crucial variables for each data set[a].

NC-SF-B	NC-SF-NB	NC-WS-B	NC-WS-NB
RATING	RATING	RATING	RATING
LENGTH_WORD	LENGTH_WORD	LENGTH_WORD	LENGTH_SENT
SYLL_PER_WORD	WORD_PER_SENT	STR_POS	WORD_PER_SENT
CLI	STR_POS	STR_NEG	ARI
FRES	STR_NEG	WEAK_POS	CLI
STR_POS	WEAK_POS	WEAK_NEG	FOG
STR_NEG	STR_SENTI	STR_SENTI	STR_POS
WEAK_POS	N_SENTI_CLASS	WEAK_SENTI	STR_NEG
STR_SENTI	S_SENTI_CLASS	N_SENTI_CLASS	WEAK_POS
WEAK_SENTI	STANFORD_NEG	S_SENTI_CLASS	WEAK_NEG
N_SENTI_CLASS	STANFORD_POS	STANFORD_NEG	STR_SENTI
S_SENTI_CLASS	RECENTCY	STANFORD_POS	WEAK_SENTI
STANFORD_VERYNEG	MONETARY	STANFORD_VERYPOS	N_SENTI_CLASS
STANFORD_NEG		NUM_PAST_VOTE	S_SENTI_CLASS
STANFORD_POS		RECENTCY	STANFORD_VERYNEG
STANFORD_VERYPOS		MONETARY	STANFORD_NEG
RECENTCY			STANFORD_POS
FREQUENCY			STANFORD_VERYPOS
MONETARY			RECENTCY
			MONETARY

MO-SF-B	MO-SF-NB	MO-WS-B	MO-WS-NB
RATING	RATING	RATING	RATING
LENGTH_WORD	LENGTH_WORD	LENGTH_WORD	LENGTH_WORD
STR_POS	ARI	STR_POS	STR_NEG
STR_NEG	FOG	STR_NEG	WEAK_POS
WEAK_POS	SMOG	WEAK_POS	STR_SENTI
STR_SENTI	WEAK_POS	WEAK_NEG	WEAK_SENTI
WEAK_SENTI	STR_SENTI	N_SENTI_CLASS	STANFORD_VERYNEG
N_SENTI_CLASS	N_SENTI_CLASS	STANFORD_NEUTRAL	RECENTCY
S_SENTI_CLASS	NUM_PAST_HOTEL	STANFORD_POS	MONETARY
STANFORD_VERYNEG	RECENTCY	RECENTCY	
STANFORD_NEG	MONETARY	MONETARY	
STANFORD_NEUTRAL			
STANFORD_POS			
NUM_PAST_HOTEL			
NUM_PAST_VOTE			
RECENTCY			
MONETARY			

LV-SF-B	LV-SF-NB	LV-WS-B	LV-WS-NB
RATING	RATING	RATING	RATING
LENGTH_WORD	LENGTH_WORD	LENGTH_WORD	LENGTH_WORD
LENGTH_SENT	WORD_PER_SENT	WORD_PER_SENT	STR_POS
SYLL_PER_WORD	FOG	CLI	STR_NEG
FRES	STR_POS	FRES	WEAK_POS
SMOG	STR_NEG	STR_POS	WEAK_NEG
STR_POS	WEAK_POS	STR_NEG	STR_SENTI
STR_NEG	WEAK_NEG	WEAK_POS	WEAK_SENTI
WEAK_POS	N_SENTI_CLASS	WEAK_NEG	N_SENTI_CLASS
WEAK_NEG	S_SENTI_CLASS	N_SENTI_CLASS	S_SENTI_CLASS
WEAK_SENTI	STANFORD_NEG	S_SENTI_CLASS	STANFORD_NEG
N_SENTI_CLASS	STANFORD_POS	STANFORD_NEG	STANFORD_POS
S_SENTI_CLASS	RECENTCY	STANFORD_POS	STANFORD_VERYPOS
STANFORD_VERYNEG	FREQUENCY	STANFORD_VERYPOS	RECENTCY
STANFORD_NEG	MONETARY	RECENTCY	FREQUENCY
STANFORD_POS		FREQUENCY	MONETARY
STANFORD_VERYPOS		MONETARY	
NUM_PAST_HOTEL			
RECENTCY			
FREQUENCY			
MONETARY			

[a] NC: north region (New York and Chicago), LV: west region (Las Vegas), MO: south region (Miami and Orland), SF: Summer–Fall, WS: Winter–Spring, B: business travel, NB: non-business travel.

hotels, whereas negative opinions provide them with information concerning disadvantages and defects. Our findings in this regard are fairly consistent with existing literature. The RFM model [32] also received a consistent support across all our data subsets with recency and monetary appearing in every one of the data subsets.

Prior studies indicate that reviewer characteristics concerning their past records and activity rates on travel websites were also related to review helpfulness [27,32,56]. In our study, the date difference between the latest review and the previous review (RECENCY) and the total number of votes the reviewer has received (MONETARY) are consistently the key aspects for all types of travel. Reviewer characteristics are gaining considerable attention for review helpfulness prediction. Our findings help shed some light in our understanding of how these features are related to hotel review helpfulness.

6. Conclusions

Travel websites have become an important source for travelers to plan their trips and share their own experiences. The collective review messages made available through a travel website are like a large database that makes the knowledge accessible to the general public. However, the exponential growth of the available review messages makes it difficult for a reader to distill the information for the trip that he or she has in mind. Therefore, identifying helpful reviews accurately becomes an important issue. Added to the complexity is that most travel websites allow readers to filter review messages. In our case, TripAdvisor.com provides four filters, including geographic location, season of travel, type of travel, and language. These filters impose an artificial selection of what reviews to show for a given search. As a result, not all reviews receive an equal opportunity to be visible and be voted on for helpfulness. This study collected hotel reviews from TripAdvisor.com to study review helpfulness and its predictors in three groups (namely review quality, review sentiment, and reviewer characteristics) in conjunction with the three filters mentioned above.

Our work contributes to the literature in three ways. First, this study is one of the first to examine the interaction effects among geographic location, season of travel, and travel type. The majority of product review studies overly emphasize on optimizing the best combination of independent variables (IVs), while neglecting the fact that the IVs also interact with each other. If interactions are not taken into account, the true relationship between the selected IVs on review helpfulness may be biased, inaccurate, or even distorted. In the present study, the traditional optimizing approach is still followed, but variables were studied together with interaction effects. The results uncover additional insights that were not available before.

Second, we extended the review diagnosticity-based theory from Mudambi and Schuff [12] with the theoretical constructs from the eWOM and other literatures. Three categories of IVs (review polarity, review structure, and reviewer characteristics) adapted from multiple disciplines were tested in three models (LR, RF, and REPTree). All three models tested in the present study had shown large improvements over Mudambi and Schuff's original model. As their model was constructed primarily based on variables representing review structure (e.g., review rating and review length), an extension into other relevant aspects of product review provides a better explanatory power to predict review helpfulness. This is consistent with the expectation that reviews for experience goods (such as hotels) need to provide a multifaceted view to meet the needs of the readers in order to receive helpfulness vote.

Third, by dividing the sample into subsamples based on filters controllable by users and showing strong variations across the subsamples, we were able to demonstrate that the generalizability of the traditional approach, which does not specifically distinguish between possible groupings within the overall sample, may fall short for experience goods. This is an important theoretical implication because treating hotel reviews to be homogeneous across geographic locations, travel types, and travel season will likely obscure the true effect on the dependent variable. As a result, models generated for the overall sample will less likely have good predictive power than the individualized version that we provided in the present study. Additionally, by dividing samples based on user-controllable filters, we were able to shed some light on the boundaries for generalizability based on the individual combinations of user-controllable filters. For example, sentiment-related variables outweigh reviewer characteristics for business travelers when voting on review helpfulness. In other words, the generalizability of reviewer characteristics in the sample of business travelers may be weak. Table 5 summarizes the baseline models based on the parsimonious principle for model construction, which helps future researchers to expand on their theoretical directions.

Several practical implications may be derived from this study. First, we could construct a smart recommendation system that recommends useful reviews customized to travelers on travel websites. Travelers' needs can be identified automatically through the search combination of TravelRegion, TravelType, and TimeOfYear while they browse the travel websites. We predict that this smart recommendation system would provide reviews with higher helpful votes according to review quality, review polarity, and reviewer characteristics as well as the three previously mentioned search combinations. Such a system would conserve time that travelers spend on travel websites searching for helpful reviews for their trips and increase the usability of travel websites if they provide valuable information on the selected hotels. Second, indicators of review helpfulness may also be helpful to hotel managers; they can verify opinions in helpful reviews and subsequently strengthen the advantages and address weaknesses. The crucial advantage of using our proposed algorithm is that hotel managers may not necessarily read every review of their hotels; they may read only the reviews with higher helpful votes for their likely higher impact on sales. Third, we could construct a strategy that provides helpful reviews with higher rankings, thereby providing higher visibility. In general, reviews that have more votes are ranked higher. Such a mechanism disfavors recently published hotel reviews; these reviews will not be placed at the top because of their low rankings and cumulative votes, even if they reflect the latest hotel information. Without any filtering, hotel websites usually list the most recent reviews first regardless of the quality of review; reviewers must invest extra time to filter useful information. Considering the aforementioned challenges, travel websites should not rank online reviews simply by publication date or score; rather, they should filter the most appropriate review of each hotel automatically to reduce the time spent filtering user information, thereby increasing the chances of users using their web services.

This study is not without limitations and there are still areas that future research may help continue to advance our understanding. First, although this study collected the complete set of reviews for five cities from TripAdvisor.com, the reviews from other travel websites such as Hotel.com and Expedia.com were not considered. Since studying the complete set of reviews reduces sampling error compared to the traditional sampling approach, the representation of our sample from TripAdvisor.com may likely be similar to those of other sources of data. Future research may be conducted to confirm the difference. Second, cultural and societal variations are another area not studied in our research. Although there is a possible difference across cultural and societal norms, it is not readily accessible until the travel websites release traces of clues for a researcher to uncover the difference. Third, generally, reviews posted at an earlier time have a greater chance to be seen and voted on. However, more recent reviews are closer to the current reality, thus offering better realistic insights. Future studies may perform longitudinal observations to monitor how review helpfulness changes over time. It is also recommended to consider variables not studied in the present research. One such extension is to study variables concerning the nature of entertainment offerings and surrounding attractions of the city where a hotel resides. This is especially true when experience goods (such as hotels) are the subject of the study. The literature has been concerning variables relating to hotels, reviews, and reviewers, but very little attention has been paid on the effect of factors other than these three categories. Our work casts one of the early calls by showing that the effects of predictor variables vary between the three geographical regions. Further works in this direction will likely help advance our understanding in situational or nontraditional variables.

Acknowledgments

This research was supported in part by the Ministry of Science and Technology of the Republic of China (grant numbers MOST 104-2410-H-194-070-MY3 and MOST104-2410-H-194-109). The authors would like to thank Mr. Kuan-Ting Lu for his support on data collection and data preprocessing.

References

[1] L.J. Harrison-Walker, The measurement of word-of-mouth communication and an investigation of service quality and customer commitment as potential antecedents, J. Serv. Res. 4 (1) (2001) 60–75.
[2] Y. Yoon, M. Uysal, An examination of the effects of motivation and satisfaction on destination loyalty: a structural model, Tour. Manage. 26 (2005) 45–56.
[3] C.M.K. Cheung, M.K.O. Lee, What drives consumers to spread electronic word of mouth in online consumer-opinion platforms, Decis. Support Syst. 53 (1) (2012) 218–225.
[4] A.S. Cantallops, F. Salvi, New consumer behavior: a review of research on eWOM and hotels, Int. J. Hosp. Manage. 36 (2014) 41–51.
[5] U. Gretzel, M. Sigala, Z. Xiang, C. Koo, Smart tourism: foundations and developments, Electron. Markets 25 (3) (2015) 179–188.
[6] I. Arsal, K.M. Woosnam, E.D. Baldwin, S.J. Backman, Residents as travel destination information providers: an online community perspective, J. Travel Res. 49 (4) (2010) 400–413.
[7] H. Xie, L. Miao, P.J. Kuo, B.Y. Lee, Consumers' responses to ambivalent online hotel reviews: the role of perceived source credibility and pre-decisional disposition, Int. J. Hosp. Manage. 30 (2011) 178–183.
[8] H. Lee, A.R. Law, J. Murphy, Helpful reviewers in TripAdvisor: an online travel community, J. Travel Tour. Market. 28 (2011) 675–688.
[9] N. Yacouel, A. Fleischer, The role of cybermediaries in reputation building and price premiums in the online hotel market, J. Travel Res. 51 (2012) 219–226.
[10] A. Ghose, P.G. Ipeirotis, B. Li, Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content, Market. Sci. 31 (3) (2012) 493–520 (INFORMS.).
[11] Y.H. Hu, K. Chen, Predicting hotel review helpfulness: the impact of review visibility, and interaction between hotel stars and review ratings, Int. J. Inf. Manage. 36 (6) (2016) 929–944.
[12] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on Amazon.com, Manage. Inf. Syst. Q. 34 (1) (2010) 11.
[13] N. Hu, P. Pavlou, J. Zhang, Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication, Proceedings of the 7th ACM Conference on Electronic Commerce, New York, USA, 2006.
[14] P. Nelson, Information and consumer behavior, J. Polit. Econ. 78 (1970) 311.
[15] Z. Jiang, I. Benbasat, Virtual product experience: effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping, J. Manage. Inf. Syst. 21 (3) (2004) 111–147.
[16] P.M. Herr, F.R. Kardes, J. Kim, Effects of word-of-mouth and product-attribute information on persuasion: an accessibility-diagnosticity perspective, J. Consum. Res. 17 (1991) 454–462.
[17] H. Song, G. Li, Tourism demand modelling and forecasting—a review of recent research, Tour. Manage. 29 (2) (2008) 203–220.
[18] S. Gössling, C.M. Hall, Uncertainties in predicting tourist flows under scenarios of climate change, Clim. Change 79 (3–4) (2006) 163–173.
[19] J. Rosselló-Nadal, How to evaluate the effects of climate change on tourism, Tour. Manage. 42 (2014) 334–340.
[20] C. Goh, Exploring impact of climate on tourism demand, Ann. Tour. Res. 39 (4) (2012) 1859–1883.
[21] H.Q. Zhang, N. Kulendran, The impact of climate variables on seasonal variation in Hong Kong inbound tourism demand, J. Travel Res. (2016), doi: http://dx.doi.org/10.1177/0047287515619692.
[22] S.-M. Kim, P. Pantel, T. Chklovski, M. Pennacchiotti, Automatically assessing review helpfulness, Proceedings of the 2006 Conference of Empirical Methods in Natural Language Processing, Sydney, Australia, 2006.
[23] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, M. Zhou, Low-quality product review detection in opinion summarization, Comput. Ling. (2007) 334–342.
[24] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, Inf. Syst. Res. 19 (3) (2008) 291–313.
[25] Z. Zhang, Weighing stars: aggregating online product reviews for intelligent e-commerce applications, IEEE Intell. Syst. 23 (5) (2008) 42–49.
[26] Y. Liu, X. Huang, A. An, X. Yu, Modeling and predicting the helpfulness of online reviews, Proceedings IEEE International Conference on Data Mining, Pisa, Italy, 2008.
[27] J. Otterbacher, 'Helpfulness' in online communities: a measure of message quality, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2009, pp. 955–964.
[28] M.P. O'Mahony, B. Smyth, A classification-based review recommender, Knowledge-Based Syst. 23 (4) (2010) 323–329.
[29] C.C. Chen, Y.-D. Tseng, Quality evaluation of product reviews using an information quality framework, Decis. Support Syst. 50 (4) (2011) 755–768.
[30] A. Ghose, P.G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics, IEEE Trans. Knowl. Data Eng. 23 (10) (2011) 1498–1512.
[31] X. Yu, Y. Liu, X. Huang, A. An, Mining online reviews for predicting sales performance: a case study in the movie domain, IEEE Trans. Knowl. Data Eng. 24 (2012) 720–734.
[32] T.L. Ngo-Ye, A.P. Sinha, Analyzing online review helpfulness using a regressional relief-enhanced text mining method, ACM Trans. Manage. Inf. Syst. 3 (2) (2012) 10:1–10:20.
[33] Y. Liu, J. Jin, P. Ji, J.A. Harding, R.Y.K. Fung, Identifying helpful online reviews: a product designer's perspective, Comp. Aided Des. 45 (2) (2013) 180–194.
[34] R. Dong, M. Schaal, M.P. O'Mahony, B. Smyth, Topic extraction from online reviews for classification and recommendation, Third International Joint Conference on Artificial Intelligence, Beijing, China, 2013.
[35] T.L. Ngo-Ye, A.P. Sinha, The influence of reviewer engagement characteristics on online review helpfulness: a text regression model, Decis. Support Syst. 61 (2014) 47–58.
[36] N. Hu, N.S. Koh, S.K. Reddy, Ratings lead you to the product, reviews help you clinch it?: the mediating role of online review sentiments on product sales, Decis. Support Syst. 57 (2014) 42–53.
[37] S.-Y. Hwang, C.-Y. Lai, J.-J. Jiang, S. Chang, The identification of noteworthy hotel reviews for hotel management, Pacific Asia J. Assoc. Inf. Syst. (2014).
[38] G. Yin, L. Wei, W. Xu, M. Chen, Exploring heuristic cues for consumer perceptions of online reviews helpfulness: the case of Yelp.com, Proceedings of the 2014 Pacific Asia Conference on Information Systems, Chengdu, China, 2014.
[39] S. Lee, J.Y. Choeh, Predicting the helpfulness of online reviews using multilayer perceptron neural networks, Expert Syst. Appl. 41 (2014) 3041–3046.
[40] L. Martin, P. Pu, Prediction of helpful reviews using emotions extraction, AAAI Conference on Artificial Intelligence Twenty-Eighth AAAI Conference on Artificial Intelligence, Quebec, Canada, 2014.
[41] L. Zhu, G. Yin, W. He, Is this opinion leader's review useful? Peripheral cues for online review helpfulness, J. Electron. Comm. Res. 15 (4) (2014) 267–280.
[42] Z. Liu, S. Park, What makes a useful online review?: Implication for travel product websites, Tour. Manage. 47 (2015) 140–151.
[43] D. Weathers, S.D. Swain, V. Grover, Can online product reviews be more helpful? Examining characteristics of information content by product type, Decis. Support Syst. 79 (2015) 12–23.
[44] A.H. Huang, K. Chen, D.C. Yen, T.P. Tran, A study of factors that contribute to online review helpfulness, Comp. Hum. Behav. 48 (2015) 17–27.
[45] S.N. Ahmad, M. Laroche, How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis, Int. J. Electron. Comm. 20 (1) (2016) 76–111.
[46] A.Y. Chua, S. Banerjee, Helpfulness of user-generated reviews as a function of review sentiment: product type and information quality, Comp. Hum. Behav. 54 (2016) 547–554.
[47] B. Fang, Q. Ye, D. Kucukusta, R. Law, Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics, Tour. Manage. 52 (2016) 498–506.
[48] A. Qazi, K.B.S. Syed, R.G. Raj, E. Cambria, M. Tahir, D. Alghazzawi, A concept-level approach to the analysis of online review helpfulness, Comput. Hum. Behav. 58 (2016) 75–81.
[49] X. Wang, H.H. Teo, K.K. Wei, Simultaneity and interactivity of the effects of communication elements on consumers'decision making in ewom systems, J. Electron. Comm. Res. 16 (3) (2015) 153.
[50] M. Li, L. Huang, C.H. Tan, K.K. Wei, Helpfulness of online product reviews as seen by consumers: source and content features, Int. J. Electron. Comm. 17 (4) (2013) 101–136.
[51] R.M. Schindler, B. Bickart, Perceived helpfulness of online consumer reviews: the role of message content and style, J. Consum. Behav. 11 (2012) 234–243.
[52] A.M. Weiss, N.H. Lurie, D.J. MacInnis, Listening to strangers: whose responses are valuable, how valuable are they, and why? J. Market. Res. 45 (2008) 425–436.
[53] Q. Cao, W. Duan, Q. Gan, Exploring determinants of voting for the helpfulness of online user reviews: a text mining approach, Decis. Support Syst. 50 (2011) 511–521.
[54] H. Chen, D. Zimbra, AI and opinion mining, IEEE Intell. Syst. 25 (2010) 74–76.
[55] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, The International Conference on Language Resources and Evaluation, Valletta, Malta, 2010.
[56] Y. Pan, J.Q. Zhang, Born unequal: a study of the helpfulness of user-generated product reviews, J. Retailing 87 (2011) 598–612.
[57] N. Korfiatis, E. García-Bariocanal, S. Sánchez-Alonso, Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content, Electron. Comm. Res. Appl. 11 (2012) 205–217.

[58] J. Otterbacher, Gender, writing and ranking in review forums: a case study of the IMDb, Knowl. Inf. Syst. 35 (3) (2013) 645–664.
[59] U. Gretzel, K.H. Yoo, Use and impact of online travel reviews, Inf. Commun. Technol. Tour. 2008 (2008) 35–46.
[60] F. Bronner, R. de Hoog, Vacationers and eWOM: who posts, and why, where, and what? J. Travel Res. 50 (1) (2011) 15–26.
[61] E. Parra-López, J. Bulchand-Gidumal, D. Gutiérrez-Taño, R. Díaz-Armas, Intentions to use social media in organizing and taking vacation trips, Comput. Hum. Behav. 27 (2) (2011) 640–654.
[62] P.S. Fader, B.G.S. Hardie, K.L. Lee, RFM and CLV: using iso-value curves for customer base analysis, J. Market. Res. 42 (4) (2005) 415–430.
[63] R. Kahan, Using database marketing techniques to enhance your one-to-one marketing initiatives, J. Cons. Market. 15 (1998) 491–493.
[64] C. Marcus, A practical yet meaningful approach to customer segmentation, J. Consum. Market. 15 (5) (1998) 494–504.
[65] C.D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S.J. Bethard, D. McClosky, The natural language processing toolkit, Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, USA, 2014.
[66] E.A. Smith, J.P. Kincaid, Derivation and validation of the automated readability index for use with technical materials, Hum. Factors: J. Hum. Factors Ergon. Soc. 12 (5) (1970) 457–564.
[67] M. Coleman, T. Liau, A computer readability formula designed for machine scoring, J. Appl. Psychol. 60 (2) (1975) 283–284.
[68] J.P. Kincaid, Computer readability editing system, IEEE Trans. Prof. Commun. 24 (1) (1981) 38–42.
[69] R. Gunning, The fog index after twenty years, J. Bus. Commun. 6 (2) (1969) 3–13.
[70] G.H. McLaughlin, SMOG grading: a new readability formula, J. Read. 12 (8) (1969) 639–646.
[71] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Stroudsburg, USA, 2003.
[72] J. Wiebe, E. Riloff, Creating subjective and objective sentence classifiers from unannotated texts, Computational Linguistics and Intelligent Text Processing, Springer, Berlin Heidelberg, 2005, pp. 486–497.
[73] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software, ACM SIGKDD Explor. Newsl. 11 (1) (2009) 10.
[74] T.K. Ho, Random decision forests, Proceedings of 3rd International Conference on Document Analysis and Recognition (vol. 1), Montreal, Canada, 1995.
[75] R. Diao, Q. Shen, Feature selection with harmony search, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 42 (6) (2012) 1509–1523.
[76] M. Hu, B. Liu, Mining and summarizing customer reviews, Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 168–177.

Ya-Han Hu is currently an Associate Professor of the Department of Information Management at the National Chung Cheng University, Taiwan. He received a PhD degree in Information Management from the National Central University of Taiwan in 2007. His current research interests include text mining and information retrieval, clinical decision support systems, and recommender systems. His research has appeared in Decision Support Systems, Journal of the American Society for Information Science and Technology, IEEE Transactions on Systems, Man, and Cybernetics, Artificial Intelligence in Medicine, Applied Soft Computing, Computers in Human Behavior, Data & Knowledge Engineering, Expert Systems, Knowledge-Based Systems, Information Systems and e-Business Management, Journal of Information Science, Journal of Clinical Epidemiology, Methods of Information in Medicine, Online Information Review, and Journal of Systems and Software.

Kuanchin Chen is a Professor of Computer Information Systems at the Western Michigan University. Dr. Chen's research interests include electronic business, social networking, project management, privacy & security, online behavioral issues, business analytics, and human–computer interactions. He has published articles in journals and other academic publication outlets, including Information Systems Journal, Decision Support Systems, Information & Management, IEEE Transactions on Systems, Man, and Cybernetics, Internet Research, Journal of Database Management, Communications of the Association for Information Systems, Electronic Commerce Research and Applications, Journal of Global Information Management, DATA BASE for Advances in Information Systems, Decision Sciences Journal of Innovative Education, and many others. Dr. Chen serves on the editorial review boards of several academic journals.

Pei-Ju Lee is currently an Assistant Professor of the Department of Information Management at the National Chung Cheng University, Taiwan. She received her PhD degree in Information Sciences from University of Pittsburgh in 2015. Her current research interests include information fusion, data mining, database management, human-computer interaction, and human factor.

Please cite this article in press as: Y.-H. Hu, et al., The effect of user-controllable filters on the prediction of online hotel reviews, Inf. Manage. (2017), http://dx.doi.org/10.1016/j.im.2016.12.009