ARTICLE IN PRESS

International Journal of Hospitality Management xxx (xxxx) xxxx

Contents lists available at ScienceDirect

International Journal of Hospitality Management

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

Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach

Swagato Chatterjee*

IIT Kharagpur, Kharagpur, West Bengal, 721302, India

ARTICLE INFO

Keywords:
Online reviews
Sentiment mining
Emotion mining
Helpfulness
Polarity

ABSTRACT

Although online hotel reviews (OHR) help consumers in better decision–making, and service providers in better service design and delivery, they are hard to manage due to their high volume, velocity, and veracity. This paper focuses on the drivers of helpfulness of textual OHR, for which we have used text-mining techniques to find the sentiment content, polarity, and emotions; we have also used econometric and machine learning techniques to explain and predict its helpfulness. We found that content and title polarity lead to OHRs being less helpful, whereby this negative relationship gets accentuated with higher sentiment content. On the other hand, while negative emotion with low arousal makes OHR helpful, high arousal makes it less helpful. It has also been noted that after controlling for polarity, sentiment, and emotions, longer reviews are less helpful. Higher quantitative rating, recency of OHR and a reviewer's past expertise make a review more helpful. Additionally, machine-learning techniques have been found to predict 'review' helpfulness marginally better than econometric techniques. This study contributes to OHR literature in terms of its performance, and would also help decision makers in OHR management strategy.

1. Introduction

Consumers today have become more interconnected with each other than ever before with the booming business of e-marketplaces. Stories/incidents of their experiences and opinions about both products and services bind the consumer community together, whereby such expressions also lead influencing purchase decisions of future/prospective consumers. Online hotel reviews (OHR) is one such instance where such expressions and opinions are shared between the consumer communities in the specific context of hotels.

Online hotel reviews (OHRs) help prospective consumers to assess the service design and delivery of a hotel per se, while deciding on which would be the one to provide the highest value (Yang et al., 2018). Though the hotel provides service descriptions in online e-commerce channels, user-generated contents such as OHRs are perceived to be more trustworthy, thereby drawing the attention of a large pool of 'prospective' consumers (Bickart and Schindler, 2001; Chen and Xie, 2008). In fact, online reviews is the second most trusted source of information for consumers, just behind recommendations from family and friends (Grimes, 2012); thus, they do have a strong influence on the consumers' purchase decisions and thereby an organization's sales (Senecal and Nantel, 2004; Godes and Mayzlin, 2004; Liu, 2006; Zhu and Zhang, 2010; Filieri and McLeay, 2014). However, as OHRs have high volume, velocity, and veracity, they are often difficult to analyze for getting actionable insights.

Extant literature has tried to find what effectively makes a review helpful; in fact, studies have used survey-based methods to find the impact of informational and normative influences in the e-WOM on the adoption of online consumer reviews in the purchase decision-making process (Filieri, 2015). Others have focused on volume, sentiment, and polarity of reviews to predict review helpfulness (Mudambi and Schuff, 2010; Salehan and Kim, 2016); whilst some others have derived review helpfulness from reviewer characteristics (Ngo-Ye and Sinha, 2014). However, a comprehensive study encompassing both review characteristics (qualitative and quantitative) and reviewer characteristics has not been done yet. In this study, we try to focus on the relationship between quantitative ratings and information generated from qualitative textual data with review helpfulness. We have also considered the reviewer credibility as a predictor of review helpfulness. Based on user-generated content from a popular review website Trip Advisor, we have created explanatory and predictive models to find how the above-mentioned variables can influence and predict review helpfulness.

This study is important for a number of reasons: first, if one could correctly predict a potentially popular review, it could certainly help both the hotels and consumers alike. Whilst hotels could act based on whether the review is positive or negative, they could also look for

* Correspondence to: Vinod Gupta School of Management, IIT Kharagpur, Kharagpur, West Bengal, 721302, India.
E-mail address: swagato@vgsom.iitkgp.ac.in.

https://doi.org/10.1016/j.ijhm.2019.102356
Received 3 January 2019; Received in revised form 16 July 2019; Accepted 31 July 2019

Please cite this article as: Swagato Chatterjee, International Journal of Hospitality Management, https://doi.org/10.1016/j.ijhm.2019.102356

genuine feedback from such reviews, which in turn would lead them to enhance service quality. On the other hand, the consumer could use this mechanism to shortlist reviews, which are new but potentially helpful, which they could use for their purchase decision–making thereby reducing the risks associated with the purchase. Second, it is important for the hotel to know what kind of review characteristics lead to more helpfulness and popularity, thus for instance, if a hotel wants its 'satisfied' consumers to provide helpful reviews in order to enhance its branding and value proposition, it could look to motivate the 'satisfied' customers to write such reviews. Lastly, a reviewer could also know how to write good and helpful reviews from this study. With the advancement of the internet, review writing and blogging has become a business for many. Various bloggers such as travel bloggers, food bloggers etc., are paid based on how popular and helpful their reviews are. Therefore, this research will help such business too.

In the next part of this paper, we first discuss extant literature, and how the current study contributes to the same. Next, we develop an explanatory model along with hypotheses related to the same, which in turn has to be tested in the empirical study. Further, we discuss the validation process of the model proposed, and also create predictive models for review helpfulness. The penultimate section discusses the results in light of the theoretical contributions and managerial implications, followed by the limitations and future scope of the study.

2. Literature review

Explaining review helpfulness has been an important area of study for researchers in the areas of marketing, information systems, decision support system etc. As a matter of fact, a number of studies have tried to explain the review helpfulness from the basic information about the review, ratings, product types etc. For instance, Danescu-Niculescu-Mizil et al. (2009) focused on average rating, and deviance from it, as drivers of review helpfulness. Mudambi and Schuff (2010) have tried to explain review helpfulness using product type, review word count, rating etc. However, not all these studies have focused on the information available within a textual review, although it is expected to be a major driver of review helpfulness. Yang et al. (2017) did try to shed some light on reviewing helpfulness using both imagery and textual formats as predictors; however, even they have not explored the sentimental and emotional aspects expressed within the text.

Another stream of literature did use sentiment-mining techniques to use information generated from textual reviews to predict helpfulness of OHR. Baek et al., 2012 for instance, have only focused on negative word percentage along with other variables such as ratings, ranking, product type, price, word count etc. while explaining review helpfulness. Schindler and Bickart (2012) on the other hand, used proportion of positive-negative statements in order to explain review helpfulness, but were unable to find any strong support for the same. More recently, Salehan and Kim (2016), basing themselves on the theory of selective attention, the theory of selective perception along with the attribution theory found that review length, review sentiment and review polarity were major predictors; but still, a 'holistic' view of helpfulness of online reviews still seemed to be missing.

As regards performance drivers of OHR, extant research, again to some extent did focus on the same; but they weren't able to provide an automated scalable solution, which the service providers in turn could use to automate helpful online review detection or prediction (Salehan and Kim, 2016). Apart from human subject-based categorization of online reviews, which up to now has been the major contributor in getting an insight into the drivers of OHR (Salehan and Kim, 2016), there hasn't been much progress; in fact, this has remained a major limitation within extant literature, especially in this area of research.

Our study is similar and yet different from some of studies covered in more ways than one. First, ours is one of the very few studies, which focus on both explanatory and predictive models while studying helpfulness of OHRs. While explaining what makes a review helpful is important, prediction of 'potential helpfulness' of a new review, based on the data available is also equally important for both the hotel and prospective consumers; such studies not being available in extant literature, our study fills this gap, and therein lies its importance and originality. We have used a combination of econometric and machine learning techniques to explain as well as predict OHR helpfulness, while most of the earlier studies have used only the econometric technique. We added on machine-learning techniques, as they often provide a better predictive power, which extant literature of OHR helpfulness did not explore. Thus, introducing machine-learning techniques while creating a scalable model for automated detection is indeed an important addition to extant literature. This will not only help the hotel management to detect potentially helpful reviews automatically, but it would also look to harness the knowledge of predictive power of various algorithms in the helpfulness prediction problem, which at a later date could be empirically tested.

Moreover, what makes our study unique is that in our models, we've considered using qualitative data of textual reviews (i.e. review length, review sentiment, review polarity, title polarity, review emotions etc.), quantitative data from ratings (i.e. overall quantitative rating), alongside data about the reviewer (i.e. total helpful votes in past, time distance between experience and review posting). This, certainly is more exhaustive than most of the empirical studies available in this domain. Thus our study makes a mark in the extant literature of helpfulness of OHR.

We may add that this study also focuses on in-depth natural language processing techniques while analyzing textual data. Extant literature in this domain has only focused on polarity and sentiment content of the data (Salehan and Kim, 2016); however, one must note that emotions do play a very important role in consumer decision-making and service evaluation process (Laros and Steenkamp, 2005). We have thereby used text-mining techniques to find emotion scores of different valence, and different levels of arousal from the review data, minimizing thereby the research gap of non-consideration of consumer emotions.

Table 1 gives a comparative analysis of our study with past studies on similar domain.

3. Hypothesis development

3.1. Sentiment and polarity for the review text and review title

Vivid description in reviews are considered to be more trustworthy and helpful, as consumers try to create inference about a product and service quality from the information obtained from such reviews (Filieri, 2016). Therefore, consumers find reviews which are neutral in nature with both positive and negative views to be highly trustworthy (Filieri, 2016). Such reviews are expected to be high in total sentiment content, as both positive and negative sentiments are elaborated in such reviews. Extant literature has suggested that a good review should have both negative and positive content, and therefore sentiment words from both valences (Salehan and Kim, 2016). This is because, following the framing effect, two-sided arguments are expected to be more persuasive than one-sided arguments (Crowley and Hoyer, 1994). In fact, reviews with both positive and negative opinions are found to be more helpful (Cao et al., 2011). Therefore, irrespective of valence, high sentiment content within a review makes it more balanced and full of information. Such information in turn, reduces the search cost of future consumers, making the review more helpful (Liu and Park, 2015). Therefore an elaborate review which contains a high amount of sentiment would have a higher importance in the eyes of the consumers.

H1. *The higher the total sentiment content in the text of the review, the more helpful the review is perceived to be.*

Polarity has been defined as the extent by which a text is biased towards any direction of valence- i.e. positive or negative (Geetha et al.,

Table 1
Studies on helpfulness of OHR.

Paper	Methodology	Qualitative Review Data	Quantitative Rating Data	Reviewer Data	Explanatory Model	Predictive Model
Danescu-Niculescu-Mizil et al. (2009)	Empirical	Plagiarized Review	Deviation from average rating, Signed deviation	No	No	No
Mudambi and Schuff (2010)	Empirical	Review length	Rating	Total Votes	Tobit Regression	No
Baek et al. (2012)	Empirical	Review length, Negative words (%)	Rating, Ranking	Total votes	Hierarchical Regression	No
Schindler and Bickart (2012)	Empirical	Number of statements, Proportion of positive-negative statements, Proportion of product vs review descriptions	No	No	Regression	No
Salehan and Kim (2016)	Empirical	Review length, Review sentiment and Review polarity	No	No	Binomial Regression	No
Yang et al. (2017)	Empirical	Review length, Readability, Imagery data	No	No	Hierarchical Regression	No
Our Study	Empirical	Review length, Review sentiment, Review polarity, Title polarity, Review emotions	Overall quantitative rating	Total Helpful Votes in Past, Time distance between experience and review	Poisson Regression, Negative Binomial Regression	Poisson Regression, Negative Binomial Regression, Artificial Neural Network, Random Forest, Support Vector Machine

2017). Highly polar customer reviews are considered to be 'extreme reviews', i.e., extremely positive or extremely negative. Extant literature articulates that a consumer generally perceives extreme reviews to be untrustworthy (Filieri, 2016), as they tend to be skewed. Extreme positive valence reviews for instance, are written often in a 'marketing language' and not in a typical 'consumer writing style' (Filieri, 2016); herein, prospective consumers assume that such reviews are probably posted by the seller, or a competitor or a highly critical customer (Filieri, 2016; Salehan and Kim, 2016). Moreover, it should be noted that extreme sensational content would lead to consumer reactance, resulting thereby in an unfavorable judgment about the review (Salehan and Kim, 2016). Such distrust and unfavorable judgment towards extreme reviews lead to lesser diagnosticity and review helpfulness; thus, we posit:

H2. *The higher the polarity of the text of the review, the less helpful the review is perceived to be.*

Total sentiment content within a review text is good only when its polarity is low. Highly biased review with huge sentiment content in the text reduces the credibility and the informational value of the review which ultimately makes the review less helpful (Clare et al., 2018). A review, which is high in total sentiment content and polar, will tend to lead to high negative judgment from prospective consumers, and at the end of the day, would not be helpful at all, being perceived as 'untrustworthy'. However, reviews with high sentiment content and with neutral or low polarity would be judged as 'best reviews' (Salehan and Kim, 2016), leading us to our next hypothesis:

H3. *The interaction of review sentiment content and review polarity will have a negative relationship with review helpfulness. As the total amount of sentiment increase, the negative relationship between polarity and review helpfulness gets weaker.*

Consumers generally tend to have high search costs while evaluating a service or a product (Wang and Sahin, 2017). Therefore, they look for such information which effectively reduces the search cost from the consumer's perspective (Liu and Dukes, 2016). The title of the review gives a brief idea about what is written in the review itself (Salehan and Kim, 2016); thus, the title acts as an important source of quick information, which consumers could use to reduce search costs. However, as discussed above, highly polar title sentiments could lead to distrust towards the review, thereby making it less helpful. This effect is expected to be stronger for title polarity than review polarity, as title acts a stronger cue for service evaluation due to its ability to be concise and to the point. Hence we hypothesize:

H4. *The higher the title polarity score is, the less helpful the review is perceived to be.*

3.2. Emotions in the review text

Consumers are often skeptical about reviews posted online. They often think that such reviews are posted by the organization and its agencies and the reviews are not trustworthy (Filieri, 2016). Therefore, reviews with vivid negative emotions are considered to be more trustworthy. A paid review is less likely to have a vivid description of negative emotions. Thus, negative emotions increase the diagnosticity and trustworthiness of a review, making it more believable. However, other than valence, another important dimension of emotions is the level of arousal. Cavanaugh et al. (2016) have divided emotions based on valence and arousal. For instance, anger and fear have been categorized as high arousal negative emotions. On the other hand, sadness has been categorized as low arousal negative emotion. A review with high arousal may often lead to less information available, as the cognitive ability of a person faces hindrance under high arousal (Filieri, 2016; Salehan and Kim, 2016). Therefore, emotions associated with negative valence, and high arousal leads to lesser trustworthiness and helpfulness. We thereby posit:

H5. *The higher the emotions associated with negative valence and low arousal are expressed in a textual review, the more helpful the review is perceived to be.*

H6. *The higher the emotions associated with negative valence and high arousal are expressed in a textual review, the less helpful the review is perceived to be.*

3.3. Review length

The length of a review has been found to be one of the major predictors of the performance of a review per se (Mudambi and Schuff, 2010; Schindler and Bickart, 2012). Longer reviews for instance are expected to contain lots of information, while shorter ones are expected to be shallow, essentially a broad-level description of consumers' experiences. The information content of longer reviews reduces the search cost of consumers due to its increased diagnosticity of the stimuli (Johnson and Payne, 1985). Moreover, the persuasion power of a longer review increases, as multiple arguments are presented (Schwenk, 1986); this in turn, results in increased confidence level of the consumer, who now, is able to take a decision based on these reviews (Tversky and Kahneman, 1974). We thereby conclude that longer reviews are expected to be more helpful, leading us to posit:

H7. *Length of the review has a positive relationship with the review helpfulness*

3.4. Overall quantitative rating

Consumers often use non-compensatory purchase decision rules while evaluating alternatives (Johnson and Meyer, 1984); in such cases, products or services, which have a lower value than a cut-off in terms of an attribute, are not considered anymore for purchase decisions (Swait, 2001). The overall quantitative rating is one such attribute. It has often been observed that consumers tend to use filters to reduce their search costs while evaluating alternatives (Olbrich and Holsing, 2011). Various websites, such as Trip Advisor for instance, uses overall quantitative rating as one such filter. Therefore, with an increase of the overall quantitative rating, a review becomes more relevant to future purchasers who, in turn would use information from such reviews to help them in their decision-making. Moreover, it may be noted that often consumers who read a review already have a positive opinion about the service. A high overall rating would be more aligned to existing opinions of the reader, making it in turn more useful (Huang et al., 2015). On the other hand, there are also counter arguments suggesting that low ratings effectively make reviews more helpful. For instance, consumers often tend to depend more on low-rating reviews, as they have higher diagnostic value and higher depth (Ahluwalia et al., 2000; Mudambi and Schuff, 2010). In this study, we try to re-look at this phenomenon in existing literature, leading us to posit:

H8. *The higher the overall quantitative rating is, the more helpful the review is perceived to be.*

3.5. Time distance between experience and review posting

Consumers look for relevant information in online reviews because their purchase decisions would be affected by the experiences of other consumers. However, while some reviewers post the review soon after their experience of the product of the service, others tend to post reviews much later. It has been observed that the time distance between the experience and the review puts consumers in a higher construal level, which in effect, tends to reduce the negativity in their reviews, making them more abstract thereby (Huang et al., 2016). It should also be noted that higher time distance between the experience and the review can also lead to forgetting certain information about the experience; thereby this is less helpful to prospective consumers. We thus posit:

H9. *The higher the time distance between the experience and the review posting is, the less helpful the review is perceived to be.*

3.6. Past helpful votes of the reviewer

Source credibility is one of the major predictors of usefulness of word of mouth communication (McGinnies and Ward, 1980; Eagly and Chaiken, 1993). In the absence of relevant information about the source of information, consumers use information from a reviewer's profile, such as cues for ensuring credibility of the 'experienced' consumer (Park et al., 2014; Filieri, 2015). Moreover, the reviewer rating system adopted by various review websites also helps consumers judge the credibility of the author of a review (Cheung et al., 2009). For instance, Trip Advisor has introduced a badge system to identify the expertise of the reviewers (Filieri, 2015), whereby it also provides information about how many past cities the reviewer visited, how many hotels s/he visited, how many helpful votes the reviewer received in the past etc. All this information in an amalgamated form leads to higher credibility of the reviewer.

Moreover, reviews are written expressions of the experience a consumer had; thus, the usefulness of the review also depends on how well a reviewer can effectively articulate his/her experience in a written format. A basic indicator of such capability is how many helpful votes the reviewer has received in the past; this indicator does not only say how much the review is credible, but also says how popular s/he is, or how well s/he can express himself/herself in the reviews. For both of these above-mentioned reasons, we hypothesize that:

H10. *The higher the past helpful votes received by the author is, the more helpful the review is perceived to be.*

Fig. 1 gives the visual representation of the theoretical framework developed above.

4. Empirical study

4.1. Data description

Our data were collected from a renowned travel review website named Trip Advisor. We first scraped all the hotels listed in the review website, and then have randomly chosen 40 hotels for our analysis; reviews posted for these 40 hotels from 2005 to 2014 were gathered, resulting in 942 observations. The dataset contained the overall quantitative rating (in 5-point scale – 1 meaning very low and 5 meaning very high) along with the textual review given by reviewers. The textual reviews contained both the review content and the review title. We arrived at the 'review volume' by counting the number of words in each review. The data also contained the date of the review along with the travel date. The time distance between these two variables gave the measure of time distance between experience and review posting. We also collected helpful votes received by the reviews, and the total number of helpful votes received by the reviewer in the past.

Additionally, the data also contained insights about attribute-wise rating, travel type (family or business), reviewer data, such as the total number of reviews posted, the total number of hotel reviews posted, the total number of cities visited etc. However, these variables were later dropped from our explanatory models as they were found to be highly correlated with variables in the theoretical model.

4.2. Data processing

We created insights out of the textual reviews by using text-mining

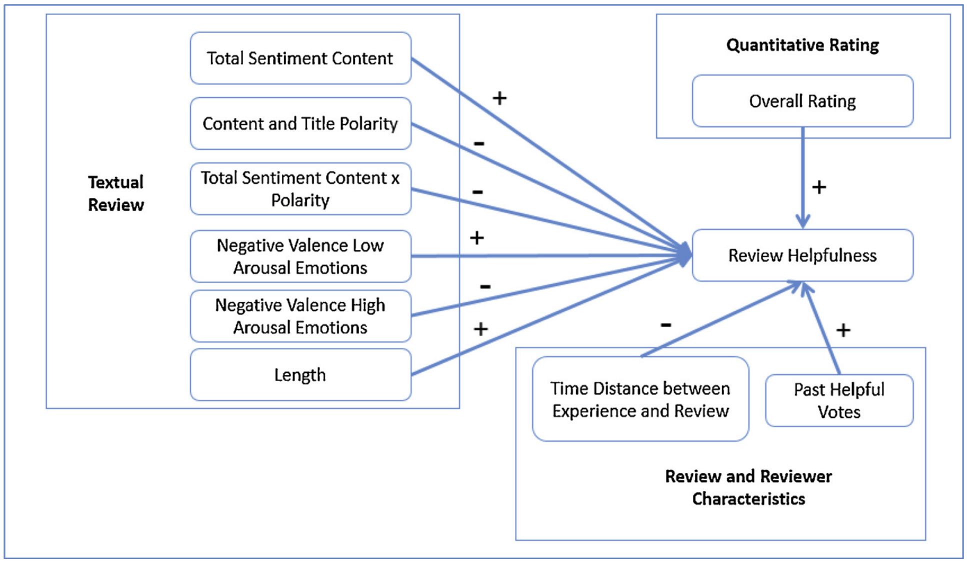

Fig. 1. Theoretical Model of Review Helpfulness.

techniques, which were later used in the models. The first step of such a text mining process was finding the overall sentiment of the textual review. Lexicons developed by researchers in the domain of computational linguistics has been used in marketing, information systems, and data science literature for finding out the sentiments of the texts (Dang et al., 2010; Taboada et al., 2011; Mostafa, 2013). The lexicons helped in categorizing the words either as being positive and/or negative. For sentiment analysis, we used the AFINN lexicon by Nielsen (2011), which lists English terms with an integer between -5 (negative) and +5 (positive). Following Salehan and Kim (2016), review polarity was calculated as the sentiment scores of all the words within a review; the same method was also used to find the title polarity. Other popular methods used for finding polarity included finding the ratio of both positive-valence words and negative-valence words (Geetha et al., 2017). However, we assume that our method is better equipped than the alternative method, as we provide the differential valence score to words with different levels of extremeness, unlike Geetha et al. (2017). Moreover, Geetha et al. (2017) have created polarity as a dichotomous variable, while we considered it as continuous variable.

Another important variable that has been used in our theoretical model is total sentiment score, i.e. the total number of words related to positive or negative sentiment available in a review. For this, we used NRC Word-Emotion Association Lexicon (also called EmoLex) created by Mohammad and Turney (2013), which gives the count of positive and negative sentiment words available within a review. The total sentiment score of the review is the total count of positive and negative words available within a review per se.

Along with overall sentiment and polarity, important insights may be generated from emotions expressed within the text too. While sentiment provides the overall valence of the word, emotions provide an in-depth knowledge of feelings and the mental states of individuals. We used the NRC Word-Emotion Association Lexicon (also called EmoLex) for emotion scores, as it provides the number of emotions such as anger, anticipation, disgust, fear, joy, sadness, surprise, trust (Mohammad and Turney, 2013). We checked the frequency in which the words related to the above emotions that occur in the text. We have first chosen negative emotions based on our theoretical model; out of which we specifically chose disgust, fear, and sadness, as they were found to not correlate with other variables in the model. Table 2 gives the brief details about the variables in the model post data processing.

Table 2
Descriptive Statistics of the Variables from the Theoretical Model.

	Mean	SD
Review Length	142.14	105.90
Polarity	15.07	11.52
Title Polarity	4.21	2.41
Overall Quantitative Rating	4.02	1.06
Time Distance Between Experience and Review	75.06	125.86
Disgust	0.77	1.38
Fear	1.32	1.68
Sadness	1.24	1.66
Helpful Votes in this Review	1.63	3.58
Total Past Helpful Votes of the Reviewer	2.52	2.39
Review Sentiment	12.3	8.46

4.3. Explanatory models

The following model as described in Equation 1 and 2 have been used to explain the review helpfulness in the data:

$$\textit{Review Helpfulness} = f(\beta_0 + \beta_1 \text{ x } \textit{Review Length} + \beta_2 \text{ x } \textit{Review Polarity} + \beta_3 \text{ x } \textit{Review Sentiment} + \beta_4 \text{ x } \textit{Title Polarity} + \beta_5 \text{ x } \textit{Overall Quantitative Rating} + \beta_6 \text{ x } \textit{Time Distance Between Experience and Review} + \beta_7 \text{ x } \textit{Disgust} + \beta_8 \text{ x } \textit{Fear} + \beta_9 \text{ x } \textit{Sadness} + \beta_{10} \text{ x } \textit{Total Past Helpful Votes of the Reviewer}) \quad (1)$$

$$\textit{Review Helpfulness} = f(\beta_0 + \beta_1 \text{ x } \textit{Review Length} + \beta_2 \text{ x } \textit{Review Polarity} + \beta_3 \text{ x } \textit{Review Sentiment} + \beta_4 \text{ x } \textit{Title Polarity} + \beta_5 \text{ x } \textit{Overall Quantitative Rating} + \beta_6 \text{ x } \textit{Time Distance Between Experience and Review} + \beta_7 \text{ x } \textit{Disgust} + \beta_8 \text{ x } \textit{Fear} + \beta_9 \text{ x } \textit{Sadness} + \beta_{10} \text{ x } \textit{Total Past Helpful Votes of the Reviewer} + \beta_{11} \text{ x } \textit{Review Polarity} \text{ x } \textit{Review Sentiment}) \quad (2)$$

We used both Poisson regression and negative binomial regression, and have done a comparative analysis of the results. We also did a negative binomial regression to take care of the over dispersion of the review helpfulness variable. We have added the mathematical details of these models in the supplementary material. We have also checked for possible quadratic effects in the model and included the results in supplementary material.

At first, we checked the correlation matrix, and ensured that the variables are not correlated with each other. Table 3 gives the correlation matrix of the independent variables mentioned in Equation 1 and 2. As we observed that some of the variables have a relatively high correlation, we also checked for VIF, wherein we found that the

Table 3
Correlation Matrix of Independent Variables.

	RC	OS	PL	TP	OR	TD	THV	DSG	FR	SAD
RC	1.00	0.66	0.39	−0.13	−0.19	−0.09	0.06	0.50	0.63	0.65
OS	0.66	1.00	0.42	−0.13	−0.15	−0.08	0.06	0.52	0.64	0.66
PL	0.39	0.42	1.00	0.24	0.41	−0.02	−0.04	−0.12	0.07	0.06
TP	−0.13	−0.13	0.24	1.00	0.37	0.02	−0.05	−0.27	−0.13	−0.13
OR	−0.19	−0.15	0.41	0.37	1.00	0.07	−0.08	−0.41	−0.32	−0.24
TD	−0.09	−0.08	−0.02	0.02	0.07	1.00	0.05	−0.10	−0.10	−0.06
THV	0.06	0.06	−0.04	−0.05	−0.08	0.05	1.00	0.03	0.06	0.07
DSG	0.50	0.52	−0.12	−0.27	−0.41	−0.10	0.03	1.00	0.56	0.59
FE	0.63	0.64	0.07	−0.13	−0.32	−0.10	0.06	0.56	1.00	0.67
SAD	0.65	0.66	0.06	−0.13	−0.24	−0.06	0.07	0.59	0.67	1.00

RC = Review Count, OS = Overall Review Sentiment, PL = Review Polarity, TP = Title Polarity, OR = Overall Quantitative Rating, TD = Time Distance Between Experience and Review, THV = Total Helpful Votes by the reviewer, DSG = Disgust, FR = Fear, SAD = Sadness.

Table 4
Results of Regression Models Explaining Review Helpfulness.

	Poisson Regression 1	Poisson Regression 2	Negative Binomial Regression 1	Negative Binomial Regression 2	Hypothesis Supported
AIC Value	3161.7	3158.9	2106.9	2108.7	
(Intercept)	0.19ˆ	0.1ˆ	−0.10ˆ	−0.12ˆ	
RC	−0.00***	−0.00***	0.00ˆ	0.00ˆ	Not H7
OS	−0.00ˆ	0.00ˆ	−0.00ˆ	0.00ˆ	Not H1
PL	−0.04***	−0.05***	0.02ˆ	0.02ˆ	H2
TP	−0.11***	−0.11***	−0.10***	−0.11***	H4
OR	0.10**	0.09*	0.14˜	0.14˜	H8
TD	−0.00*	−0.00*	−0.00ˆ	−0.00ˆ	H9
DSG	0.14***	0.12***	0.14*	0.13*	H5
FR	−0.11**	−0.12***	−0.14*	−0.14*	H6
SAD	0.06*	0.06*	0.09˜	0.10˜	H5
THV	0.06***	0.06***	0.07**	0.07**	H10
OS x PL		0.00*		0.00ˆ	H3

*** means $p < 0.001$; ** means $p < 0.01$; * means $p < 0.05$, ˜ means $p < 0.1$, ˆ means NS.
RC = Review Count, OS = Overall Review Sentiment, PL = Review Polarity, TP = Title Polarity, OR = Overall Quantitative Rating, TD = Time Distance Between Experience and Review, THV = Total Helpful Votes by the reviewer, DSG = Disgust, FR = Fear, SAD = Sadness.

multicollinearity problem is nonexistent in all the models.

Table 4 summarizes the results found after analyzing the above-mentioned models. The Akaike Information Factor (AIC) value for the negative binomial (NB) models is better than the Poisson regression, suggesting thereby NB models have better goodness of fit than the Poisson model. Moreover, introducing the interaction term between review polarity and review sentiment content reduced the AIC in both the models, thereby improving the models. This, in turn suggests that such interaction effect is important to consider while explaining review helpfulness.

We did not find any significant relationship of total sentiment content with review helpfulness; however, in the Poisson models, polarity has been found to have negative relationships with review helpfulness. Thus, while H1 was not supported, H2 was. The Poisson model also suggests a significant positive relationship between the interaction of review sentiment content and review polarity with review helpfulness, which in turn suggests that as the total amount of sentiment increases, negative relationship between polarity and review helpfulness gets weaker, supporting thereby H3.

All models found a significant negative relationship between title polarity and review helpfulness, thereby supporting H4. Negative and low arousal emotions, such as disgust or sadness had a positive relationship with review helpfulness. However, negative and high arousal emotions, such as fear was found to have a negative relationship with review helpfulness, thereby supporting H5 and H6. According to the Poisson models, the overall quantitative rating has been found to have a positive and significant relationship with review helpfulness, supporting H8. On the other hand, time distance between the experience and the review posting has been found to have a negative and significant relationship with review helpfulness, supporting thereby H9. All models found a significant positive relationship between past helpful votes received by the author, and review helpfulness, supporting thereby H10.

As per the Poisson models, the length of the review has been found to have a negative relationship with review helpfulness. This result is the exact opposite to our hypothesis H7. In the discussion part, we discussed the implications of the results in greater details. We checked for possible endogeneity in the model, and did not find any significant correlation or even partial correlation between independent variables and residuals, thereby suggesting exogeneity.

4.4. Predictive models

To predict review helpfulness using the variables from our theoretical model, we have tried multiple techniques, stemming from both econometric and machine learning domains. The predictive power of these techniques has been analyzed using standard evaluation techniques, whereby we tested three different predictive models: the first, did not include any sentiment and emotion analysis data; the second, had only sentiment data on the review along with the title; while the last one, had all variables including sentiment, polarity, and other emotion-related variables.

Along with the two regression techniques (i.e. Poisson regression and NB regression), machine learning techniques, generally used for predictive models include: artificial neural network (ANN), random forest (RF) and support vector machine (SVM). ANN consist of a variety of (computational) neurons that comes in multiple interconnected layers including input, hidden and output layers. Such a network is made to mimic the human neural network, whereby each neuron has an activation function to process inputs coming from connections, with each connection having certain weights assigned to it. The weighted sum of outputs from neurons in earlier layers goes through such activation function, and are transmitted by the output neuron (Han et al., 2011). During the training process, weights assigned to the connections and the activation functions are changed through iterations such that the final output is close to the actual dependent variable (Nisbet et al., 2009). In this study, we apply a feed-forward neural network using back-propagation. The most frequently used activation function is the sigmoid function (Fuller et al., 2009). We also used the same. The 'nnet' package has been used in R software for the analysis (Venables and Ripley, 2002). The package can simulate a feed-forward neural networks with a single hidden layer with a backpropagation algorithm as the training algorithm (Werbos, 1994). Following Raczko and Zagajewski (2017), we used the following parameters: 24 neurons in the hidden layer, decay value as 0.0005 and training stop at 0.0001.

We also used another popular machine learning technique named Random Forest developed by Breiman (2001). This method can handle high dimensional data well and can resist overfitting (Breiman, 2001). In RF, multiple decision trees are developed using different training data, which are a sample of the overall data available for training. Such sampling is done by selecting a subset of the training data with repetition (Satapathy et al., 2016). Repeated rehashing the accompanying steps for every terminal node of the tree is done to develop a RF tree till the minimum node size is arrived at. Finally, a subset of the nodes from each tree is selected to find the results in such nodes. Moreover, the results of multiple trees are ensemble to find the result of the random forest (Satapathy et al., 2016). Such ensemble of classification trees are done such that each tree votes on the class assigned to a given sample, with the most frequent answer winning the vote (Sun and Schulz, 2015). We have used 'randomForest' package in R (Liaw and Wiener, 2002). We have used the following two parameters: the number of predictors taken into consideration at each fork of the tree = 24 and the number of random trees assembled during model building = 500 (Raczko and Zagajewski, 2017).

Another machine learning technique used, is SVM; it finds the maximum margin hyperplane that maximizes the distance between observations from different groups (Kotsiantis et al., 2007; Vapnik, 2013). Linear separation of the observations is often impossible; thus, SVM uses non-linear kernel functions. By and large, this technique tries to find the optimal hyperplane in n-dimensional classification space, which results in highest margin between classes (Vapnik, 2013). Extant literature often reports SVM to be a better classifier than others (Huang et al., 2002; Ghosh et al., 2014). We used the 'e1071′ package in R for this analysis (Meyer et al., 2015). We have used a Radial Basis Function as our kernel and grid-search heuristic based parameter selection process following past literature (Siering et al., 2018).

We divided the data into 70:30 ratio of training and testing data, where 70% of the data was used for training the models, and the rest 30% was used to test the models. To evaluate our models, we calculated the root mean square error (RMSE) of the predictions coming out of each technique. To reduce the impact of the sampling during the training and testing data creation, we ran the operation (as mentioned above) 100 times, and reported the mean and standard deviations of the RMSE of the models. Table 5 details the same. Based on our findings, SVM performs better than all other models followed by the regression techniques. ANN and RF performed worst out of the ones tested. As the sentiment and emotion-related variables are considered to predict OHR, the performance of the models are found to be dropped. Although such drop is not significant in any of the techniques. This suggests that the prediction of the review helpfulness can be achieved without the sentiment analysis too.

5. Discussion

This study has focused on explaining and predicting OHR helpfulness based on quantitative and qualitative information generated from a review. Based on data collected from a popular online hotel review website, we have done Poisson regression and negative binomial regression to find the explaining power of various variables towards review helpfulness. The results suggested that total sentiment content in the review text has no impact on review helpfulness. This result is contrary to existing knowledge, which says reviews that are neutral in nature, are highly trustworthy (Filieri, 2016; Salehan and Kim, 2016) and thereby, are expected to be more helpful. On the other hand, polarity reduces review helpfulness; this result is in line with existing literature, which says extreme reviews are untrustworthy, and often written by the seller, or a competitor or a highly critical customer (Filieri, 2016; Salehan and Kim, 2016); therefore, consumers have unfavorable judgments about such reviews. Our results also suggested that the negative impact of polarity worsens when the sentiment content is high; this too is supported by extant literature, which says reviews with high sentiment content and with neutral or low polarity will be judged as the best reviews (Salehan and Kim, 2016).

While checking for the impact of title polarity, we found that it reduced the 'helpfulness' of a review. A title is important as long as it gives a brief idea of the textual review, and reduces the information search cost, while evaluating the alternative stage of the purchase decision-making process (Liu and Dukes, 2016; Salehan and Kim, 2016; Wang and Sahin, 2017). However, a polar title would lead to distrust, thereby making a review less helpful.

We categorized the emotions expressed within the review text based on valence and arousal, following Cavanaugh et al. (2016). We found that negative emotions with low arousal has in effect higher helpfulness scores. This is because, positive textual reviews are less trustworthy than negative reviews, due to the consumer's perception that the reviews are written in 'marketing language' and are often fake (Filieri, 2016). Therefore, emotions like disgust and sadness lead to higher trustworthiness, increasing thereby the helpfulness of the review. However, emotions like anger, which are high in arousal and negative in valence, signals less usage of cognitive effort while writing the review, in turn reducing the 'believability' factor (Filieri, 2016; Salehan and Kim, 2016). Therefore, emotions associated with negative valence and high arousal leads to lesser trustworthiness and helpfulness.

The study suggests that overall quantitative rating has a positive and significant relationship with review helpfulness. Such ratings help consumers in better and faster decision making while they follow non-compensatory purchase decision rules while evaluating alternatives (Johnson and Meyer, 1984; Olbrich and Holsing, 2011). Thus our result is supported by the theories in extant literature. We also found that the time distance between the experience and the review posting has a negative and significant relationship with review helpfulness. The time distance between the experience and the review, leads to a higher construal level in the consumers' minds, and this is what reduces the negative outcome of the reviews. Moreover, reviews posted soon after the experience, is certainly richer in information; thus, our results corroborate with extant literature.

The results suggest a significant positive relationship between past helpful votes received by the author and review helpfulness. Reviewer demographics such as past helpful votes, reviewer badge levels etc. provide credibility of the reviewer (Cheung et al., 2009; Filieri, 2015), which is important for WOM (McGinnies and Ward, 1980; Eagly and Chaiken, 1993). The helpful votes also is a signal of the reviewer's ability to articulate and express his/her experience in great detail. Therefore, the positive relationship between past helpful votes received by the author and review helpfulness is justified.

We found that the higher length of the review text leads to lower helpfulness. However, extant literature has found opposite results (Mudambi and Schuff, 2010; Schindler and Bickart, 2012). The reasons suggested are higher diagnosticity of the stimuli (Johnson and Payne, 1985), better persuasion power (Schwenk, 1986), and increasing confidence (Tversky and Kahneman, 1974). Therefore longer reviews are expected to be more helpful. However, the results obtained by us is

Table 5
Details of Root Mean Square Error for Various Predictive Analysis Techniques after 100 iterations.

	No sentiment and Emotion Data		With Sentiment Data		With Sentiment and Emotion Data	
	Mean	SD	Mean	SD	Mean	SD
Poisson Regression	1.59	0.19	1.62	0.19	1.64	0.19
Negative Binomial	1.6	0.2	1.63	0.2	1.65	0.19
Artificial Neural Network	1.87	0.13	1.87	0.12	1.89	0.12
Random Forest	1.83	0.13	1.87	0.11	1.87	0.11
Support Vector Machine	1.55	0.18	1.55	0.18	1.57	0.17

opposite; this could be because, post considering the information available in the text in the form of emotions, sentiments, valence etc., the length of the review does not add to any diagnosticity or persuasion power. Moreover, a lengthy text requires higher information processing, thereby increasing the search costs of consumers. Thus, after controlling the information context, length of a textual review would be negatively related to review helpfulness.

We also focused on the predictive power of the explanatory variables when used in various econometric and machine learning techniques. We found that SVM outperforms other models followed by the regression techniques. ANN and RF performed worst out of the ones tested. The results also show low predictive power of the textual components, suggesting in turn that although the insights from textual reviews may explain review helpfulness significantly, they often cannot predict review helpfulness of future reviews. Therefore, it is suggested that a predictive model should include only quantitative rating data of the consumers.

5.1. Theoretical contribution

The current study has a number of theoretical contributions. First, this study is one of the very few studies, which focuses on both explanatory and predictive models while studying helpfulness of OHRs (Siering et al., 2018). We have shown that the performance of econometric and machine learning methods in predicting review helpfulness are comparable. On the other hand, econometric methods also allow explaining the underlying reason of why a review is more helpful than other. Therefore, we provide a comparative analysis of the advantages and disadvantages of using econometric and machine learning methods as the method of data analysis and inference generation, contributing to the literature of applied statistics and computation too.

Second, this paper is one of the very few papers, which combine information generated from both qualitative data of the textual review (review length, review sentiment, review polarity, title polarity, review emotions) and quantitative data from the rating (overall quantitative rating) while studying review helpfulness. We also considered the effect of contextual covariates (i.e. total helpful votes in past, time distance between experience and review posting) in our models, which in fact, made our model more exhaustive. Thus, our study contributes to extant literature of review helpfulness of OHR by giving a comprehensive empirical framework (Mudambi and Schuff, 2010; Baek et al., 2012; Schindler and Bickart, 2012; Salehan and Kim, 2016).

Third, the paper uses text mining techniques to find the sentiments and emotions expressed within a review text, along with their impact on review helpfulness. Extant literature on OHR helpfulness has only studied the impact of sentiment content on review helpfulness (Salehan and Kim, 2016). However, the impact of emotions in this area of research is new. Moreover, we have also found that the impact of such emotional contents in the text varies based on the type of emotions (Cavanaugh et al., 2016). More specifically, emotions with negative valence and low arousal lead to higher review helpfulness. However, emotions with negative valence and high arousal lead to lower review helpfulness. These findings are an extension of consumer psychology and emotion literature in the context of OHR, and a noble contribution in this domain of research (Filieri, 2016; Salehan and Kim, 2016).

5.2. Managerial implications

The paper is also important for managers for a number of reasons. First, it gives an idea about how to predict potentially popular reviews; this, we believe would help the sellers take corrective actions against potentially harmful reviews and promote actions towards positive reviews, which potentially are helpful. On the other hand, a consumer can also take better decisions if they know which reviews are potentially not helpful or misleading and which are not; this, in turn would help them in better decision-making, leading thereby to post-purchase satisfaction.

Second, information about the drivers of review helpfulness is important for sellers. This information will help them in the review solicitation process, which they can use to promote their business. Drivers of review helpfulness are also important for a reviewer who can get information about how to write a good and helpful review. In this digital age, such information is important for many. Moreover, digital platforms, which collect consumer reviews can also focus on the popularity of their platform by promoting more helpful reviews and demoting less helpful votes. Therefore, knowledge about the drivers of the helpfulness of the review is important for digital platforms too.

Lastly, this paper explores the process of integrating qualitative data and quantitative data obtained from consumer reviews. This is also an important aspect for managers who often rely on qualitative methods for exploratory analysis and quantitative methods for in-depth analysis. These two methods are often separate and sequential. However, a methodology, which can integrate both in management decision-making process makes the decision making information rich and more effective.

5.3. Limitations and future score

The paper has remained limited in terms of the data source. The data has been collected from only from one website. Future research should try and integrate data from multiple online sources to find the impact of drivers of review helpfulness. Moreover, we did not focus on the demographics of the reviewer in depth. Variables such as the cultural background, the origin of the reviewer, gender, age etc. can also impact review helpfulness. Cultural and mental congruity between the reviewer and the reader may make a review more helpful; as a matter of fact, the psychological frame of a reviewer could also impact review helpfulness. Future research should focus on such unanswered questions and focus on other service contexts, while our study majorly focused on hotel reviews. For instance, the impact of sentiment and emotions on review helpfulness may be very different in services such as airlines than that of hotels. While fear and trust are expected to be most important for airlines, the relative importance of such aspects might be very different for the hotel. Future research should focus on such contextual effects on the impact of drivers of review helpfulness.

6. Conclusion

This study has focused on explaining and predicting OHR helpfulness based on quantitative and qualitative information generated from a review. While total sentiment content in the text has no relationship with review helpfulness, text polarity and title polarity have a negative relationship, which gets prominence when the sentiment content is high. Moreover, reviews with negative emotions and low arousal, such as disgust and sadness, have higher helpfulness scores. However, high arousal negative emotions such as anger leads to lesser helpfulness. After controlling for textual emotions, sentiments and valence etc., the length of the review has a negative impact on review helpfulness. While overall quantitative rating has a positive relationship with review helpfulness, the time distance between the experience and the review posting has a negative relationship. Also, reviewers with higher number of past helpful votes get higher helpfulness votes, and textual information has lower predictive power than quantitative ratings in OHR.

The current study is an important addition to extant literature of review helpfulness of OHRs, and we believe that it will play a major role in bridging the gap between existing knowledge and future research.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijhm.2019.102356.

References

Ahluwalia, R., Burnkrant, R.E., Unnava, H.R., 2000. Consumer response to negative publicity: the moderating role of commitment. J. Mark. Res. 37 (2), 203–214.
Baek, H., Ahn, J., Choi, Y., 2012. Helpfulness of online consumer reviews: readers' objectives and review cues. Int. J. Electron. Commer. 17 (2), 99–126.
Bickart, B., Schindler, R.M., 2001. Internet forums as influential sources of consumer information. J. Interact. Mark. 15 (3), 31–40.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Cao, Q., Duan, W., Gan, Q., 2011. Exploring determinants of voting for the "helpfulness" of online user reviews: a text mining approach. Decis. Support Syst. 50 (2), 511–521.
Cavanaugh, L.A., MacInnis, D.J., Weiss, A.M., 2016. Perceptual dimensions differentiate emotions. Cogn. Emot. 30 (8), 1430–1445.
Chen, Y., Xie, J., 2008. Online consumer review: word-of-mouth as a new element of marketing communication mix. Manage. Sci. 54 (3), 477–491.
Cheung, M.Y., Luo, C., Sia, C.L., Chen, H., 2009. Credibility of electronic word-of-mouth: informational and normative determinants of on-line consumer recommendations. Int. J. Electron. Commer. 13 (4), 9–38.
Clare, C.J., Wright, G., Sandiford, P., Caceres, A.P., 2018. Why should I believe this? Deciphering the qualities of a credible online customer review. J. Mark. Commun. 24 (8), 823–842.
Crowley, A.E., Hoyer, W.D., 1994. An integrative framework for understanding two-sided persuasion. J. Consum. Res. 20 (4), 561–574.
Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., Lee, L., 2009. How opinions are received by online communities: a case study on amazon.com helpfulness votes. April. Proceedings of the 18th International Conference on World Wide Web 141–150.
Dang, Y., Zhang, Y., Chen, H., 2010. A lexicon-enhanced method for sentiment classification: an experiment on online product reviews. IEEE Intell. Syst. 25 (4), 46–53.
Eagly, A.H., Chaiken, S., 1993. The Psychology of Attitudes. Harcourt Brace Jovanovich College Publishers.
Filieri, R., 2015. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. J. Bus. Res. 68 (6), 1261–1270.
Filieri, R., 2016. What makes an online consumer review trustworthy? Ann. Tour. Res. 58, 46–64.
Filieri, R., McLeay, F., 2014. E-WOM and accommodation: an analysis of the factors that influence travelers' adoption of information from online reviews. J. Travel. Res. 53 (1), 44–57.
Fuller, C.M., Biros, D.P., Wilson, R.L., 2009. Decision support for determining veracity via linguistic-based cues. Decis. Support Syst. 46 (3), 695–703.
Geetha, M., Singha, P., Sinha, S., 2017. Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tour. Manag. 61, 43–54.
Ghosh, A., Fassnacht, F.E., Joshi, P.K., Koch, B., 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. Int. J. Appl. Erath Observ. Geoinform. 26, 49–63.
Grimes, M., 2012. Nielsen: Global Consumers' Trust in 'Earned' Advertising Grows in Importance. Retrieved from https://www.nielsen.com/us/en/press-room/2012/nielsen-global-consumers-trust-in-earned-advertising-grows.html on 16th November, 2018. .
Godes, D., Mayzlin, D., 2004. Using online conversations to study word-of-mouth communication. Mark. Sci. 23 (4), 545–560.
Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier.
Huang, A.H., Chen, K., Yen, D.C., Tran, T.P., 2015. A study of factors that contribute to online review helpfulness. Comput. Human Behav. 48, 17–27.
Huang, C., Davis, L.S., Townshed, J.R.G., 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sens. 23, 725–749.
Huang, N., Burtch, G., Hong, Y., Polman, E., 2016. Effects of multiple psychological distances on construal and consumer evaluation: a field study of online reviews. J. Consum. Psychol. 26 (4), 474–482.
Johnson, E.J., Meyer, R.J., 1984. Compensatory choice models of noncompensatory processes: the effect of varying context. J. Consum. Res. 11 (1), 528–541.
Johnson, E.J., Payne, J.W., 1985. Effort and accuracy in choice. Manage. Sci. 31 (4), 395–414.
Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: a review of classification techniques. Emerg. Artif. Intell. Appl. Comput. Eng. 160, 3–24.
Laros, F.J., Steenkamp, J.B.E., 2005. Emotions in consumer behavior: a hierarchical approach. J. Bus. Res. 58 (10), 1437–1445.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.
Liu, Y., 2006. Word of mouth for movies: its dynamics and impact on box office revenue. J. Mark. 70 (3), 74–89.
Liu, Z., Park, S., 2015. What makes a useful online review? Implication for travel product websites. Tour. Manag. 47, 140–151.
Liu, L., Dukes, A., 2016. Consumer search with limited product evaluation. J. Econ. Manag. Strategy 25 (1), 32–55.
McGinnies, E., Ward, C.D., 1980. Better liked than right: trustworthiness and expertise as factors in credibility. Pers. Soc. Psychol. Bull. 6 (3), 467–472.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2015. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, R Package Version 1.6-7. http://CRAN.R-project.org/package=e1071.
Mohammad, S.M., Turney, P.D., 2013. Crowdsourcing a word–emotion association lexicon. Comput. Intell. 29 (3), 436–465.
Mostafa, M.M., 2013. More than words: social networks' text mining for consumer brand sentiments. Expert Syst. Appl. 40 (10), 4241–4251.
Mudambi, S.M., Schuff, D., 2010. Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. MIS Quarterly. pp. 185–200.
Ngo-Ye, T.L., Sinha, A.P., 2014. The influence of reviewer engagement characteristics on online review helpfulness: a text regression model. Decis. Support Syst. 61, 47–58.
Nielsen, F..Å., 2011. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. Workshop on Making Sense of Microposts: Big Things Come in Small Packages. pp. 93–98.
Nisbet, R., Elder, J., Miner, G., 2009. Handbook of Statistical Analysis and Data Mining Applications. Academic Press.
Olbrich, R., Holsing, C., 2011. Modeling consumer purchasing behavior in social shopping communities with clickstream data. Int. J. Electron. Commer. 16 (2), 15–40.
Park, H., Xiang, Z., Josiam, B., Kim, H., 2014. Personal profile information as cues of credibility in online travel reviews. Anatolia 25 (1), 13–23.
Raczko, E., Zagajewski, B., 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. Eur. J. Remote. Sens. 50 (1), 144–154.
Salehan, M., Kim, D.J., 2016. Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics. Decis. Support Syst. 81, 30–40.
Satapathy, S.M., Acharya, B.P., Rath, S.K., 2016. Early stage software effort estimation using random forest technique based on use case points. Iet Softw. 10 (1), 10–17.
Schindler, R.M., Bickart, B., 2012. Perceived helpfulness of online consumer reviews: the role of message content and style. J. Consum. Behav. 11 (3), 234–243.
Schwenk, C.H., 1986. Information, cognitive biases, and commitment to a course of action. Acad. Manag. Rev. 11 (2), 298–310.
Senecal, S., Nantel, J., 2004. The influence of online product recommendations on consumers' online choices. J. Retail. 80 (2), 159–169.
Siering, M., Deokar, A.V., Janze, C., 2018. Disentangling consumer recommendations: explaining and predicting airline recommendations based on online reviews. Decis. Support Syst. 107, 52–63.
Sun, L., Schulz, K., 2015. The improvement of land cover classification by thermal remote sensing. Remote Sens. 7, 8368–8390.
Swait, J., 2001. A non-compensatory choice model incorporating attribute cutoffs. Transp. Res. Part B Methodol. 35 (10), 903–928.
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M., 2011. Lexicon-based methods for sentiment analysis. Comput. Linguist. 37 (2), 267–307.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.
Vapnik, V., 2013. The Nature of Statistical Learning Theory. Springer Science & Business Media.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics With S, fourth ed. Springer, New York ISBN 0-387-95457-0, 462.
Wang, R., Sahin, O., 2017. The impact of consumer search cost on assortment planning and pricing. Manage. Sci. 64 (8), 3469–3970.
Werbos, P., 1994. The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting (Adaptive and Learning Systems for Signal Processing, Communications and Control Series). pp. 342. John Wiley & Sons, New York.
Yang, S.B., Hlee, S., Lee, J., Koo, C., 2017. An empirical examination of online restaurant reviews on Yelp. com: a dual coding theory perspective. Int. J. Contemp. Hosp. Manage. 29 (2), 817–839.
Yang, Y., Park, S., Hu, X., 2018. Electronic word of mouth and hotel performance: a meta-analysis. Tour. Manag. 67, 248–260.
Zhu, F., Zhang, X., 2010. Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. J. Mark. 74 (2), 133–148.