Electronic Markets (2020) 30:791–804
https://doi.org/10.1007/s12525-020-00397-5

RESEARCH PAPER

Exploring thematic composition of online reviews: A topic modeling approach

Vamsi Vallurupalli[1] · **Indranil Bose**[1]

Received: 23 December 2018 / Accepted: 7 January 2020 / Published online: 21 January 2020

Abstract
Online reviews are a critical component of the retail business ecosystem today. They help consumers share feedback and readers make informed choices. As such, it is important to understand the mechanism driving the creation of reviews and identify factors which make them useful for readers. Extant work in this field has largely ignored the distribution of thematic content in reviews and its role in review diagnosticity. This article attempts to bridge the gap. A novel approach is proposed to explore the distribution of thematic content in reviews, in terms of underlying topics, and test its impact on influence of reviews. The approach is illustrated through a case study using data from Yelp. Implications of the study for theory and practice are discussed.

Keywords Latent Dirichlet allocation · Online reviews · Review influence · Thematic content · Topic modeling · Yelp

JEL Classification M31 Marketing

Introduction

Online reviews have become ubiquitous in consumer facing businesses today. They allow consumers to share feedback regarding a product and readers to make informed purchase decisions. The importance of reviews has been highlighted in both academic research and industry reports. Anderson (2014) has reported that about 90% of consumers trust online reviews as much as personal recommendations. Based on a survey, Podium (2018) has reported that reviews impact purchase decisions for 93% of consumers. Also, it has been found that two-thirds of consumers are willing to pay more if they are assured of a better experience. It has therefore become critical for businesses to monitor online reviews written for their product(s). Correspondingly, it is important for academic researchers to understand the mechanism driving the creation, consumption, and overall impact of reviews on a business (Corley et al. 2013).

Despite extensive research on online reviews in recent years, there has been little focus on the thematic composition of reviews and the association between thematic content and review influence. Thematic composition of a document is defined by distribution of underlying topics, and is a representation of granular information contained in a document. A detailed examination of the thematic composition of reviews can help online retailers and review sites identify the most relevant reviews to be displayed to consumers, which in turn may boost consumer value and satisfaction. As such, a generic approach to explain the thematic composition of reviews and examine its impact is important and a strong need for e-commerce providers.

Towards this end, a novel process-based approach using Latent Dirichlet Allocation (LDA) has been proposed in this article. The approach is exploratory in nature, and inductively identifies topics embedded in the reviews. Using the theoretical lens of information diagnosticity and quality cues, the topical composition is used to discover cues affecting the purchase decision of the consumer. Correspondingly, an investigation of the relationship between the discovered topics and review influence is included as a part of the approach. To summarize, the approach can help answer the following questions for a given set of reviews.

Responsible Editor: Steven Bellman

✉ Indranil Bose
bose@iimcal.ac.in

Vamsi Vallurupalli
vallurupalliv13@email.iimcal.ac.in

[1] Indian Institute of Management Calcutta, Diamond Harbour Road, Kolkata 700104, India

Q1: What are the major topics underlying the set of reviews?
Q2: How does thematic composition of a review affect its influence?

To illustrate the utility of the proposed approach, the article examines the distribution and impact of thematic content on a set of restaurant reviews collected from Yelp. Three underlying topic clusters are discovered. The first provides a generic overview of the restaurant, the second is related to service, and the third cluster is related to specific food items. The thematic composition of reviews is found to vary based on the characteristics of restaurants, reviewers, and their evaluation of the restaurants. The discussion related to specific food items is found to be positively associated with the diagnosticity of a review, whereas generic information is found to have a negative association. The discussion related to service is found to have no effect.

The article is organized as follows. The next section presents an overview of the extant literature in the field. Following it, the proposed approach is presented and discussed. A case study, illustrating the application and utility of the approach is presented next. A discussion of the important findings in the case study and their implications are presented in the following sections. The article concludes with a discussion of the limitations and scope for future research in this area.

Background and literature review

This article draws upon the theorical ideas of diagnosticity and quality cues and makes a contribution to literature on online review influence. A brief discussion of the two theoretical lenses and important findings from previous studies has been presented below.

Diagnosticity

Diagnosticity is defined as the extent to which consumers gain information about a product from an entity, and are able to make purchase decisions (Kempf and Smith 1998). In other words, diagnosticity is a measure of the impact of an entity on the consumers' purchase decisions. Diagnosticity has been a popular construct in the information systems and marketing literature. Kempf and Smith (1998) have used the construct to investigate the effectiveness of product trials in judging brand attributes. Likewise, Jiang and Benbasat (2004) have used diagnosticity to explain the impact of virtual product experience on consumer's decision making in an online environment. Pavlou et al. (2007) and Dimoka et al. (2012) have identified higher diagnosticity to be important mechanism to mitigate uncertainty in online transactions and online commerce respectively. In the context of online reviews, diagnosticity has been referred to as "helpfulness" or "usefulness", depending on the terminology used by the e-commerce site or review platform that provided the data for the study. In this article, generic terms "review influence" or "review diagnosticity" have been used for the purpose.

Quality cues and attributes

The idea of quality cues and quality attributes provides an important conceptual lens to understand reviews. Quality cues are "informational stimuli that are, according to the consumer, related to the quality of the product, and can be ascertained by the consumer through the senses prior to consumption" (Steenkamp 1990). Consumers compare between available alternatives, and make purchase decisions based on the observable quality cues. In the context of online commerce, it may be noted that product or service characteristics are often unknown before consumption, and consumers have to rely on reviews posted by other consumers to make purchase decisions. Reviews act as a significant source of quality cues in this context. Quality attributes, on the other hand, are benefits provided by the product. Unlike quality cues, quality attributes can be ascertained only after the consumption of the product. Reviews may be viewed as consumer feedback on the quality related attributes of a product.

Previous studies on online review influence

The literature in the field of review influence has largely been guided by the idea of diagnosticity. In an early article, Mudambi and Schuff (2010) have proposed that longer reviews have higher information content and higher diagnostic value. They have found that longer reviews are more influential than shorter reviews. The finding was corroborated in subsequent studies (Baek et al. 2012; Lee and Choeh 2016). Likewise, Weathers et al. (2015) have showed that reviews of products which contained both positive and negative information about a product had a higher diagnosticity and were more influential than neutral reviews as well as reviews containing only positive or negative information.

Examining the association between review valence and its influence, Baek et al. (2012) and Chen and Lurie (2013) have shown that reviews with a lower rating score are more influential compared to reviews with a higher rating score. Likewise, it has been shown that the influence of a review increases with an increase in the proportion of negative words in the text of the review (Baek et al. 2012; Kuan et al. 2015). In addition to diagnosticity, the idea of negativity bias has been used to explain these observations.

Other factors which have been established as drivers of review influence include readability, linguistic attributes of a message, and characteristics of a reviewer. Ghose and Ipeirotis (2011) have shown that the influence of a review increased with an increase in the readability of the review. Kuan et al. (2015), however, have found a negative relationship between the two variables. Examining the impact of linguistic attributes, Ghose and Ipeirotis (2011) have shown that spelling errors in a review negatively affected its influence. Huang et al. (2013) have shown that attribute based reviews were more helpful for search goods, while experience based reviews were more helpful for experience goods. Weathers et al. (2015) have found that reviews of experience goods that contained both positive and negative information about a product were more influential than neutral reviews as well as reviews containing only positive or negative information. The effect has not been observed for search goods. Otterbacher and Arbor (2009) have shown that a review written by a reviewer with a better rank is more influential. Likewise, Ghose and Ipeirotis (2011) have observed that the influence of a review is positively associated with the average number of positive votes received by the previous reviews written by a reviewer. And Forman et al. (2008) have shown that the disclosure of the reviewer's real name and location is positively associated with the influence enjoyed by a review.

The important articles in the field have been summarized in Table 1 below.

Table 1 Extant research on online review influence

Variables	Articles	Findings	Data Source
		Textual components	
Review length	(Mudambi and Schuff 2010; Baek et al. 2012; Lee and Choeh 2016)	The length of a review is positively associated with its influence.	Amazon (all three articles)
Sentiment / emotional content	(Baek et al. 2012; Kuan et al. 2015) (Yin et al. 2014)	The influence of a review increased with an increase in the number of negative words in the review. The perceived anxiety affected review influence more than perceived anger.	Amazon (Baek et al. 2012; Kuan et al. 2015) and Yahoo! Shopping (Yin et al. 2014)
Readability	(Ghose and Ipeirotis 2011; Kuan et al. 2015)	Inconsistent findings. Ghose and Ipeirotis (2011) have shown that review influence increased with an increase in readability, whereas Kuan et al. (2015) have shown that review influence decreased with an increase in the readability of the review.	Amazon (Both articles)
Linguistic attributes	(Ghose and Ipeirotis 2011) (Huang et al. 2013) (Weathers et al. 2015)	The spelling errors in a review reduced the influence of a review. Attribute based reviews were more influential for search goods and experience based reviews were more helpful for experience goods. The reviews of experience goods which contained both positive and negative information were more helpful than reviews which were neutral or contained only positive or negative information.	Amazon (Ghose and Ipeirotis 2011; Weathers et al. 2015) and experiments ((Huang et al. 2013)
		Non-textual components	
Rating score / Rating extremity / Rating deviation	(Baek et al. 2012; Chen and Lurie 2013; Wu 2013) (Baek et al. 2012; Kuan et al. 2015) (Mudambi and Schuff 2010; Pan and Zhang 2011; Kuan et al. 2015)	Inconsistent findings. Baek et al. (2012) and Chen and Lurie (2013) have shown that review influence varied negatively with rating score. Wu (2013) have shown that the effect vanished when the quality of review was controlled. Inconsistent findings. Baek et al. (2012) have shown that review influence decreased with an increase in the rating deviation, whereas Kuan et al. (2015) have observed that review influence increased with an increase in the deviation of rating score from average rating. Review extremity increased the influence of a review and the effect was stronger for experience goods than search goods.	Yelp (Chen and Lurie 2013) and Amazon (other articles)
Reviewer credibility and information disclosure	(Otterbacher and Arbor 2009; Ghose and Ipeirotis 2011) (Forman et al. 2008)	The reviews written by a reviewer with higher rank or with better response to previous posts were more influential. The extent of reviewer information disclosure was positively associated with the influence of a review.	Amazon (All articles)
Price	(Baek et al. 2012)	The impact of review length and negative words on review influence was more for higher priced products and the impact of rating deviation and reviewer characteristics on review influence was more for lower priced products.	Amazon

Based on the above discussion, an important gap in the literature may be noted. Information content has repeatedly been highlighted as a key determinant of online review influence. However, measures used to represent information content have been weak, often based only on the count of words in the review text. While word count or review length may represent the amount of information contained in a review, it is a weak proxy, if at all, for the quality and type of information conveyed by a review.

In this study, we propose that diagnosticity depends not only on the quantity of information, but also on the type of information in a review. For instance, while purchasing a phone, a customer may seek information about various aspects of the product, including its look and feel and functional features. As such, we suggest that review information needs to be represented on the basis of the granular concepts that it represents, and not just on the count of words it contains. In this study, topic modelling has been used to summarize review content into individual topics (representing concepts) and test the impact of topics on the overall influence of a review. Topic modelling allows richer analysis than the standard text mining approach of representing information through keywords (Zhang et al. 2010), and has been discussed in detail in the following section.

Proposed approach

Both quantitative and qualitative research methods have been used to study different aspects of online social communities in general, and online reviews in particular. There are inherent major drawbacks, though, in either of these approaches. Quantitative research methods have predominantly focused on review components, which are easily measurable (e.g., rating score, length etc.). Qualitative research methods, on the other hand, have relied on content analysis to manually examine a set of reviews and identify factors affecting influence. These methods have often been criticized for being subjective, non-replicable, and non-scalable (Elo and Kyngäs 2008). Topic modeling can act as a bridge method, and address limitations of the two research approaches. A detailed discussion of using topic modeling, and its relevance for studying online review influence has been presented below.

Topic modeling

Topic modeling is an information summarization technique, and can be viewed as a semi- automated content analysis method (Wallach 2006; Debortoli et al. 2016). Similar to content analysis, topic modeling also consists of five major steps: problem identification, data collection, coding, analysis, and interpretation (Schmiedel et al. 2018). First, the research question intended to be addressed by topic modeling is finalized. Next, during the data collection phase, text data relevant for the research question is collected and stored as a corpus. A major difference between manual content analysis and topic modeling arises during the coding and analysis phases. In the former, coding categories are decided apriori and values corresponding to the pre-defined categories are filled for each text document. But in topic modeling, categories are identified using the unsupervised machine learning approach. The interpretation phase in topic modeling involves explaining the functional significance of identified categories or "topic clusters", and using the clusters for further analysis (e.g., in PCA or regression).

Topic modeling offers several advantages compared to manual content analysis, as well as traditional statistical techniques. First, traditional techniques like k-means clustering work only on numeric data. Topic modeling methods, on the other hand, allow identification of clusters from textual data, and thereby enable a researcher to inductively gain deep insights from the text. Second, by allowing an inductive examination of large text corpus, topic modeling allows constructs as well as relationship between constructs to "emerge" from data. This makes it a suitable tool for theory development through exploratory research, unlike traditional quantitative techniques which are mostly used for testing of theories. And third, topic modeling allows analysis of naturally occurring data (e.g., blogs, reviews etc.), unlike traditional quantitative approaches (e.g., surveys), where data is "artificially" generated in a controlled environment to support a study. As such, data used in topic modeling is "less biased though social expectations that can influence data that are created in research situations" (Schmiedel et al. 2018).

Topic modeling is of particular significance in the study of online review influence. It has been established by previous articles in the field that textual content of reviews plays a significant role in determining influence (Mudambi and Schuff 2010; Baek et al. 2012). While review length, as a measure of quantity of information provides an information cue, a detailed analysis of the association between thematic elements in a review and review diagnosticity is also needed. Furthermore, as the information content of reviews varies for different products and product types, and apriori theoretical identification of specific themes affecting review influence is not feasible, a generic method allowing automatic discovery of information from text is needed. Finally, since the volume of reviews is generally large, a scalable computing method is preferable over manual analysis. Topic modelling allows richer analysis than the standard text mining approach of representing information through keywords (Zhang et al. 2010). This is because traditional text mining relies directly on the analysis of word count and does not differentiate between the use of individual words in different contexts. In a topic modelling algorithm like LDA the meaning of words is inferred on the basis of their co-occurrence with other words.

This helps to capture the contextual meaning of words (DiMaggio et al. 2013). A process-based approach of topic modeling, as discussed above and used in this article, has been presented in Figure 1 below and discussed later.

Data collection

First, relevant data needs to be collected and stored. In addition to review text, this includes contextual data, particularly the details of products or businesses for which reviews have been written and reviewers who wrote the reviews.

Data preparation / cleaning

In text analysis, it is important to appropriately clean the dataset and prepare it in a way so that text-based algorithms can be applied on it. As the first step in data cleaning, whitespaces in all strings constituting the text corpus are eliminated. Whitespaces or blank spaces often alter individual strings by getting attached either before or after the strings. Elimination of whitespaces ensures that same strings are not interpreted as different by the topic modelling algorithm. Next, punctuations and digits are removed from the corpus. Punctuations add clarity to the text by separating individual elements within the sentences. They do not add any new meaning or modify the meaning of existing textual content. Therefore, for the purpose of examining the topical content in a text, punctuations have little significance and are eliminated from the textual corpus. Likewise, digits and numbers, while adding specificity to the meaning being conveyed by the text (e.g., used for specifying exact price instead of mentioning a "high price" item), do not add additional meaning. Therefore, in the context of topic modelling, digits and numbers are also redundant and are eliminated from the text corpus.

All words in the document are converted to a common case (lower / upper). This ensures that the same words written in a different case do not get interpreted differently. A text corpus invariably consists of words which occur very commonly in any document. For instance, words "the" and "can" occur in most textual documents. These commonly occurring words, referred to as "stopwords" are eliminated from the text corpus. Elimination of stopwords is important and useful, as it allows relevant keywords to emerge. Also, it makes the processing of text faster. Following this, all words remaining in the text corpus are stemmed. In Natural Language Processing stemming refers to the idea of retaining only the word stem for each word in the text corpus. For instance, words "eating" and "ate" after stemming are reduced to the same word stem "eat". Stemming ensures that words with similar connotations are not interpreted differently.

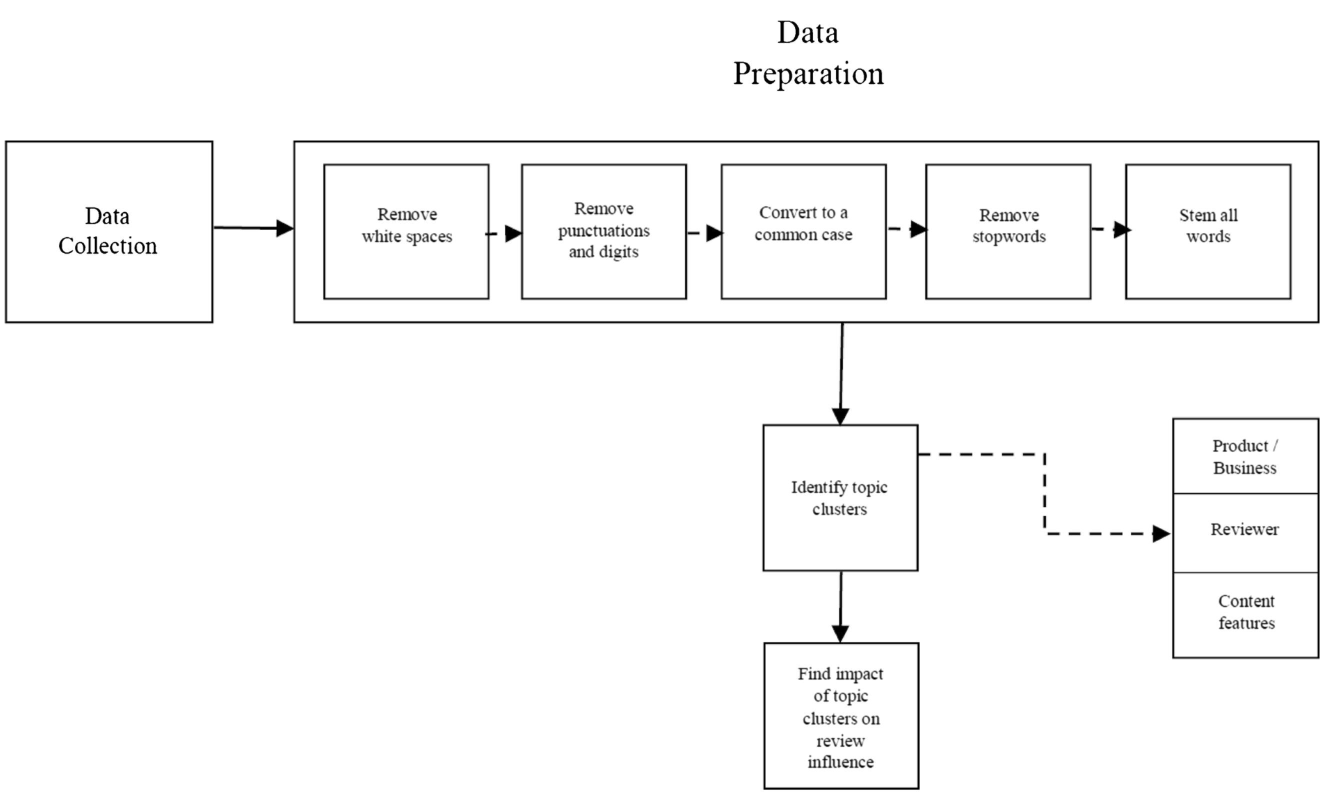

Fig. 1 An approach for exploring topical composition of reviews

An example to illustrate the process has been presented below.

"Original review: "Very satisfied! Great food! Quiet! Perfect! New to Fountain Hills. This is the place we were looking for".

After removing whitespaces: "Very satisfied! Great food! Quiet! Perfect! New to Fountain Hills. This is the place we were looking for".

After removing punctuations and digits: "Very satisfied Great food Quiet Perfect New Fountain Hills This is the place we were looking for".

After converting to lower case: "very satisfied great food quiet perfect new fountain hills this is the place we were looking for".

After removing stopwords: "very satisfied great food quiet perfect new fountain hills this place looking"".

Finding topical information

To identify topic clusters, Latent Dirichlet Allocation (LDA) is used (Blei et al. 2003). LDA is a popular topic modeling algorithm, and has frequently been used in studies across disciplines to examine topics underlying a natural language textual corpus (Shi et al. 2016; Dyer et al. 2017; Guo et al. 2017; Lugmayr and Grueblbauer 2017). Specifically, in the execution of LDA, a document is considered to be composed of independent topics. The number of topic clusters to be identified in the text needs to be specified by the user. The relative composition of each topic cluster in the document is calculated by the algorithm and is called "Gamma". For example, with the number of topic clusters specified as 2, the LDA may compute gamma of topic cluster 1 as 0.60 and topic cluster 2 as 0.40. This indicates that 60% of the document is composed of topics related to topic cluster 1, while 40% of the document is composed of topics related to topic cluster 2.

Two important measures calculated by the LDA help in interpretation of the identified clusters. For each term in the text corpus, the probability of the term appearing in the discussion related to the underlying topic clusters is identified. This value is called 'Beta'. The terms with the highest value of beta corresponding to a topic cluster appear more frequently in the discussion pertaining to the topic cluster, compared to other terms. Second, for every term, the relative probability of appearing in one topic cluster compared to all other topic clusters is calculated, and is called 'Beta Spread'. Mathematically, for a term T1, 'Beta Spread' between two topic clusters, say C1 and C2, is calculated as follows.

$$\text{Beta Spread (T1)}_{C1,\,C2} = \log(\text{Beta (T1)}_{C1}) / \log(\text{Beta (T1)}_{C2}).$$

The terms with a higher value of beta spread between a pair of topic clusters are more likely to appear in the first topic cluster, while terms with a lower value of beta spread are more likely to appear in the second topic cluster. Analyzing reviews containing terms with the highest or lowest values of these measures can help in interpreting the identified clusters (Robinson and Silge 2017).

Comparison of topical content across sub-samples

Based on a review of literature, three categories of variables have been identified to sub-sample the review data: the first is related to product / business, the second is related to reviewers and the third is related to characteristics of review content. Based on the identified topic clusters and the relative composition of the clusters, the topical composition of reviews in individual sub-samples are compared.

Influence of topical content

Finally, a regression model is used to find if and how the identified topic clusters associate with the influence that is generated by a review. This step can determine the type of information related to a product or business that help readers make purchase decisions.

Case study

To illustrate the application of the proposed approach, we identify and explain the distribution of thematic content in an online review dataset of restaurants collected from Yelp. The dataset is made available by Yelp, as part of the Yelp dataset challenge 2018 (Yelp 2018). Review data collected from Yelp is increasingly being used in academic research pertaining to online reviews (Luca and Zervas 2016; Banerjee et al. 2017; Vallurupalli and Bose 2017) and is found suitable for this study. A detailed step-wise discussion of the application of this approach is presented below.

Data collection

From the available dataset, reviews of a randomly selected sample of 200 restaurants from the state of Arizona are used in this study. Overall, the dataset consisted of 13,456 reviews.

Data preparation

Following the proposed approach, the whitespaces are eliminated from the text. Following this, all digits and punctuations are removed. All strings are then converted into lower case. Next, the stopwords are eliminated, and finally, individual strings in the text corpus are stemmed. Each of these steps is executed using the in-built functions provided by the "tm" package in R.

Model building and insights

The first part of the approach relates to identification and interpretation of topic clusters in the review text corpus. This is done as discussed below.

Identification of underlying topic clusters

Latent Dirichlet Allocation (LDA) is used to identify topic clusters underlying the reviews. LDA is used thrice, specifying the number of output clusters as 2, 3, and 4 respectively. For the interpretation of the identified topic clusters, and to identify the differences between the clusters, beta and beta spread of terms in the corpus are calculated, as discussed in the previous section. With the number of output clusters specified as 3, the topic clusters are found to be most interpretable.[1] In the interest of parsimony, the rest of this section focuses on the output of LDA with 3 topic clusters. Following the proposed approach, two measures, 'Beta' and 'Beta spread' are used for interpretation of the topic clusters. For the output of LDA with 3 topic clusters, the terms with the highest positive and negative beta values are shown in Table 2. The terms with the highest and the lowest value of beta spread between topic clusters are shown in Table 3.

A detailed content analysis of reviews containing the words in Tables 2 and 3 is carried out to understand the contextual meaning of terms and determine the business significance of the identified clusters. It is found that topic cluster 1 primarily provides generic information about the restaurant, topic cluster 2 relates to service-related information, while topic cluster 3 relates to information on specific food items. This may also be observed from Appendix, which shows some of the reviews containing the above terms.

Next, the analysis focuses on the topical composition of the overall set of reviews and individual sub-samples of reviews. This is discussed in detail in the following paragraphs.

Topical distribution across different individual sub-samples

The textual components of reviews commonly explored in previous studies include length, readability, emotional content, and linguistic attributes of review text. Likewise, rating score, reviewer credibility, reviewer information disclosure, and price constitute the non-textual components. The linguistic attributes of the review text and the extent of information disclosure by reviewers are obtained through a manual coding of the data. An important objective of the approach is to discover the quality attributes in reviews, and understand how they vary across samples. For this purpose, the following questions are addressed: How does topical distribution vary between (1) reviews of different length, (2) reviews of different readability, (3) reviews with different levels of affect or emotional content, (4) reviews with different rating scores, (5) reviews of businesses with different price range, and (6) reviews written by reviewers with different levels of popularity?

The dataset used in this study has identified the *price range* of all restaurants as falling in one of the three categories: 1, 2, or 3, where the higher digit represented higher priced restaurants. Overall, the dataset has 3503 reviews of restaurants with price range 1, 9554 reviews of restaurants with price range 2, and 399 reviews of restaurants with price range 3 respectively. The restaurants with *price range* 1 are identified as 'Low price' restaurants, whereas those with *price range* 2 or 3 are identified as 'High price' restaurants.

The length of a review is measured by the number of words in the review. Review readability is measured by the average number of words per sentence in the review text. With more words, a sentence is believed to become more complex and less readable. This is a well-known metric to measure readability (Pennebaker et al. 2007). Emotional content is measured by the total number of positive and negative words in the review (Ghose and Ipeirotis 2011; Goes et al. 2014). All three attributes are computed using Linguistic Inquiry and Word Count (LIWC), a popular text analysis tool (Pennebaker et al. 2007). The reviews with a star rating score higher than the median star rating / length / emotional content are marked as 'High rating' / 'High length' / 'High affect', whereas those with a rating score less than or equal to the median star rating are marked as 'Low rating' / 'Low length' / 'Low affect' reviews. The reviews where the average words per sentence is less than or equal to the median words per sentence for all reviews are marked as 'More readable'. In the same vein, the reviews where average words per sentence is more than the median words per sentence are marked as 'Less readable'. The number of fans of a reviewer is used as a proxy for the popularity of the reviewer. The reviews written by reviewers with number of fans more than the median number of fans are marked as 'Reviews by more popular reviewers', whereas those written by reviewers with number of fans less than the median number of fans are marked as 'Reviews by less popular reviewers'.

For each of the features, the mean and the standard deviation of gamma values for the identified topic clusters is calculated for reviews in the 'High' and 'Low' categories. A t test is used to determine if the difference in gamma values of topic clusters between the categories is significant. The results are shown in Table 4 and summarized in Table 5.

[1] In the output of the LDA with 2 clusters, one of the topic clusters represented food related topics, while the other cluster was found to be a combination of service and ambience related topics. In the output of the LDA with 4 clusters, the first topic cluster represented food and the second cluster represented ambience. The third and fourth topic clusters represented service-related topics. As such, the output of the LDA with 3 clusters, where the three clusters represented food, service, and ambience respectively, was found to be most interpretable and ideal for further analysis.

Table 2 Terms with the highest beta values for the three topic clusters

Topic cluster 1		Topic cluster 2		Topic cluster 3	
Term	Beta value	Term	Beta value	Term	Beta value
place	0.0462	order	0.0253	chicken	0.0162
food	0.0311	time	0.0191	delicious	0.0101
friend	0.0186	just	0.0154	taste	0.0094
service	0.0139	wait	0.0100	salad	0.0087
price	0.0115	table	0.0093	dish	0.0086

Next, the impact of topic clusters on review influence is examined. The descriptive statistics related to the variables used for this analysis are presented in Table 6.

First, a model without topic clusters is developed to serve as baseline, as shown below.

Review influence (Total votes) = ß1*Rating deviation + ß2*Review length + ß3*Reviewer popularity + ß4*Negative affect + ß5*Review age + ß6*Review count + Error.

The variables that are known to affect review influence from extant literature are used as independent variables in the above model. Following extant literature (Kuan et al. 2015; Wan 2015), time elapsed (in days) since the review is written and the number of reviews written for a product are

Table 3 Terms with the highest beta spread between different topic clusters

Topic cluster 1 versus topic cluster 2			
Highest		Lowest	
Terms	Beta spread	Terms	Beta spread
always	13.7	order	−14.7
price	13.7	just	−13.9
best	13.4	wait	−13.3
staff	13.4	table	−13.2
happy	13.2	server	−13.0
Topic cluster 2 versus topic cluster 3			
Terms	Beta spread	Terms	Beta spread
time	14.4	menu	−13.5
drink	13.5	hot	−12.3
wait	13.4	cheese	−12.0
people	12.9	love	−11.9
minute	12.8	home	−11.3
Topic cluster 1 versus topic cluster 3			
Highest		Lowest	
Terms	Beta spread	Terms	Beta spread
place	15.6	just	−12.9
friend	14.3	meal	−12.7
price	13.6	made	−11.8
staff	13.3	offer	−10.9
hour	13.2	order	−10.6

added as control variables. Review readability, as measured by the average number of words per sentence in the review is found to be correlated with review length and is dropped from the model.

In subsequent models, each of the topic clusters is separately entered and the results are compared with the baseline results. The topic cluster values are moderately correlated with each other. To avoid multicollinearity, the topic clusters are not used together as independent variables in any of the models. The results of the regression are summarized in Table 7.

The results show that generic information is negatively associated with review influence, while discussion related to food items is positively associated with the influence of a review. The discussion related to service is found to have no significant impact on influence.

Discussion and analysis

The application of the proposed approach has allowed the categorization of thematic content in reviews into three topic clusters. Since online reviews reflect customer feedback of a product or service, the identified topic clusters can be thought of as dimensions along which the customers evaluate restaurants. The identification of topic clusters, and the quantification of the extent of discussion for each of the clusters, has allowed contextual analysis to determine if and how the expectations of customers vary across restaurants. Also, it has helped understand how the content created varies with the characteristics of reviewers and their specific evaluation or feedback of the restaurant. Finally, the use of regression has helped determine how the topic clusters affect the helpfulness of a review, thereby illustrating a relationship between the nature of information in a review and its influence on customers.

Several important inferences can be drawn from the findings. It is observed that the discussion related to food had a positive impact while generic discussion had a negative impact on the influence of a review. The discussion related to service had no observable impact. This indicates that the decision to visit a restaurant is more often informed by information provided in reviews about food than service issues or generic information about the restaurant. From the perspective of quality cues and information diagnosticity, it may be inferred that cues related to food make a review more diagnostic, compared to cues related to service or general characteristics of a restaurant. The finding shows how the use of topic modelling algorithms can illustrate the application of theoretical ideas related to textual content.

Likewise, important findings regarding the quality attributes valued by consumers also emerge from the analysis. Sub-sample analysis that is based on characteristics of business has indicated that in reviews of high priced restaurants, there is a higher composition of service and food related issues, whereas in reviews of low priced restaurants, there is a higher composition of generic information. This hints that customers have specific and high expectations regarding both food and service quality attributes in high priced restaurants compared to low priced restaurants. It is also observed that reviews with a higher rating consist more of topics related to food and less of topics related to service compared to reviews with lower ratings. This indicates that compared to food, poor service is more likely to lead to a lower rating score. It indicates a dissatisfaction towards service-related quality attributes is more likely to lead to a negative evaluation of a restaurant.

The topics discovered are consistent with the theory of goods and services continuum. The theory suggests that output from all businesses have characteristics of both goods and services, the former being different from the latter based on typical characteristics of tangibility, perishability, separability, and standardization (Lovelock 1983; Zeithaml et al. 1985; Evans and Berman 2002; Bearde et al. 2007; Gale 2007). Since restaurants, which form the focus of this study, are expected to fall in the middle of the continuum, topics related to both aspects should be discovered from the reviews. In line with this, in our case, separate topic clusters for "goods" and "services" aspects of the business were discovered, which provide a qualitative validation of the approach.

It may be noted that quality cues in reviews, used by consumers to make decisions, as well as quality attributes which reflect consumers' expectation from a product or service are specific to a product or service and cannot be generalized. Furthermore, consumers' needs and tastes may evolve over time, and hence, the cues and attributes will likely change over time. In such a scenario, an unsupervised knowledge discovery approach is needed for understanding customer preferences and decision-making process, and topic modeling provides the most suitable alternative. The use of topic modeling in this study led to discovery of cues and attributes valued by consumers, and thereby illustrated this point.

The findings from case analysis, as discussed above, make a contribution to review influence literature in two major ways. First, it extends the idea of diagnosticity in the study of online review influence by illustrating the important role of embedded topics in influence. And second, it illustrates the

Table 4 Mean and standard deviation of gamma values of topic clusters for different categories of reviews

High length versus low length	Topic 1 (Generic) Gamma (Mean)	Topic 2 (Service) Gamma (Mean)	Topic 3 (Food) Gamma (Mean)
High length	0.317 (0.079)	0.348 (0.096)	0.335 (0.097)
Low length	0.354 (0.048)	0.319 (0.046)	0.327 (0.049)
	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01
High readability versus low readability			
High readability	0.347 (0.569)	0.321 (0.058)	0.332 (0.063)
Low readability	0.324 (0.076)	0.347 (0.090)	0.329 (0.089)
	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01
High rating versus low rating (review)			
High rating	0.353 (0.059)	0.314 (0.054)	0.332 (0.064)
Low rating	0.318 (0.071)	0.412 (0.090)	0.329 (0.088)
	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01
High affect versus low affect (review)			
More positive	0.352 (0.062)	0.305 (0.050)	0.343 (0.070)
Less positive	0.325 (0.069)	0.351 (0.084)	0.324 (0.080)
	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01
High price versus low price (business)			
High price	0.334 (0.072)	0.335 (0.079)	0.331 (0.081)
Low price	0.337 (0.059)	0.332 (0.071)	0.330 (0.069)
	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01	Dif. sig. at p < =0.01
More popular versus less popular (reviewer)			
More Popular	0.340 (0.0667)	0.329 (0.073)	0.330 (0.074)
Less Popular	0.337 (0.0650)	0.339 (0.080)	0.323 (0.072)
	Dif. not sig.	Dif. sig. at p < =0.01	Dif. not sig.

Table 5 Summary of sub-sampling results

Sub-sampling variable	Findings
Length	Longer reviews have a higher composition of service and food related topics, and a lower composition of generic information compared to shorter reviews.
Readability	More readable reviews have a higher composition of generic information and food related topics and a lower composition of service-related topics compared to less readable reviews.
Rating score	Reviews with high rating have a higher composition of generic information and food related topics, and a lower composition of service-related topics compared to reviews with low rating.
Affect	Positive reviews have a higher composition of generic information and food related topics, and a lower composition of service-related topics compared to less positive reviews.
Price range	Reviews of high-priced restaurants have a higher composition of service and food related topics, and a lower composition of generic information compared to reviews of low-priced restaurants.
Reviewer popularity	Reviews written by more popular reviewers have a lower composition of service-related topics compared to reviews written by less popular reviewers. The composition of other topics written by both category of reviewers is similar.

role of topic modeling in the study of online reviews in general and review influence in particular. While the specific findings are relevant only for restaurants, the approach can be extended to online reviews for any product or service. Furthermore, since online review data is often available in the public domain, it may be possible to compare the findings with other players in the industry. A better understanding of the nature of information contained in the reviews and its impact on users can help managers understand and serve the needs of the customers in a more effective way.

Implications

The article makes important contributions to the literature on information systems. First, the article contributes to the literature in the field of online review creation by offering a novel way to analyze thematic content in a review dataset. The proposed approach in this article enables a more comprehensive analysis of thematic content by identifying individual topics embedded in reviews as well as examination of the differences in topical composition between different review sub samples. It extends the literature on online reviews that has focused on mining product features and / or determination of customer satisfaction from online reviews (Ma et al. 2013; Xianghua et al. 2013; Guo et al. 2017). Second, the article adds to a growing body of literature in the field of online review influence by illustrating the importance of analyzing thematic content in determining the usefulness of a review. While several articles have investigated the determinants of online review influence (as shown in Table 1), the role of thematic content in impacting review influence has so far remained unexplored. This gap is bridged by this article. And third, the proposed approach extends the knowledge about text mining and natural language processing techniques in understanding information systems related phenomenon, particularly online reviews (Buettner 2017; Lugmayr and Grueblbauer 2017). The need for the same has been noted in extant research (Debortoli et al. 2016), and the article contributes by illustrating the application of topic modelling to better understand the impact of online reviews on consumers.

Table 6 Descriptive statistics of variables used in the analysis

Variables	Min	Max	Mean	Std. dev
Total votes	0	75	1.848	3.794
Topic cluster 1 (Generic)	0.110	0.631	0.335	0.068
Topic cluster 2 (Service)	0.081	0.739	0.334	0.077
Topic cluster 3 (Food)	0.078	0.809	0.331	0.077
Rating deviation	0.000	3.633	0.996	0.719
Review length	1	986	110.181	107.912
Review readability	1.000	198.670	14.483	9.185
Number of fans	0.000	594.000	4.725	17.797
Negative affect	0.000	42.860	1.128	1.936
Review age (in days)	0	3673	1336.913	925.879
Review count	3	450	176.173	142.686

Likewise, the article has important implications for practitioners. The proposed approach can be used by retailers, manufacturers, and service-based businesses to monitor the feedback given by customers in greater detail, and with little human intervention. This can be done by automatically identifying the key themes in reviews using topic modelling, and periodically reporting them to the business end users. Second, the approach can assist in market research efforts by identifying expectations of customers across different business types. As topic modelling inductively identifies key themes from data naturally and

Table 7 Results obtained from the regression analysis

Variables	Model 1 (t value) Baseline	Model 2 (t value) Baseline + Topic 1 Cluster	Model 3 (t value) Baseline + Topic 2 Cluster	Model 4 (t value) Baseline + Topic 3 Cluster
(Constant)	-3.173^{**}	1.300	-2.619^{***}	-4.754^{***}
Rating deviation	4.018^{***}	3.518^{***}	3.946^{***}	4.564^{***}
Review length	17.524^{***}	15.244^{***}	16.975^{***}	17.373^{***}
Reviewer popularity	25.540^{***}	25.546^{***}	25.407^{***}	25.155^{***}
Negative affect	2.026^{*}	1.548	2.048^{*}	2.345^{**}
Review age	-8.211^{***}	-8.614^{***}	-8.206^{***}	-8.495^{***}
Review count	1.126	1.330	1.087	0.942
Topic cluster 1 (Generic)	NA	-3.999^{***}	NA	NA
Topic cluster 2 (Service)	NA	NA	−0.321	NA
Topic cluster 3 (Food)	NA	NA	NA	3.645^{***}

$*p < 001$; $**p < 0005$; $***p < 0001$. (Sample size: 13,456).

wilfully generated by consumers, it may be more effective than standard market research efforts using pre-defined questions and scales. And third, by identifying topics which make reviews useful to a reader, e-commerce and review platforms can make specific recommendations to reviewers to create more useful content. They can also effectively identify reviews which need to be shown to the readers, which is particularly important if the number of reviews for a product or business is large.

Conclusion

With the increasing popularity of analytics, there has been a shift in traditional approaches of marketing in favour of data-driven techniques. As such, there is great interest in the academic community to develop novel analytics driven approaches to support marketing efforts across different industries and contexts. This article has contributed to this field by proposing and illustrating the use of a novel approach to examine the composition of online reviews and its impact of influence.

As in other articles, a few limitations may be observed, which point towards scope for future work in the area. First, the generalizability of findings obtained from topic modeling is limited to the data source being considered. As such, findings obtained from topic modeling in one context cannot be applied to other contexts. The approach has been illustrated using data from a single platform. In future studies, the utility of the approach across multiple platforms can be established. In particular, while the current article has focused on service-based reviews, the utility of the proposed approach can be examined and validated for product reviews in future. Second, the identification of constructs and establishing relationship between constructs needs subjective interpretations on the part of researchers. As such, like manual content analysis, topic modeling is also not fully automatic and replicable. The approach relies on the determination of a suitable theme for identification of clusters by the authors.It is not clear if a large group of experts interpreting the same content will reach same or similar conclusions. In future studies, the consistency of findings for different users (particularly industry experts) can be examined. Third, we have considered several review characteristics and studied their impact of the discovery of topics from reviews. However, in order to keep the models parsimonious we have considered only select characteristics that have been shown to be influential in extant research. In future, a more comprehensive list of review and reviewer characteristics may be considered. This can include linguistic characteristics like spelling errors as well as the extent of reviewer information disclosure. Fourth, we have observed that topics representing 'goods' and 'service' result in different impact on review influence. However, further research is needed to examine why this difference is observed and what theoretical explanation can be provided for this result. Future researchers are encouraged to explore advanced theoretical lenses as well semi-structured interviews of the readers of the reviews to provide a deeper understand of this finding. It is hoped that the proposed approach will lead to an increased interest in investigating the thematic content of reviews, and future studies will refine and enhance the approach making it more accurate, useful, and robust.

Appendix

Some sample reviews containing words with highest value of beta for different topics and highest value of beta spread between different pairs of topics have been presented below. It may be noted that specific terms and not entire reviews were assigned to different topic clusters. We have listed reviews in the appendix only to highlight the reviews containing considerable proportion of terms in the three identified topic clusters.

Topic 1 – Generic

*"Very satisfied! Great **food**! Quiet! Perfect! New to Fountain Hills. This is the **place** we were looking for".*

*"Atmosphere is cool and laid back, nice drink menu, friendly staff! Just moved to the neighborhood and really like the **place** ...".*

*"Amazing authentic **food**. I highly recommend the buffet. Friendly **staff**. Definitely coming back."*

*"**Staff** is great, **food** is great, beer selection is super great. They have PBR in a can every day all day for $1.25."*

*"Im in town for business for a few days. This **place** is close to my hotel so I popped in to try it. \n\nFantastic gyros, friendly **service** and reasonably **priced**. \n\nDont forget to pick up some baklava!*

Topic 2 – Service

"*... This **time** the service was even more slow. 15 minutes passed before we even got our water. 45 minutes later our food still hadnt arrived. The waitress kept speeding by our **table** and ignoring us ...*".

"*... At this point I look around the restaurant and I see some lady customer get up from her **table**, walks over to a waitress and screams at her saying shes walking out because her family too was still **waiting** for their food after over an hour. Within 15 minutes like a trickle effect I watch 3 more **tables** walk out havent even gotten their food yet either ...*".

"*... Aside from the layout problems with the lines, the line did move fairly quickly and we were able to place our **orders** in a fair amount of **time**. After we placed our orders we sat down at one of the **tables** and **waited**. While **waiting** we noticed more layout issues. The spot where the **orders** come up is on the opposite side of the store from the **tables**, so if you are sitting down its difficult to hear if they are calling your number ...*".

"*... We **waited** 10 min before anyone noticed us. A waiter came over and took drink **orders**, which we received quickly, but no sugar or silverware. We were ready to **order** but could not get the waiters attention ...*".

"*... At the bar we **waited** an unreasonable amont of **time** to be served. The bartender completely ignored us and only went to people who yelled for him or frantically waived their arms signaling him over.\n\nAfter a long **wait** I **ordered** a long Island Iced Tea ...*"

Topic 3 – Food

"*... My husband had the Chefs Taco **Tasting** Platter. It was five different tacos with two salsas. I didnt **taste** the tacos but my husband said he couldnt **taste** what taco was **chicken** from the beef and pork and liked the vegetable tacos best. (that is very unusual for him) He also said they had too much raw cabbage in them for his **taste** ...*".

"*... Our entrees were all excellent. Salmon, Short ribs and steak. The Salmon had this awesome crispy skin, **delicious** blue cheese potatoes. Steak was excellent. Short ribs had a cumin citrus sauce that was so good ...*".

"*... For my meal I had the National **Dish** of Malaysia the Ri Nasi Lemak. It was a pile of rice with fried **chicken**, hard boiled egg, cucumbers, and spicy chili anchovy sauce. The **chicken** was **tasty** and seemed infused with mild flavors ...*".

"*... This Puerto Rican inspired eatery, which is excellent for small groups and sharing, also has some heartier **dishes** on its menu such as the Pork Pernil or Ropa Vieja, which are a 12 hour slow cooked pork and beef pot roast ...*".

> "*... Basically it was a grilled **chicken salad**. The **salad** was made with crisp fresh greens, canned mandarins, LOTs of Carrots and Lots of crisp tortilla strips. The **chicken** on it was very **tasty** but slightly dry. I thought the cashews were very good they seemed real sweet but had a spicy kick at the end and the ginger dressing was good too. Overall it was a nice **salad** that I would order again but would ask for them to go lighter on the tortilla strips ...*
> "

References

Anderson, M. (2014). 88% of consumers trust online reviews as much as personal recommendations. Retrieved July 6, 2018, from https://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce, 17*, 99–126. https://doi.org/10.2753/JEC1086-4415170204.

Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems, 96*, 17–26. https://doi.org/10.1016/j.dss.2017.01.006.

Bearde, W. O., Ingram, T. N., & Raymond, L. (2007). *Marketing*. McGraw-Hill/Irwin: Principles and Perspectives.

Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993.

Buettner, R. (2017). Predicting user behavior in electronic markets based on personality-mining in large online social networks. *Electronic Markets, 27*(3), 247–265. https://doi.org/10.1007/s12525-016-0228-z .

Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research, 50*(4), 463–476. https://doi.org/10.1509/jmr.12.0063.

Corley, J. K., Jourdan, Z., & Ingram, W. R. (2013). Internet marketing: A content analysis of the research. *Electronic Markets, 23*(3), 177–204. https://doi.org/10.1007/s12525-012-0118-y.

Debortoli, S., Müller, O., Junglas, I., & vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems, 39*(1), 7–35. https://doi.org/10.17705/1CAIS.03907.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics., 41*(6), 520–906. https://doi.org/10.1016/j.poetic.2013.08.004.

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On product uncertainty in online markets: Theory and evidence. *MIS Quarterly, 36*(2), 395–426. https://doi.org/10.2307/41703461.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K texual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics, 64*(2–3), 221–245. https://doi.org/10.1016/j.jacceco.2017.07.002.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x.

Evans, J. R., & Berman, B. (2002). Marketing: Marketing in the 21st century. Atomic Dog Pub Inc.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research, 19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193.

Gale, T. (2007). Goods and services. Retrieved from https://www.encyclopedia.com/finance/finance-and-accounting-magazines/goods-and-services

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering, 23*(10), 1498–1512. https://doi.org/10.1109/TKDE.2010.188.

Goes, P. B., Lin, M., & Yeung, C. m. A. (2014). "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research, 25*(2), 222–238. https://doi.org/10.1287/isre.2013.0512.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management, 59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009.

Huang, L., Tan, C.-H., Ke, W., & Wei, K.-K. (2013). Comprehension and assessment of product reviews: A review-product congruity proposition. *Journal of Management Information Systems, 30*(3), 311–343. https://doi.org/10.2753/MIS0742-1222300311.

Jiang, Z., & Benbasat, I. (2004). Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. *Journal of Management Information Systems, 21*(3), 111–147. https://doi.org/10.1080/07421222.2004.11045817.

Kempf, D. S., & Smith, R. E. (1998). Consumer processing of product trial and the influence of prior advertising: A structural modeling approach. *Journal of Marketing Research, 35*(3), 325–338. https://doi.org/10.2307/3152031.

Kuan, K. K. Y., Hui, K.-L., Prasarnphanich, P., & Lai, H.-Y. (2015). What makes a review voted? An empirical investigation of review voting in online review systems. *Journal of the Association for Information Systems, 16*(1), 48–71. https://doi.org/10.17705/1jais.00386.

Lee, S., & Choeh, J. Y. (2016). The determinants of helpfulness of online reviews. *Behaviour and Information Technology, 35*(10), 853–863. https://doi.org/10.1080/0144929X.2016.1173099.

Lovelock, C. H. (1983). Classifying services to gain strategic marketing insights. *Journal of Marketing, 47*(3), 9–20. https://doi.org/10.2307/1251193.

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science, 62*(12), 3412–3427. https://doi.org/10.1287/mnsc.2015.2304.

Lugmayr, A., & Grueblbauer, J. (2017). Review of information systems research for media industry–recent advances, challenges, and introduction of information systems research in the media industry. *Electronic Markets, 27*(1), 33–47. https://doi.org/10.1007/s12525-016-0239-9.

Ma, B., Zhang, D., Yan, Z., & Kim, T. (2013). An LDA and synonym lexicon based approach to product feature extraction from online consumer product reviews. *Journal of Electronic Commerce Research, 14*(4), 304–314.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185–200.

Otterbacher, J., & Arbor, A. (2009). "Helpfulness" in online communities: A measure of message quality. Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI '09, 955–964. https://doi.org/10.1145/1518701.1518848

Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing, 87*, 598–612. https://doi.org/10.1016/j.jretai.2011.05.002.

Pavlou, Liang, & Xue. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly, 31*(1), 105–136. https://doi.org/10.2307/25148783.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). LIWC2007: Linguistic inquiry and word count.

Podium (2018). *2017 state of online reviews*. Retrieved from http://learn.podium.com/rs/841-BRM-380/images/2017-SOOR-Infographic.jpg

Robinson, D., & Silge, J. (2017). Text mining with R. O'Reilly Media.

Schmiedel, T., Müller, O., & vom Brocke, J. (2018). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*. https://doi.org/10.1177/1094428118773858.

Shi, Z. M., Lee, G. M., & Whinston, A. B. (2016). Toward a better measure of business proximity: Topic modeling for industry intelligence. *MIS Quarterly, 40*(4), 1035–1056. 10.1145. https://doi.org/10.25300/MISQ/2016/40.4.11.

Steenkamp, J. B. E. M. (1990). Conceptual model of the quality perception process. *Journal of Business Research, 21*(4), 309–333. https://doi.org/10.1016/0148-2963(90)90019-A .

Vallurupalli, V., & Bose, I. (2017). Temporal changes in the impact of drivers of online review influence. In *Proceedings of the 28th Australasian Conference on Information Systems*.

Wallach, H. (2006). Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning, ACM* (pp. 977–984).

Wan, Y. (2015). The Matthew effect in social commerce. *Electronic Markets, 25*(4), 313–324. https://doi.org/10.1007/s12525-015-0186-x .

Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems, 79*, 12–23. https://doi.org/10.1016/j.dss.2015.07.009 .

Wu, P. F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology and Marketing, 30*, 971–984. https://doi.org/10.1002/mar.20660 .

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems, 37*, 186–195. https://doi.org/10.1016/j.knosys.2012.08.003 .

Yelp (2018). Yelp Dataset Challenge. Retrieved May 1, 2018, from https://www.yelp.com/dataset/challenge

Yin, D., Bond, S., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly, 38*(2), 539–560.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1985). Problems and strategies in services marketing. *Journal of Marketing, 49*(2), 33–46. https://doi.org/10.2307/1251563 .

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics, 1*(1–4), 43–52. https://doi.org/10.1007/s13042-010-0001-0 .

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.