Online sentiment analysis in marketing research: a review

Meena Rambocas and Barney G. Pacheco
Department of Management Studies, The University of the West Indies, St. Augustine, Trinidad and Tobago

Received 9 May 2017
Revised 19 October 2017
Accepted 14 December 2017

Abstract

Purpose – The explosion of internet-generated content, coupled with methodologies such as sentiment analysis, present exciting opportunities for marketers to generate market intelligence on consumer attitudes and brand opinions. The purpose of this paper is to review the marketing literature on online sentiment analysis and examines the application of sentiment analysis from three main perspectives: the unit of analysis, sampling design and methods used in sentiment detection and statistical analysis.

Design/methodology/approach – The paper reviews the prior literature on the application of online sentiment analysis published in marketing journals over the period 2008-2016.

Findings – The findings highlight the uniqueness of online sentiment analysis in action-oriented marketing research and examine the technical, practical and ethical challenges faced by researchers.

Practical implications – The paper discusses the application of sentiment analysis in marketing research and offers recommendations to address the challenges researchers confront in using this technique.

Originality/value – This study provides academics and practitioners with a comprehensive review of the application of online sentiment analysis within the marketing discipline. The paper focuses attention on the limitations surrounding the utilization of this technique and provides suggestions for mitigating these challenges.

Keywords Online marketing, Qualitative research, Quantitative research, Methodology, Text mining, High technology marketing

Paper type Literature review

Introduction

In recent years, consumers' willingness to share consumption experiences online, coupled with the technology to analyze "big data", offer marketing managers an unprecedented opportunity to collect market intelligence (Erevelles *et al.*, 2016). Through online sentiment analysis, hereafter referred to as sentiment analysis, researchers can systematically extract and classify consumer emotions about products and services expressed in social network discussions and online postings to track brand attitudes and emerging market trends. While sentiment analysis presents tremendous opportunities to interpret a large body of data collected in a naturalistic setting, concerns have been expressed about the technique's accuracy and practicality (Gonçalves *et al.*, 2013). Moreover, apprehensions over online data volume, fragmented data sources, content bias and user exploitation have exposed the technique to critical scrutiny.

In light of these challenges, it is somewhat surprising that researchers have not devoted more attention to evaluating the overall feasibility of using sentiment analysis as a tool for online marketing research. Our study serves to fill this gap by reviewing the literature on the application of sentiment analysis in the marketing discipline. The review is specific to the literature published in scholarly peer-reviewed marketing journals between 2008 and 2016, which coincides with the technique's general usage within the marketing discipline.

Journal of Research in Interactive Marketing
Vol. 12 No. 2, 2018
pp. 146-163

2040-7122
DOI 10.1108/JRIM-05-2017-0030

The current study makes two unique contributions to the field of marketing research in an interactive environment. First, it is one of the few papers to review the application of sentiment analysis in marketing research comprehensively. Second, the paper focuses attention on the limitations surrounding the utilization of this technique for marketing research and provides suggestions for more effective use.

Overview of sentiment analysis

While reviewing the literature, it is apparent that a misunderstanding often exists about what constitutes sentiment analysis. To provide conceptual clarity, sentiment analysis first needs to be distinguished from the broader literature on online text mining. With text-mining applications, researchers structure a large body of data from various online sources into numerous topics or themes which *emerge* from the body of textual data. In this regard, text mining is similar to traditional content analysis, since it allows researchers to efficiently extract, classify and manage a large body of data to identify hidden patterns or trends (He *et al.*, 2013).

In contrast, sentiment analysis refers to the application of machine learning techniques to evaluate and classify attitudes and opinions on a specific topic of interest (Rambocas and Gama, 2013). Sentiment analysis focuses on extracting emotions from the online text but classifies specific problem areas into *predefined mutually exclusive categories* (Liu, 2012). These categories imply bi-polar classifications of emotions (positive and negative) and are typically represented by numeric codes for subsequent statistical analyses.

A major advantage of sentiment analysis is that it collects and analyzes online comments in real time. This is especially significant to researchers, given the exponential updating of user-generated content on social media platforms. With sentiment analysis, researchers can also automatically extract high-quality data on emotional expressions that are measurable, objective and consistent. Given these advantages, it is no surprise that sentiment analysis has attracted the interest of both academics and practitioners. However, previous research on this analytical tool has primarily focused on the technique's methodological properties by identifying efficient methods for extracting textual content from a body of online data via natural language processing, computational linguistics and text analytics (Günther and Furrer, 2013). From a marketing research perspective, however, extracting and classifying online text through sentiment analysis remains a relatively new area with burgeoning potential.

The scope and approach of the review

The objective of this review is to examine the use of sentiment analysis in the marketing literature published between 2008 and 2016. The studies identified in this paper were sourced using a combination of computerized and manual search methods. We first surveyed several online scientific databases including Business Source Complete, Proquest and Emerald, and conducted an issue by issue search of the top-ranked marketing journals. We also searched for articles using Google Scholar with the search term "sentiment analysis". Finally, we used a snowballing procedure where the references of each article on sentiment analysis were examined to identify additional studies, a search technique consistent with Babić Rosario *et al.* (2016). This broad search yielded a total of 21,456 articles.

Next, the title, abstract and keywords of each article were examined to determine whether it was relevant to the application of online sentiment analysis or simply contained keywords such as "sentiments" or "emotions". Articles not related to the application of online sentiment analysis, defined as "uses natural language processing, computational linguistics and text analytics to identify and extract the content of interest from a body of textual data" (Rambocas and Gama, 2013, p. 4), were excluded from the data set. Articles in conference papers and non-peer-reviewed journals were also eliminated. This resulted in a reduced list of 257 articles.

As Table I indicates, these articles were published across disciplines in nine subject matters, namely, Communications, Computer, Education, Engineering, Finance, Health, Marketing, Political Science and General literature review. A comparison over the review period 2008-2016 shows the preponderance of sentiment analysis articles were concentrated in the computer-related discipline (72 per cent of published work). The remaining 28 per cent of publications were distributed across several other disciplines, with Marketing accounting for 22 articles or roughly 9 per cent of this total. Figure 1 shows the publication trends over the review period.

Characteristics of the marketing articles reviewed

The authors selected the 22 marketing articles and evaluated them on four criteria:

(1) utilized sentiment analysis and not text mining or other social media analytics;
(2) applied sentiment analysis in the study of marketing-related issue/s from the perspective of the consumer, business or both;
(3) published in a peer-reviewed academic journal; and
(4) empirical in nature, with a large body of data and utilized statistical tests for data analysis.

Non-English publications were not included in this review. Only 12 articles met all four criteria (Table II), which is perhaps indicative of how recent the technique is within the marketing discipline. Many of the studies included in our sample were published in relatively high-ranking marketing journals such as *Journal of Marketing*, *Journal of Marketing Research*, *Marketing Science* and *Journal of the Academy of Marketing Science*. Of the 12 articles reviewed, the majority of studies were led by researchers from the USA (five), followed by Germany (two), The Netherlands (one), England (one), South Korea (one), Taiwan (one) and Denmark (one). The relatively low number of sentiment analysis articles published in marketing journals (average 1.33 per year) sharply contrasts with the emergence of the internet as an important tool for both consumers and marketers and is much lower than expected. We suspect that this trend might be attributable to the challenges in using the technique by researchers in the marketing discipline.

Discipline	Frequency	(%)
Communication	26	10.12
Computer	185	71.98
Education	1	0.39
Engineering	2	0.78
Finance	5	1.95
Health	3	1.17
Marketing	22	8.56
Political Science	1	0.39
Review	12	4.67
Total	*257*	*100*

Table I.
Comparison of published work on sentiment analysis from 2008 to 2016

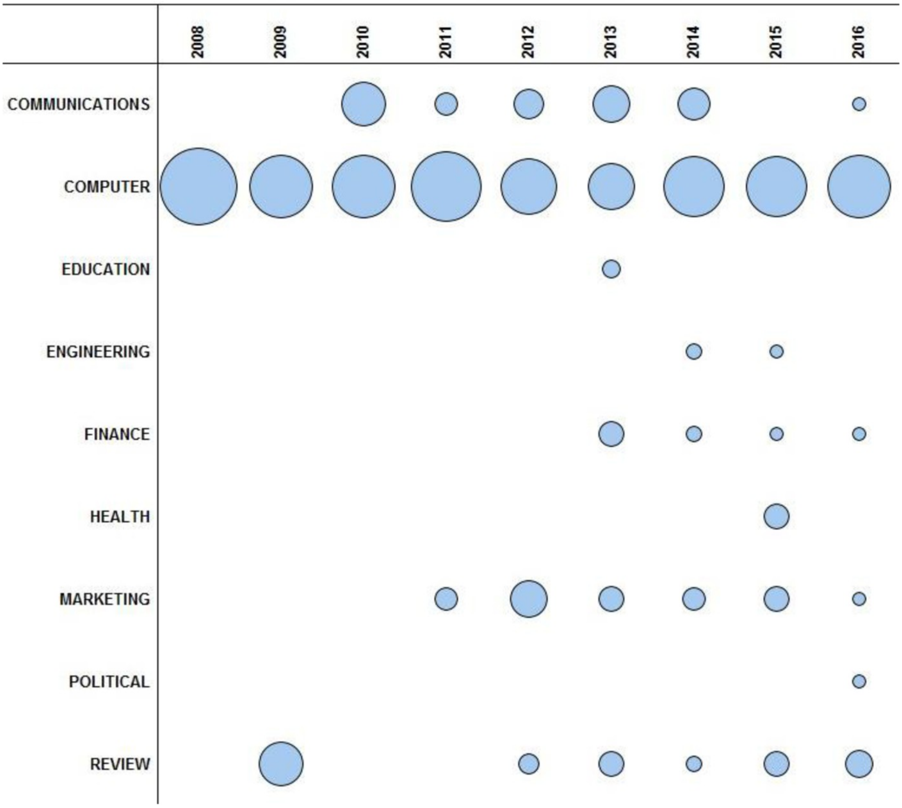

Figure 1.
Sentiment analysis publication trend by discipline from 2008 to 2016

Journal/year	2008	2009	2010	2011	2012	2013	2014	2015	2016	Grand total
Journal of Marketing						1	1			2
Journal of Marketing Research							1	1		2
Marketing Science				1	1					2
International Journal of Electronic commerce					1			1		2
Academy of Marketing Studies Journal					1					1
Academy of Marketing Science								1		1
Corporate Communications: An International Journal						1				1
The Journal of Consumer Affairs									1	1
Total	*0*	*0*	*0*	*1*	*3*	*2*	*2*	*3*	*1*	*12*

Table II.
Summary of the sentiment analysis publications published in marketing journals

Categorization of the articles

The contents of each article were categorized into three predetermined themes, which form the basis for subsequent discussions. The themes followed the categorization scheme adopted by Sousa *et al.* (2008) and sought to summarize the research design of the articles reviewed. These themes are:

(1) unit of analysis;

(2) sampling design (size, venue and sample characteristics) and

(3) methods used in sentiment detection and statistical analysis.

Following the methods used in Cummins *et al.* (2013) and Rodriguez *et al.* (2014), we prepared a data file that listed each article's title, authors, journal, volume, issue, abstract, unit of analysis, sampling characteristics, data venues, sample size, extraction methods, statistical analysis, challenges and key findings. This technique allowed us to summarize the information into a smaller set of categories and identify patterns across the articles reviewed. The authors examined and discussed each category and then proceeded with independent coding. Only minor discrepancies in coding were discovered, which were resolved through discussions. The properties of the 12 articles are presented in Table III.

Unit of analysis

The majority of studies (nine) focused on comments generated by individual users who purchased goods and services for personal and household consumption. The textual content focal object of interest related to products within the consumer market. These studies surveyed a broad range of tangible categories including books, cars, personal computers, phones, footwear, toys, clothing and accessories and personal care items. Service-related industries were also evident with analyses conducted on movies, telecommunications, mobile applications, music downloads, video games, hotels, airlines and retail stores. Two studies focused on business-related cases. The first investigated the accuracy of automated Twitter sentiment coding by third-party research companies and the second evaluated sentiments toward an organization's corporate communication strategies. Finally, there was one study that used sentiment analysis to examine boycott messages and quantify the intensity of emotions expressed.

Our review revealed that the majority of the studies focused on user-generated comments within a single industry or product category. For instance, Schweidel and Moe (2014) evaluated two leading brands in a single sector – the enterprise and development software sector. Likewise, Ludwig *et al.* (2013) extracted user-generated content on books from the fiction and non-fiction categories such as academia, religion, philosophy, humor, mystery and crime, romance, horror, science fictions, among others. Similarly, Tang *et al.* (2014) surveyed the automotive industry and extracted user-generated comments on 39 brands, while Hennig-Thurau *et al.* (2015) focused on the movie industry. Focusing on a single product category may be viewed as a pragmatic decision by researchers, given the sheer volume of user-generated content on social media.

There were, however, a few notable exceptions to the single-category approach among the studies included in this review. Homburg, Ehm and Artz (2015), for instance, extracted reviews from both a firm's sponsored online community of do-it-yourself individuals and online travel-related forums. Tirunillai and Tellis (2012), on the other hand, analyzed an extensive range of reviews on personal computing, phones, personal digital assistants, footwear, toys and data storage. Similarly, Baek *et al.* (2012) collected reviews on 28 product-related categories, and Sonnier *et al.* (2011) collected comments from a technology firm who

Journal	Title of article	Authors	Data source	Size of sample	What was sentiment analysis used for	Key findings
Journal of Marketing	More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates	Ludwig *et al.* (2013)	Amazon.com	18,682	Investigated the influence of textual properties of consumer reviews on online retailer's conversion rates	Nonlinear relation between positive affective content and conversion rates. Linguistic style and content of the review influences changes in consumers' conversion behavior
Journal of Marketing	Is Neutral Really Neutral? The Effects of Neutral User-Generated Content (UGC) on Product Sales	Tang *et al.* (2014)	Youtube and Facebook	592,310	Specify the performance implications of neutral UGC on product sales by differentiating mixed-neutral UGC, which contains an equal amount of positive and negative claims, from indifferent-neutral UGC, which includes neither positive nor negative claims	Positive and negative UGC provide opportunities for consumers to process product-related information, whereas both mixed- and indifferent-neutral UGC affect consumers' motivation and ability to process positive and negative UGC. UGC amplifies the effects of positive and negative UGC, whereas indifferent-neutral UGC attenuates them. The effects of neutral UGC on product sales thus are not truly neutral, and the direction of the bias depends on both the type of UGC and the distribution of positive and negative UGC
Journal of Marketing Research	Listening in on social media: A joint model of sentiment and venue format choice	Schweidel and Moe (2014)	Converseon extracted data from blogs, product reviews, discussion streams, social network sites and micro blogs	7,565	Modeled the sentiment expressed in user-generated comments across venues	Significant variations in sentiments expressed across venues
Journal of Marketing Research	Measuring and managing consumer sentiment in an online community environment	Homburg *et al.* (2015)	Online forums	115,000	Examined how consumers react to firm's active participation in consumer-to-consumer conversations in an online community setting	Reactions were significantly different and based on both customers functional and. social needs. Customers expressed diminishing returns as firm generated information increases
Marketing Science	A dynamic model of the effect of online communications on firm sales	Sonnier *et al.* (2011)	Proprietary Webcrawler technology	Average 2,572 per day	Investigate the effect of the volume of positive, negative and neutral online comments on daily sales performance	Significant effect of positive, negative and neutral online communications on daily sales performance

(*continued*)

Table III.
Characteristics of the published work on sentiment analysis

Table III.

Journal	Title of article	Authors	Data source	Size of sample	What was sentiment analysis used for	Key findings
Marketing Science	Does chatter really matter? Dynamics of user-generated content and stock performance	Tirunillai and Tellis (2012)	amazon.com; Epinions. com; Yahoo Shopping (yahoo.com)		Investigate whether there is a relationship between chatter and the stock market performance of the firm?	Chatter volume has the strongest relationship with returns and trading volume. This is followed by negative chatter which increases volatility (risk) in returns
International Journal of Electronic commerce	Helpfulness of online consumer reviews: readers' objectives and review cues	Baek *et al.* (2012)	Amazon.com	75,226	Investigate the factors that determine the helpfulness of online reviews and which factors are more important for helpful online reviews	The helpfulness of online reviews is determined by both peripheral cues (review rating and reviewer's credibility) and central cues (content of reviews, influence the helpfulness of reviews)
International Journal of Electronic commerce	What in consumer reviews affects the sales of mobile apps: A Multifacet Sentiment Analysis Approach	Liang *et al.* (2015)	Seventy-nine paid and seventy free apps from an iOS app store	79 paid and 70 free apps	Examine the effect of textual consumer reviews on the sales of mobile apps	The study found that although consumers' opinions on product quality occupies a larger portion of consumer reviews, their comments on service quality have a stronger unit effect on sales rankings
Academy of Marketing Studies Journal	Assessing the accuracy of automated Twitter sentiment coding	Davis and O'Flaherty (2012)	Twitter	767	Evaluated the automated sentiment coding accuracy and misclassification errors of six leading third-party companies	The study showed that automated sentiment coding had limited reliability and appears with limits for simple statements where a keyword is used
Journal of the Academy of Marketing Science	Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies	Hennig-Thurau *et al.* (2015)	Twitter	4,045,350 tweets about the 105 movies	Empirically test the impact of microblogging word-of-mouth (MWOM) on consumers adoption of new movies	Negative MWOM reviews impact on early adoption but not positive MWOM reviews, which indicates a negativity bias for MWOM
Corporate Communications: An International Journal	CSR communication strategies for organizational legitimacy in social media	Colleoni (2013)	Twitter	Network of 9,589 users and 326,000 CSR-related tweets	Investigate the effectiveness of which online social media corporate communication strategies is more effective to create convergence between corporations' corporate social responsibility (CSR) agenda and stakeholders' social expectations	Neither the engaging nor the information strategies leads to alignment. The findings show that, even when engaging in a dialogue, communication in social media is still conceived as a marketing practice to convey messages about companies
The Journal of Consumer Affairs	Consumer boycott behavior: An exploratory analysis of Twitter feeds	Makarem and Jae (2016)	Twitter	Extracted 14,685 and randomly selected 2,000	Examine and quantify the emotional intensity of boycott messages	Consumer boycott messages are more commonly motivated by instrumental motives. However, non-instrumental motives have higher emotional intensity

sold a variety of goods on the online market. These studies adopted a wider approach to the analysis and broadened the scope of the analysis to include the impact of product type on the nature of customer sentiments.

In relation to the reasons for selecting a specific product category, the majority of authors based their arguments on the specificity of the research context. For instance, Tang *et al.* (2014) indicated that the automobile industry contributed significantly to the US economy. Also, the US automobile sector has invested a considerable amount of resources into social media marketing aimed at building long-term relationships with customers and promoting sales initiatives. A similar rationale was advanced by Hennig-Thurau *et al.* (2015) who selected movies as the product of interest. On the other hand, Sonnier *et al.* (2011) explained that their selection was based on the firm's unique interest in social media monitoring and data collection. Additionally, Liang *et al.* (2015) justified the significance of studying mobile apps by highlighting the rapid development of smartphones and business opportunities for mobile applications.

However, some studies were more purposive in their approach and matched the selection strategy with the research designs. For instance, Baek *et al.* (2012) drew on the scholarly contribution of Nelson (1974) to classify 28 products as either search or experience goods. Davis and O'Flaherty (2012), on the other hand, used textual data from Twitter to assess the accuracy of automated coding and misclassification errors for positive, negative and neutral comments about a fictitious product (beer). The purposive approach allowed the researchers flexibility and control over the tweets extracted.

Sampling design

Sentiment analysis provides marketers with the opportunity to collect a vast amount of textual content from large samples of participants, an opportunity fully capitalized on by the 12 papers reviewed. More specifically, the data set of the 12 articles ranged from 767 to 4,045,350 cases. The median number of cases analyzed with sentiment analysis was 66,841. Apart from reducing sampling error, these large samples give analysts the options to use sophisticated statistical data analysis techniques.

Despite the capacity of sentiment analysis to extract user-generated content from multiple online sources, our review revealed that the majority of studies utilized a single-venue approach to extract text, with Amazon and Twitter being the most popular. Amazon is reportedly the preferred venue for data collection because of the easy accessibility of information and the high frequency of updates by the retailer. The unique advantages of amazon.com are discussed by Ludwig *et al.* (2013) who identified the site's unique ability to trace customers reviews and consumers conversion, which facilitates insight into the impact of consumer sentiments on actual purchase behavior. The information generated from the site also allowed the authors to control for the effects of extraneous variables such as price, review volume and review helpfulness.

Focusing on the automobile industry, Tang *et al.* (2014) investigated the performance implications of neutral user-generated extracted content on product sales data from YouTube and Facebook. The authors drew on the Stelzner (2012) social media marketing report, which confirmed that 92 per cent of the automobile companies use Facebook and 57 per cent use YouTube to share information with users, to justify the relevance of the sampling strategy employed. Also, Facebook and YouTube are two of the largest social media and video sharing sites with a significant amount of user-generated content.

Unlike previous studies, Sonnier *et al.* (2011) used proprietary Web Crawler technology to generate a daily log of online comments relevant to the company of interest and its products and services. This approach allowed the authors to canvass more internet platforms for a wider spectrum of reviews. Likewise, Schweidel and Moe (2014) acknowledged the impact of venue on social media monitoring. Using a social media data set provided by a leading online social media listening platform (*Converseon*), the authors extracted 7,565 comments across a wide sample of website domains and demonstrated how sentiments from different social media venues could vary in both strength and subject matter.

Methods used in sentiment detection and statistical analysis

Sentiment detection requires appraising and extracting only the emotionally laden content such as personal expressions, opinions and feelings from the textual data set. The studies reviewed in this paper employed either manual or automated detection mechanisms. Manual sentiment detection requires human input into the analysis and has the advantage of accommodating emoticons, abbreviations, sarcasm and slangs. Makarem and Jae (2016) note that emotional intensity may be detected from peripheral cues hidden in messages such as message length (long messages are typically associated with greater consumer engagements and emotions); capital letter words (an expression for shouting) and the number of profanities or insult words that may be undetected by automatic detection.

Manual coding can also accommodate language idiosyncrasies. For instance, Liang *et al.* (2015) noted that the Part-of-Speech system in Taiwan is different from mainland China and English where many words that would be considered adjectives in both languages would be transitive verbs laden with sentiments. Although advantageous in many ways, manual coding can reflect individual subjectivity, bias and misinterpretations. Also, manual coding is very time-consuming and costly, with researchers spending several weeks coding and processing data into categories (Makarem and Jae, 2016).

Conversely, automated sentiment analysis is more commonly used by researchers. The method relies on algorithms to process the volume of text. For instance, Tang *et al.* (2014) adopted a dictionary of affective words from SentiStrength2, while Ludwig *et al.* (2013) used the Linguistic Inquiry and Word Count program (LIWC), which calculates the proportion of words in text reviews that match predefined dictionaries. LIWC is flexible in nature and facilitates the analysis of files in multiple languages quickly and efficiently (Scholand *et al.*, 2010). These programs, however, require training documents or data corpus, which can be generated from the original data set.

Regarding statistical analysis, the majority of studies complemented sentiment analysis with econometric linear and non-linear modeling, panel data modeling and time series analysis. However, while linear regression is the most popular analytical approach, other multivariate analysis techniques such as multiple discriminant analysis and structural equation modeling are notably absent.

The application of sentiment analysis in marketing research

In reviewing the application of sentiment analysis, we found the majority of articles focused on quantifying the effect of online textual comments on corporate financial performance as measured by sales, preferential consumer behavior and corporate stock performance. In almost all cases, the research models were causal and driven by a strong theoretical underpinning.

For instance, Sonnier *et al.* (2011) considered the effect of the volume of positive, negative and neutral user-generated comments on sales. The authors' study was among the first in the marketing literature to model the dynamic effects of online communication and found positive feedback has the greatest effect on sales followed by negative and neutral comments. In an extension of that work, Tang *et al.* (2014) showed that mixed-neutral comments intensify the impact of positive and negative comments while indifferent-neutral comments diminish the effect. The findings suggest that mixed-neutral comments are associated with credibility and honesty and aid in consumers deliberate processing of product-related information.

Additional evidence of the influence of online comments on sales of a specific product is provided by Liang *et al.* (2015) who used sentiment analysis to categorize customer feedback and model the influence of electronic word of mouth on the sale of mobile phone applications. The authors concluded that sentiments on overall product's quality and service attributes would effectively predict overall sales but comments on service have a stronger effect.

Drawing on human communication theories, Ludwig *et al.* (2013) used dynamic panel data modeling to dissect text reviews into extreme positive and negative changes and advanced the notion that affect and communication style can increase consumers' product conversion rates via better rapport, credibility and shared perceptions among online users. This effect existed even after controlling for customers rating, weekly review quantity, affective content variation, price discounts and reviewer expertise. Similarly, Hennig-Thurau *et al.* (2015) found that unlike positive reviews, negative reviews impacted on consumers' early adoption of movies – a relationship rooted in the theory of information diagnosticity and prospect theory. The authors also included a series of control variables, namely, movie hype, production budget and studio.

In a departure from the focus on sales as the primary variable of interest, Makarem and Jae (2016) used sentiment analysis in conjunction with content analysis to show that boycott messages driven by non-instrumental motives have higher emotional intensity. Additionally, Tirunillai and Tellis (2012) found that negative product reviews and ratings (online chatter) increased volatility in returns and significantly impacted on traded volume. Finally, in a related stream of research, Schweidel and Moe (2014) provided evidence that the effect of online sentiments on a brand's stock market performance may vary depending on the social media venue where the comments are posted.

In summary, most of the articles reviewed used rigorous, comprehensive research designs that were primarily driven by pre-defined research frameworks. In this regard, sentiment analysis was primarily used in explanatory or causal research designs rather than in descriptive or exploratory studies.

Challenges of applying sentiment analysis in marketing research

Despite the technique's potential benefits, it is still relatively new to the marketing discipline with evolving methodological properties. Given the methodology's embryonic stage, questions on its relevance and applicability to marketing research are at the forefront, and there is growing recognition of the various application challenges. This section groups these challenges into three broad categories (technical, practical and ethical) and discusses the implications of each.

Technical limitations

Liu (2012) described the technical limitations of sentiment analysis as "multifaceted" based on object identification, feature extraction and opinion grouping. In object identification and feature extraction, sentiment analysis only extracts and classifies comments related to the study's focal object. However, the views expressed in textual logs often refer to many different issues, sometimes having an indirect link to the problem at hand. These are typically excluded in the analysis which compromises the accuracy and validity of the sentiment analysis. The CEO and President of Beyond the Arc, Steven Ramirez, attributed these misclassifications to the early development of machine learning as well as the "cultural factors, linguistic nuances and differing contexts which make it extremely difficult to turn a string of written text into a simple pro and con statement" (Mullich, 2012). However, the classification hurdles may be more pronounced in some product categories than others. Venkat Viswanathan – the CEO and founder of LatentView (a data analytics company that works with Fortune 500 companies) – suggested that the classification accuracy is higher among consumer electronics products, primarily because their distinct features and relatively limited number of main features simplifies classification.

Finally, technical limitations relating to opinion groupings refer to lower accuracy and high response errors in classifying textual data. Online texts are notorious for unstructured textual content laden with grammatical and spelling errors, which make sentiment detection more difficult. The problems are exacerbated by the use of multiple phrases, nouns or linguistic variances (slangs) to describe features and attributes. Even the context of the text should be taken into account since it can affect the accuracy of classification. Conversations are typically domain-specific, and opinions can be misclassified from positive to negative depending on the conversation context. A detailed discussion of errors in classifications can be found in the study conducted by Davis and O'Flaherty (2012).

Criticisms are also rooted in the classification approach given that the technique classifies data into mutually exclusive groups. While the limited groups are simple and easy to present, the classification ignores the diversity and richness of the online comments in addition to excluding data that straddle multiple categories or are less readily assigned and thus not captured by the analysis. Davis and O'Flaherty (2012) report that statements without keywords, or statements in which keywords are reversed through negation or context, are more likely to be miscoded. Also, neutral statements are problematic for coders and brand managers are urged to independently verify the level of accuracy across a broad range of statement types.

Researchers also face challenges in assessing brand sentiments across social media venues, given that media platforms appeal to different audiences and have different usage patterns. Very often, the data are single sourced (drawn from a single type of site), which fails to consider the diversity of media appeal, interest and attention (Ordenes *et al.*, 2017). For instance, sites such as Yelp.com and Epinons.com host reviews on consumer-related product and services including restaurants, nightlife, shopping and even medical providers. Other sites like Twitter and Facebook are more generic and host a broad range of data on an unlimited number of subject matters. Additionally, in their investigations of sentiments toward brands, Schweidel and Moe (2014) found that blogs have the highest percentage of positive comments, while forums have the highest percentage of negative comments. Moreover, the authors noted that the quantity and nature of discussions on blogs and microblogs exhibit temporal inconsistency, while forums tend to be more consistent over time.

The demographic profile of users on the various social media platforms also differs. For instance, Instagram is particularly appealing to a younger (18-29 years) African American and Latino population, while LinkedIn usage rates are higher in a more mature population who have graduated from college and Pinterest appeals to middle-class professional women between the ages 25 and 34 (Duggan, 2015). These differences are likely to influence the nature of the comments found on these sites and thus the sentiments that are extracted. This challenge is especially significant given that most marketing studies using sentiment analysis have extracted data from a single online source and those that draw data from multiple sources usually do so in a fragmented manner. Even if data are extracted from multiple sites, sampling bias may still exist given that most researchers usually consolidate responses across platforms and ignore the response variances based on where the data come from (Schweidel and Moe 2014). Therefore, the efficiency gains from gathering and classifying information from a single internet source often masks population differences across diverse media platforms that can challenge the validity of the research findings.

Practical limitations

The issues of research cost and accuracy are real concerns for market researchers. Organizations planning to use sentiment analysis either acquire expensive software and infrastructure or pay significant consulting fees to external companies. This could be a disincentive to using this technique for smaller firms who cannot afford expensive third-party services. In addition to the high research cost, the burgeoning volume of online data creates administration and data storage challenges. Moreover, although the abundance of user-generated content can provide useful insights, they are usually laden with short and irregular phrases, which hinder fast and effective sentiment classification (Saif *et al.*, 2012).

Additional challenges in using sentiment analysis in real-life applications include problems with coding text using automatic coding systems (Davis and O'Flaherty, 2012). The authors noted that the misclassification is especially important when lengthy sentences are extracted or when sentences do not contain keywords. Furthermore, misclassification and errors in coding are also common when neutral opinions are expressed. Misclassification therefore presents difficulties in measuring and tracking brand-related sentiments and may distort a study's conclusions.

In addition to misclassifications and inaccuracies, the evidence points to variations in both textual style and content based on gender (Thelwall *et al.*, 2010; Otterbacher, 2013). Boldon and Carter (2013) also described how age differences can bias the findings of sentiment analysis, given that older customers are less likely to share information on social media as opposed to a younger more digital-centered group. However, despite these potential confounds, it is clear from the articles reviewed that authors have often not considered the influence of demographics on their analysis and findings.

The dramatic increase in the number of deceptive or fake reviews has also emerged as another practical issue that challenges the trustworthiness of sentiment analysis (Heydari *et al.*, 2015). Deceptive reviews are created sinister motives designed to mislead businesses and consumers. To the extent that the opinions expressed publicly on websites and other social media websites do not represent authentic sentiments, they threaten the accuracy of the conclusions derived from sentiment analysis.

The inefficacy of sentiment analysis in cross-cultural research is another reality that researchers confront when using sentiment analysis as a research tool. It is instructive, for instance, that all the articles in our review, with the notable exception of Liang *et al.* (2015), relied on English language text for their analysis. This imbalance may perhaps be attributed to the widely available lexicons in English, most of which offer strong and reliable convergence between human and automatic coding (Ludwig *et al.*, 2013). Additionally, most of the tools and classification documents used are in English, which impedes analysts' abilities in building multi-language classifiers and conducting cross-cultural analyses (Liu, 2012).

Although improvements in automated technology make language translation easier, automatic translators can be unreliable. Research shows that automated translation often fails to identify and adapt to varying cultural idiosyncrasies and can compromise the validity of the results (Lotz and Van Rensburg, 2016). Specifically, automated translators fail to consider cultural differences in expressions and can overlook subtle nuances in the way opinions are communicated. These inefficiencies widen the gap between actual meaning and translated meaning and may cause misclassifications. In this regard, researchers should exercise caution in using automated translation when collecting opinions on foreign customers or conducting cross-cultural comparisons based solely on online linguistic expressions.

Ethical limitations

Although there is ample literature available on ethical standards for online research, sentiment analysis raises several fundamental ethical concerns, which to a large extent have been under-reported in the articles reviewed. One main concern hinges on the right to user privacy since the technology enables researchers to surreptitiously collate a comprehensive set of personal and confidential information about an individual through their online activity that may not have otherwise been made public (Nunan and Di Domenico, 2013). Although this data are useful to marketers interested in building a more accurate customer profile, questions on commercial exploitation arise, especially if third-party companies contact users via unsolicited direct marketing initiatives. The technique therefore fails to adequately consider users' right of refusal to participate in or withdraw from market research programs.

Further, the trustworthiness of the companies responsible for data extracting can also be an important issue, given the sensitive nature of personal data, which, if exposed, may cause irreparable damage to consumers (Dumas *et al.*, 2014; Hasan *et al.*, 2013). In this regard, researchers have a duty to maintain the highest ethical standards in their research activities to protect participants from harm and safeguard their privacy wherever possible.

Recommendations for sentiment analysis application in marketing research

This review highlighted several constraints of sentiment analysis that can compromise the validity of the information generated. For instance, our findings indicated that researchers often rely on a single venue for data extraction rather than using a combination of sources such as blogs, discussion boards and social network sites. This practice increases the potential for sampling bias to emerge since it does not account for user heterogeneity across social media platforms. Marketers and users of sentiment analysis should therefore view data extracted from these narrow sources with skepticism and insist on extracting data from a wider cross-section of users across multiple venues who display varying demographic and psychographic characteristics. Schweidel and Moe (2014) warn, however, that merely aggregating data from multiple forums will not automatically reduce sampling bias and may continue to yield misleading results, especially if comments are simply combined. It is therefore necessary for future researchers to consider using robust statistical analysis that can either control or explicitly account for the impact of the data source on their results.

In addition to sampling considerations, the high cost of contracting specialized vendors to track brand-specific comments and code consumer sentiments across a wide range of social media platform can be prohibitive, especially for smaller companies. Smaller companies (who are under-staffed and under-resourced) may thus want to consider open-source text-analytics tools that can be used for sentiment analysis (e.g. Watson Natural Language; Python NLTK; RapidMiner). Although many of these online interfaces have restricted functions, it allows users to run common queries on any topic of interest and analyze text-based sentiments.

Sentiment analysis in a multilingual environment also remains a challenge, primarily because of the inadequacy of online language translators and the fact that publicly available lexicons are usually only available in English. While online translators help bridge the language divide, they cannot understand language context and consequently often fail to provide an accurate rendition of the original text. However, improvements in machine learning technology are now showing dramatic improvements in translation. Grimes (2014) suggested that cross-cultural analysis is imminent given the recent advancements in natural language processing, stylistic analysis and profile extraction within the academic and commercial environments. However, to date, the application of such technologies remains sparse. Efforts still need to be devoted to developing training documents and foreign language data corpus which can be used to enhance the applicability of sentiment analysis to a broader range of material.

As a more short-term measure, Gopaldas (2014) suggested that the cultural divide can be bridged by employing new types of skills in the data environment. The author suggested that big data companies like Facebook and Google should consider hiring graduates from clinical psychology and cultural anthropology to complement scientists and statisticians, to absorb multimodal data and recognize linguistic variances (humor, irony and sarcasm) and capture more holistic market sentiments. Also, researchers have started to address the lexicon gaps. For instance, Chen and Skiena (2014) integrated a variety of linguistic resources to produce "high-quality" lexicons for 136 different languages and concurred that similar work is being done in natural language processing tasks with specific dictionaries and seed words. However, the authors admit that more work must be done in the technical areas of learning modifiers, negations terms and sentiment attributions.

Researchers can also address some of the methodological limitations of automated text processing by integrating human analysis to classify sentiments according to the language context as well as interpret the valence of a sentiment from the text. The benefits of human coders in an automated sentiment analysis environment are outlined by Venkat Viswanathan, CEO and founder of *LatentView* (a data analytics company that works with Fortune 500 companies). According to Viswanathan, "Some topics and conversations are easy to classify, some are complex [. . .] In any case, you always need humans to provide the context".

The review also presented implications from an academic and research perspective. For instance, incorporating sentiment analysis into marketing doctoral programs as a mainstream methodology may not only increase usage of the technique in academic research but will produce a cadre of trained users who can provide this service at lower cost for marketing practitioners. Expanding the pool of users will also increase competition among service providers which should further reduce costs. The analytical rigor with which the technique is applied also stands to benefit once a critical mass of users is attained, since there will be a wider understanding of the standards which should be used to judge the quality of findings produced by this technique.

Additionally, marketers and researchers should be mindful of the devastating impact of deceptive reviews on the validity of the findings. Consequently, through spam detection techniques (automated or manual), marketers should vigorously identify and isolate these predatory comments from the analysis. Admittedly, these detection methods could be very complex and may require considerable resources to develop and implement. Nevertheless, the purification of content could assist in improving the accuracy of the sentiment analysis and the overall results.

The ethical issues surrounding the execution, analysis and presentation of data remain a contentious area that requires a balance between market research goals and an individual's right to privacy, especially if the data are being extracted from private sources. Holmes (2009) suggests that posting a message on online forums or social media explaining the nature of the research and inviting volunteers can help minimize the potential damage to some online communities. This ensures that users are reminded of how and when data will be collected as well as the reasons why it is being collected. The risk of following this recommendation, however, is that it exposes the sample to self-selection bias, which is a limitation of this approach.

In the absence of specific ethical rules regarding the application of sentiment analysis, academic researchers may consider following the general ethical guidelines outlined by the institutional review board (IRB) in their institution. In relation to privacy, IRB principles dictate that the researcher should ensure that there are sufficient allowances within the research design to protect the privacy of the participants and maintain the confidentiality of the data collected. In the case of sentiment analysis, this can be done through a combination of steps that may include collecting data in anonymous environments, purging identifying information from the data set and restricting the number of personnel with access to the data.

Another critical ethical area for sentiment analysis researchers is informed consent, which guarantees voluntary participation based on participants' knowledge of the intended study. Lunnay *et al.* (2015) suggested that social media researchers (including users of sentiment analysis) should give participants the right to participate or not participate before the research starts. In the case of minors (less than 18 years), researchers should make conscientious efforts to identify minors and seek parental consent before extracting data, although this may be difficult, given that there may not be any practical way of discovering the true age of online participants.

Table IV outlines the main limitations we have identified that are experienced by researchers when using sentiment analysis in marketing research and summarizes our suggestions to mitigate their effects.

Conclusions

It is clear that the online environment provides rich and valuable information about consumer opinions, though harnessing and analyzing that data can be difficult. Improvements in computer modeling and techniques like sentiment analysis provide

Challenges	Aspects	Recommendations
Technical limitations	Accuracy, reliability and validity	Include a cross-section of product categories in the analysis Extract data from multiple venues that appeal to different consumer demographical characteristics Triangulate results with more traditional research methods
Practical limitations	Cost concerns	Free services are available, but the brand managers should be cautioned when relying on these free services for strategic decisions Invest in programs that will produce a cadre of trained users who can provide this service at lower cost for marketing practitioners
	Miscoding	Integrate human analysis with automated text processing to classify opinions
	Cross-cultural research	Cross-cultural analysis is imminent given the recent advancements in natural language processing, stylistic analysis and profile extraction within the academic and commercial environments
	Deceptive reviews	Use spam detection techniques (automated or manual) to identify and isolate these predatory comments
Ethical concerns	The right to user privacy	IRB ethical guidelines for online research should be followed to preserve the rights of consent, privacy, confidentiality and anonymity, assurances of voluntary participation and protection from harm
	Exploitation	Special effort should be taken to remove children and other vulnerable groups from the analysis

Table IV. A summary of the challenges and recommendations in applying sentiment analysis powerful mechanisms by which this information can be converted into deep insights about the attitudes held by a brand's target market. The recommendations for use provided in the current research are thus the first step to guide marketers and academics who wish to adopt this nascent technology. Hopefully, by integrating these recommendations into their research designs, academics and marketers will be better equipped to produce more enriching, meaningful and rigorous analyses. It is anticipated, therefore, that as sentiment analysis emerges as a powerful tool to understand consumer opinion, the technique will attain the methodological rigor associated with other more widely used analytic techniques and feature more prominently in future marketing research.

References

Babić Rosario, A., Sotgiu, F., De Valck, K. and Bijmolt, T.H. (2016), "The effect of electronic word of mouth on sales: a meta-analytic review of platform, product, and metric factors", *Journal of Marketing Research*, Vol. 53 No. 3, pp. 297-318.

Baek, H., Ahn, J. and Choi, Y. (2012), "Helpfulness of online consumer reviews: readers' objectives and review cues", *International Journal of Electronic Commerce*, Vol. 17 No. 2, pp. 99-126.

Boldon, R. and Carter, R. (2013), "Lost in translation/managing multi-lingual A/V and metadata in the digital supply chain", *Journal of Digital Media Management*, Vol. 1 No. 4, pp. pp. 330-335.

Chen, Y. and Skiena, S. (2014), "Building sentiment lexicons for all major languages", *ACL*, Vol. 2, pp. 383-389, available at: http://ai2-s2dfs.s3.amazonaws.com/c5e3/b065e352a93d8754b86baaf8ec20bf81a5c3.pdf (accessed 20 March 2017).

Colleoni, E. (2013), "CSR communication strategies for organizational legitimacy in social media", *Corporate Communications: An International Journal*, Vol. 18 No. 2, pp. 228-248.

Cummins, S., Peltier, J.W., Erffmeyer, R. and Whalen, J. (2013), "A critical review of the literature for sales educators", *Journal of Marketing Education*, Vol. 35 No. 1, pp. 68-78.

Davis, J.J. and O'Flaherty, S. (2012), "Assessing the accuracy of automated Twitter sentiment coding", *Academy of Marketing Studies Journal*, Vol. 16, pp. 35-50.

Duggan, M. (2015), "The demographics of social media users", *Pew Research Center: Internet, Science & Tech*, available at: www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users (assessed 20 March 2017).

Dumas, G., Serfass, D.G., Brown, N.A. and Sherman, R.A. (2014), "The evolving nature of social network research: a commentary to Gleibs", *Analyses of Social Issues and Public Policy*, Vol. 14 No. 1, pp. 374-378.

Erevelles, S., Fukawa, N. and Swayne, L. (2016), "Big data consumer analytics and the transformation of marketing", *Journal of Business Research*, Vol. 69 No. 2, pp. 897-904.

Gonçalves, P. Araújo, M. Benevenuto, F. and and Cha, M. (2013), "Comparing and combining sentiment analysis methods", *paper presented at the ACM Conference on Online Social Networks, Boston, MA,* 7-8 October 2013. available at: https://arxiv.org/pdf/1406.0032.pdf (accessed 3 February 2017).

Gopaldas, A. (2014), "Marketplace sentiments", *Journal of Consumer Research*, Vol. 41 No. 4, pp. 995-1014.

Grimes, S. (2014), "Can sentiment analysis decode cross-cultural social media", *Breakthrough Analysis*, available at https://breakthroughanalysis.com/2014/03/18/can-sentiment-analysis-decode-cross-cultural-social-media (accessed 4 May 2016).

Günther, T. and and Furrer, L. (2013), "GU-MLT-LT: Sentiment analysis of short messages using linguistic features and stochastic gradient descent", *paper presented at the Second Joint Conference on Lexical and Computational Semantic (SemEval 2013), Atlanta, Georgia,* 14-15 June, available at: www.aclweb.org/anthology/S/S13/S13-2.pdf#page=364 (accessed 1 May 2017).

Hasan, O., Habegger, B., Brunie, L., Bennani, N., and and Damiani, E. (2013), "A discussion of privacy challenges in user profiling with big data techniques: the eexcess use case", 2013 *IEEE International Congress on Big Data (BigData Congress)*, pp. 25-30, available at https://pdfs.semanticscholar.org/b5fb/e425c94c8e6477f1e4abd2af47bb1cac5f71.pdf (accessed 16 February 2017).

He, W., Zha, S. and Li, L. (2013), "Social media competitive analysis and text mining: a case study in the pizza industry", *International Journal of Information Management*, Vol. 33 No. 3, pp. 464-472.

Hennig-Thurau, T., Wiertz, C. and Feldhaus, F. (2015), "Does twitter matter? the impact of microblogging word of mouth on consumers' adoption of new movies", *Journal of the Academy of Marketing Science*, Vol. 43 No. 3, pp. 375-394.

Heydari, A., Ali Tavakoli, M., Salim, N. and Heydari, Z. (2015), "Detection of review spam: a survey", *Expert Systems with Applications*, Vol. 42 No. 7, pp. 3634-3642.

Holmes, S. (2009), "Methodological and ethical considerations in designing an Internet study of quality of life: a discussion paper", *International Journal of Nursing Studies*, Vol. 46 No. 3, pp. 394-405.

Homburg, C., Ehm, L. and Artz, M. (2015), "Measuring and managing consumer sentiment in an online community environment", *Journal of Marketing Research*, Vol. 52 No. 5, pp. 629-641.

Liang, T.P., Li, X., Yang, C.T. and Wang, M. (2015), "What in consumer reviews affects the sales of mobile apps: a multifacet sentiment analysis approach", *International Journal of Electronic Commerce*, Vol. 20 No. 2, pp. 236-260.

Liu, B. (2012), "Sentiment analysis and opinion mining", *Synthesis Lectures on Human Language Technologies*, Vol. 5 No. 1, pp. 1-167.

Lotz, S. and Van Rensburg, A. (2016), "Omission and other sins: tracking the quality of online machine translation output over four years", *Stellenbosch Papers in Linguistics*, Vol. 46 No. 0, pp. 77-97.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E.C., Wetzels, M. and Pfann, G. (2013), "More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates", *Journal of Marketing*, Vol. 77 No. 1, pp. 87-103.

Lunnay, B., Borlagdan, J., McNaughton, D. and Ward, P. (2015), "Ethical use of social media to facilitate qualitative research", *Qualitative Health Research*, Vol. 25 No. 1, pp. 99-109.

Makarem, S.C. and Jae, H. (2016), "Consumer boycott behavior: an exploratory analysis of twitter feeds", *Journal of Consumer Affairs*, Vol. 50 No. 1, pp. 193-223.

Mullich, J. (2012), "Improving the effectiveness of customer sentiment analysis", *Data Informed*, available at: http://data-informed.com/improving-effectiveness-of-customer-sentiment-analysis/ (assessed 23 February, 2017).

Nelson, P. (1974), "Advertising as information", *Journal of Political Economy*, Vol. 82 No. 4, pp. 729-754.

Nunan, D. and Di Domenico, M. (2013), "Market research and the ethics of big data", *International Journal of Market Research*, Vol. 55 No. 4, pp. 2-13.

Ordenes, F.V., Ludwig, S., De Ruyter, K., Grewal, D. and Wetzels, M. (2017), "Unveiling what is written in the stars: analyzing explicit, implicit, and discourse patterns of sentiment in social media", *Journal of Consumer Research*, available at: http://openaccess.city.ac.uk/16047/1/Villa%20Roel%20et%20al.%202017.pdf (accessed 3 May 2017).

Otterbacher, J. (2013), "Gender, writing and ranking in review forums: a case study of the IMDb", *Knowledge and Information Systems*, Vol. 35 No. 3, pp. 645-664.

Rambocas, M. and Gama, J. (2013), "The role of sentiment analysis", Working Paper [489], FEP-UP, University of Porto, April, available at: https://pdfs.semanticscholar.org/acd0/c9f75152acd2a622be442d20f96b0a3225d4.pdf (accessed 17 January 2017).

Rodriguez, M.L., Dixon, A.W. and Peltier, J. (2014), "A review of the interactive marketing literature in the context of personal selling and sales management: a research agenda", *Journal of Research in Interactive Marketing*, Vol. 8 No. 4, pp. 294-308.

Saif, H., He, Y., and Alani, H. (2012), "Semantic sentiment analysis of twitter", *11th International Semantic Web Conference (ISWC 2012), Boston, MA, 11-15 November*, available at: http://oro.open.ac.uk/34929/1/76490497.pdf (accessed 7 January 2017).

Scholand, A.J., Tausczik, Y.R. and Pennebaker, J.W. (2010), "Assessing group interaction with social language network analysis", *International Conference on Social Computing, Behavioral Modeling, and Prediction*, Springer, Berlin, pp. 248-255.

Schweidel, D.A. and Moe, W.W. (2014), "Listening in on social media: a joint model of sentiment and venue format choice", *Journal of Marketing Research*, Vol. 51 No. 4, pp. 387-402.

Sonnier, G.P., McAlister, L. and Rutz, O.J. (2011), "A dynamic model of the effect of online communications on firm sales", *Marketing Science*, Vol. 30 No. 4, pp. 702-716.

Sousa, C.M., Martínez-López, F.J. and Coelho, F. (2008), "The determinants of export performance: a review of the research in the literature between 1998 and 2005", *International Journal of Management Reviews*, Vol. 10 No. 4, pp. 343-374.

Stelzner, M.A. (2012), "Social media marketing industry report", *Social Media Examiner*, available at: www.socialmediaexaminer.com/SocialMediaMarketingIndustryReport2012.pdf (accessed 1 May 2017).

Tang, T., Fang, E. and Wang, F. (2014), "Is neutral really neutral? The effects of neutral user-generated content on product sales", *Journal of Marketing*, Vol. 78 No. 4, pp. 41-58.

Thelwall, M., Wilkinson, D. and Uppal, S. (2010), "Data mining emotion in social network communication: gender differences in MySpace", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 1, pp. 190-199.

Tirunillai, S. and Tellis, G.J. (2012), "Does chatter really matter? Dynamics of user-generated content and stock performance", *Marketing Science*, Vol. 31 No. 2, pp. 198-215.

Corresponding author
Meena Rambocas can be contacted at: Meena.Rambocas@sta.uwi.edu