Journal of Consumer Marketing

Social media sentiment analysis: lexicon versus machine learning

Chedia Dhaoui, Cynthia M. Webster, Lay Peng Tan,

Social media sentiment analysis: lexicon versus machine learning

Chedia Dhaoui, Cynthia M. Webster and Lay Peng Tan
Department of Marketing and Management, Faculty of Business and Economics, Macquarie University, Sydney, Australia

Abstract

Purpose – With the soaring volumes of brand-related social media conversations, digital marketers have extensive opportunities to track and analyse consumers' feelings and opinions about brands, products or services embedded within consumer-generated content (CGC). These "Big Data" opportunities render manual approaches to sentiment analysis impractical and raise the need to develop automated tools to analyse consumer sentiment expressed in text format. This paper aims to evaluate and compare the performance of two prominent approaches to automated sentiment analysis applied to CGC on social media and explores the benefits of combining them.

Design/methodology/approach – A sample of 850 consumer comments from 83 Facebook brand pages are used to test and compare lexicon-based and machine learning approaches to sentiment analysis, as well as their combination, using the LIWC2015 lexicon and RTextTools machine learning package.

Findings – Results show the two approaches are similar in accuracy, both achieving higher accuracy when classifying positive sentiment than negative sentiment. However, they differ substantially in their classification ensembles. The combined approach demonstrates significantly improved performance in classifying positive sentiment.

Research limitations/implications – Further research is required to improve the accuracy of negative sentiment classification. The combined approach needs to be applied to other kinds of CGCs on social media such as tweets.

Practical implications – The findings inform decision-making around which sentiment analysis approaches (or a combination thereof) is best to analyse CGC on social media.

Originality/value – This study combines two sentiment analysis approaches and demonstrates significantly improved performance.

Keywords Sentiment analysis, Social media, Consumer-generated content

Paper type Research paper

Introduction

The considerable advancements of social media during the past decade, along with the profusion of digital channels, such as social networking sites (e.g. Facebook), microblogs (e.g. Twitter) and media sharing (e.g. Instagram or Youtube), have revolutionised not only the way brands communicate with their consumers but also the roles of consumers in the marketing process. In a sense, social media gives consumers the same, if not more voice than brands, disrupting marketing processes and creating serious dilemmas and challenges for marketers (Constantinides *et al.*, 2008). Brand managers can no longer afford to ignore their consumers' important online voice (Gensler *et al.*, 2013). They are also offered new opportunities to tap into the unfettered consumer-generated content (CGC) readily available on social media platforms. With digital marketing now treated as a "many-to-many conversation" between businesses and consumers as well as among consumers themselves (Lusch *et al.*, 2010), the traditional one-way business-to-consumers transmissions is becoming obsolete.

A recent trend in the digital marketing analytics sphere is to track and analyse consumers' feelings and opinions about specific brands, products or services attributed to the CGC on social media (Hemann and Burbary, 2013). The objective is to classify positive and negative CGC, typically text-based, according to some manual or automated classification methods. For example, marketers can retrieve timely consumer feedback on a new product by evaluating consumer sentiment expressed in the comments on a Facebook post or in tweets with a specific hashtag related to the product.

Given the large volume of CGC, commonly referred to as "Big Data" that has grown along with the uptake of social media platforms, the qualitative manual analysis of consumers' sentiment conveyed in online brand-related content is no longer practical. To put this into perspective, Twitter generates over 500 million tweets each day, and there are 4.75 billion pieces of content per day on Facebook. This raises the need to develop automated tools for identifying and analysing consumer sentiment expressed in text (Wang *et al.*, 2012).

Two prominent approaches to automated sentiment analysis exist. Classification using a lexicon of weighted words (Taboada *et al.*, 2011) is a widely used approach to sentiment analysis in the marketing research community (Bolat and O'Sullivan, 2017), as it does not require any pre-processing or training of the classifier. Alternatively, the machine learning approach to sentiment

The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/0736-3761.htm**

Journal of Consumer Marketing
34/6 (2017) 480–488
© Emerald Publishing Limited [ISSN 0736-3761]
[DOI 10.1108/JCM-03-2017-2141]

Received 13 March 2017
Revised 16 June 2017
Accepted 5 July 2017

analysis, also described as a supervised learning approach, is often reported to be more accurate (Pang *et al.*, 2002; Chaovalit and Zhou, 2005) and has also been used in marketing research (Pathak and Pathak-Shelat, 2017). However, the machine learning approach requires a training phase that is either conducted by the researchers themselves or by the sentiment software provider. As each of these methods has its advantages and limitations, marketers and researchers need to carefully verify the accuracy of the classification (Brown *et al.*, 1990) to avoid acting on inaccurate data analysis outcomes (Canhoto and Padmanabhan, 2015). Furthermore, given the wide range of social media platforms and their specificities as to what type of content consumers can create (e.g. Facebook comments, Twitter tweets, the use of emoticons, emojis, hashtags, the use of abbreviations, slang language, etc.), existing sentiment analysis approaches, typically tested on well-formed English language texts, require careful validation before being used by marketers on social media data.

The purpose of this paper is to compare the lexicon-based approach and the machine learning approach to address three research questions:

RQ1. Are these two existing sentiment analysis techniques appropriate for the analysis of social media conversations?

RQ2. To what extent do the results from the two approaches differ when used on social media conversations?

RQ3. Does a combined approach improve the overall accuracy of the sentiment classification of social media conversations?

To answer these questions, we first summarize the challenges with regards to text classification methods for sentiment analysis used today on social media data. We then outline the research method and empirically evaluate the lexicon-based, machine learning and combined approaches using a large sample of CGC on Facebook brand pages.

Literature

Studying the language people use to better understand their thoughts and behaviours is not new in the social sciences (Krippendorff, 2012). Sentiment has long been measured using self-reported data in consumer surveys such as the Michigan Consumer Sentiment Surveys. However, the use of self-reported data has its limitations, as do most self-reported data. With surveys, marketing researchers rely on consumers' abilities to accurately recall their felt experiences, which may be highly variable and difficult to verbalize and reconstruct (Cooke and Buckley, 2008; Nabi and Oliver, 2009). In contrast, with experiments, there are concerns relating to the artificial circumstances in which data are gathered, which may constrain consumers' emotional responses (Nabi, 2007).

Today, social media platforms are popular vehicles to study consumer sentiment on a large scale and within a natural setting (Kivran-Swaine *et al.*, 2012) due to the significant share of online conversations expressing consumers' thoughts, feelings and opinions about products and brands (Jansen *et al.*, 2009). The analysis of sentiment in textual content often relies on simple sentiment annotation tasks during which annotators must determine whether a sentence is positive, negative or neutral (Rosenthal *et al.*, 2015; Mohammad *et al.*, 2015). Given the large volume of social media content, manual sentiment annotation is impractical.

Supported by most automated text classification tools, sentiment analysis is regularly used by marketers for a computer-supported, rapid, scalable and effective way of gauging consumer's sentiment (Murdough, 2009). Automated sentiment analysis receives increasing attention from both academia and industry (Chen and Zimbra, 2010) and has become one of the key techniques for handling large volumes of social media data. Typically, automated sentiment analysis techniques are used to classify any text-based document into predefined categories reflecting the polarity of sentiment referred to in the text. Recently, Canhoto and Padmanabhan (2015) have undertaken a comparative study of automated versus manual analysis of social media conversations. Their findings show low levels of agreement between manual and automated analysis, which is of "grave concern given the popularity of the latter in consumer research" (Canhoto and Padmanabhan, 2015, p. 1141).

Automated classification of expressed sentiment in social media conversations is challenging for several reasons. First, identifying opinions and sentiments from text-based natural language requires a deep understanding of the explicit and implicit, regular and irregular and syntactical and semantic language rules (Cambria *et al.*, 2013). Furthermore, sentiment analysis faces difficulties in using natural language processing (NLP) on unstructured text, typical of social media conversations and CGC in general. For instance, CGC content typically reflects the instant and informal nature of communication on social media (Canhoto and Padmanabhan, 2015). The content typically is a free-flowing text, casual in its word and grammar usage (Tirunillai and Tellis, 2014), commonly includes abbreviations, misspellings, emoticons, emojis and often uses SMS-like syntax, which current sentiment analysis methods do not adequately support. Additionally, particular platform features, like the 140-character limit for Twitter messages, impede the effectiveness of current automated sentiment analysis tools (Kiritchenko *et al.*, 2014). Finally, the sheer volume of social media conversations is a significant challenge. Automated technologies turn that challenge into an opportunity by obviating the need for costly and risk-prone manual analysis instead leveraging computerised procedures to draw insights from social media conversations.

Key approaches to automated sentiment classification

Selecting the right automated sentiment analysis method for social media data is crucial for achieving high accuracy in content classification. There exist two prominent approaches to text classification used for sentiment analysis: lexicon-based and machine learning. Both approaches to sentiment classification typically classify any given text into positive, negative or neutral sentiment according to the polarity of the content.

The lexicon-based approach generally relies on a dictionary of opinion words, also known as a sentiment dictionary or a sentiment lexicon, to identify and determine sentiment orientation as positive or negative. A standard lexicon like the Linguistic Inquiry and Word Count (LIWC) includes such a sentiment dictionary. The compilation of a sentiment lexicon needs to be done manually requiring considerable effort and time. While different bodies of sentiment lexicon can be created for specific subject matters, sentiment words included in most lexicon-based analysis tools are not specific to a particular topic (Godbole *et al.*, 2007). Like most lexicon-based methods, LIWC2015 typically analyses common words included in its dictionaries. Misspellings, colloquialisms, foreign words and abbreviations are usually not in the dictionaries. Although LIWC2015 includes a few words frequently used in social media and text messaging (e.g. lol, 4ever, b4) and very basic punctuation-based emoticons such as :) and ;), it does not support emojis and emoticons widely used on social media. Furthermore, the drawback of using the lexicon-based approach to sentiment analysis is that the polarity classification could vary across different domains. For example, the adjective "unpredictable" can have a positive orientation in a movie review but a negative orientation for a car's steering abilities (Turney, 2002).

The machine learning approach uses a fraction of the full data as a manually classified training data set and trains classifiers to learn by examples, thus "supervising" the classification and without relying on any prior lexicon. This approach typically trains sentiment classifiers using features such as unigrams or bigrams (Pang *et al.*, 2002) by applying different learning techniques such as Naive Bayes, maximum entropy or support vector machines. While machine learning methods that use training data sets for automating data classification are advantageous, these methods still require manual labelling of training examples, which size and quality affect the performance of the trained model. High-quality labelling of a large training data set can be time-consuming, whereas limiting the size of the training data set leads to poorer classification accuracy. Furthermore, the sampling of the training data set can have a significant impact on the performance of the trained model, depending on how many domains are represented.

The choice of which approach to use is crucial, as it impacts the accuracy of the sentiment classification and needs to be carefully aligned with the type of data being analysed (Chae, 2015). In general, using lexicon-based approaches has been shown to be less effective than machine learning models from training examples (Pang *et al.*, 2002). However, opting for machine learning and ignoring the lexical knowledge in lieu of training data may not be optimal. Several attempts to combine the two approaches have been conducted and reported in the literature, as illustrated in Table I. These studies mainly use lexicon-based sentiment classification to label data and then use that labelled data as a training data set to train a machine learning model (e.g. Sommar and Wielondek, 2015; Mudinas *et al.*, 2012; Liu *et al.*, 2011; Tan *et al.*, 2008). Combining lexicon-based and machine learning approaches in such a way avoids having to manually classify data for training purposes. While successful at improving the classification accuracy compared to lexicon-based only, these combined approaches still do not outperform machine learning approaches trained with manually classified data (Sommar and Wielondek, 2015; Mudinas *et al.*, 2012). Other attempts at combining sentiment analysis approaches (Prabowo and Thelwall, 2009) try multiple sentiment classifiers in sequence until one of them is successful at classifying sentiment either positive or negative. However, such approaches assume that sentiment classification has a binary outcome even when the content conveys both positive and negative sentiment.

Comparative evaluation of automated sentiment analysis methods

In the present research, we compare lexicon-based and machine learning approaches to automated sentiment analysis. We aim to provide evidence of any performance difference between the two approaches and to offer empirically sound guidance as to which of the two approaches is best suited to the analysis of positive or negative valence in social media conversations. We then propose a combined approach, leveraging both lexicon-based knowledge and manually labelled data as a training data set and demonstrate the superior performance of the combined approach when applied to consumer generated conversations.

Data collection and sampling

Given the emotional value consumers attach to the fashion industry, we consider luxury fashion brands as an appropriate context for the current study (So *et al.*, 2013). The Fashion 2015 Digital IQ Index® from L2 Inc was the source used to select a sample of 83 luxury fashion brands highly active on Facebook social media platform. Facebook Graph API was used to collect all posts published and their associated CGC in the form of comments on brand posts. Nine months' worth of data were collected, and the most relevant comments on each post, also called "top comments", were identified using the comment ranking algorithm introduced by Facebook back in 2012.

Top comments are crucial as they reflect not only the most meaningful comments but also the most viewed comments. Indeed, top comments are always visible under a post which means top comments are the most likely to have an impact on other consumers and thus play a role in stimulating sentiment laden brand conversations. They are also the most likely to require content analysis for marketers to gain insights into consumers' feelings, thoughts and opinions. A random sample of 850 top comments was manually classified as positive, negative or neither positive nor negative. The same sample was then classified automatically using lexicon-based and machine learning approaches and compared to the manual classification to assess their accuracy.

Sentiment classification

Lexicon-based approach to sentiment analysis

LIWC2015 (Pennebaker *et al.*, 2015), a text mining software, was used to conduct a lexicon-based sentiment analysis of the data sample. LIWC enables a computerised analysis of the word used within a text and calculates the percentage of usage of sets of words that define different linguistic categories, generating an output measure for each of these categories. Among those categories, LIWC supports a sentiment lexicon for positive and negative sentiments. LIWC has been widely used in psychology and linguistics (Tausczik and Pennebaker, 2010). For each sentiment polarity, the software calculates the relative frequency with which words related to that polarity occur in a given text sample. For example, the words "love",

Journal of Consumer Marketing
Volume 34 · Number 6 · 2017 · 480–488

Table I Previous studies combining lexicon-based and machine learning approaches to sentiment analysis

Authors	Data	Approach	Outcome
Sommar and Wielondek (2015)	Movie reviews	Use the outcome of lexicon-based classification to feed machine learning for improved performance and convenience in sentiment classification	Combined approach outperforms the lexicon-based approach, in turn being outperformed by the learning based approach
Mudinas *et al.* (2012)	Software and movie reviews	Lexicon-based output is used to train a learning-based classifier	Hybrid approach improves the accuracy of sentiment classification compared to lexicon only approach, but is less accurate than learning based methods only
Liu *et al.* (2011)	Tweets	A classifier is trained using data given by the lexicon-based approach, instead of being labeled manually	Combined approach improves recall compared to lexicon-based approach only
Prabowo and Thelwall (2009)	Movie reviews, Product reviews, MySpace comments	Multiple sentiment classifiers are used in sequence so that if one classifier fails to classify a document, the classifier will pass the document onto the next classifier, until the document is classified or no other classifier exists	The use of multiple classifiers in a sequential manner can result in better effectiveness than any individual classifier. However, documents were assigned to one sentiment only (binary classification), so that a document containing both conveying both positive and negative sentiment, was necessarily classified as either positive or negative
Tan *et al.* (2008)	Movie Reviews, Computer Reviews, Education Reviews, and House Reviews	Use a lexicon-based technique to label data; then learn a new supervised classifier based on the labeled data	The experimental results indicate that proposed scheme could dramatically outperform "learn based" and "lexicon-based" techniques

"nice" or "sweet" are counted as representatives of positive sentiment, whereas the words "hurt", "ugly" and "nasty" are counted as representatives of negative sentiment.

Machine learning method for sentiment analysis
RTextTools is a machine learning package in R for automatic text classification. The package includes several algorithms for ensemble classification including maximum entropy, random forests, SVM, bagging, decision tree, etc. The objective of using a machine learning technique is to train classifiers from examples to perform the category assignments automatically. Because categories may overlap, each category is treated as a separate binary classification problem and content can belong to several categories simultaneously. This is commonly known as a supervised learning problem.

Half of the manually classified 850 top comments were used as a training data set and the other half were reserved as a testing data set, as it is recommended to use two different data sets for training and testing purposes. All machine learning algorithms supported by RTextTools R package were used to train models using the training data set and test them using the testing data set. For each supervised learning algorithms, the training data set was fed into the algorithm to train and test two classifiers, one for positive sentiment and one for negative sentiment. Each distinct word, emoji or emoticon corresponds to a feature, with the number of times a feature occurs in the document as its value. The resulting representation scheme, generated by the RTextTools package in R, is a term matrix of 221 terms from the training data set such as not only "cute", "elegant", "horrible", etc. but also emojis and emoticons. Each of the trained classifiers uses a subset of those terms, automatically selected and weighted by the corresponding supervised learning algorithm.

The best performing classifiers were obtained using maximum entropy modelling for predicting positive sentiment and the bagging method for predicting negative sentiment. Maximum entropy modelling, or Maxent, uses a low-memory multinomial logistic regression with support for semi-automated text classification (Jurka, 2012). In the bagging classification approach (Breiman, 1996), each tree is constructed from a bootstrap (Efron and Tibshirani, 1994) sample drawn with replacement from the training data set. Maxent and bagging have been successfully applied to NLP (Charniak, 1996) and are suitable for text categorisation such as consumer comments on Facebook brand posts. In the remainder of the paper, these two top-performing machine-learning algorithms, Maxent for positive sentiment classification and Bagging for negative sentiment classification, are referred to as the machine learning approach.

Performance measures

We evaluated the performance of the two sentiment analysis approaches using a standard performance measure from the information retrieval literature (Van Rijsbergen, 1979; Sebastiani, 2002). Using the testing data set of manually pre-classified CGC, along with the automated classification of the same data set, we constructed two-by-two contingency tables of the counts of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). On the one hand, true positives (true negatives) are the number of instances in which CGCs were accurately classified as positive (negative) by automated methods, using the manual classification as reference (correct classification). On the other hand, false positives (false negatives) are the number of instances in which CGCs were inaccurately classified as positive (negative) by automated methods. Note that content not classified as positive is not necessarily classified as negative. In fact, the same content can possibly be classified as both positive and negative if it refers to both valences at the same time.

To measure the performance of each approach, we used two commonly adopted measures of classification effectiveness, namely, Precision $p = TP/(TP + FP)$ and Recall $r = TP/(TP + FN)$. Precision (p) of an automated classification method also known as positive predictive value is the fraction of CGC for which automated and manual classifications match. A higher precision results from an automated classification that has a closer match with the manual classification. Recall (r) of an automated classification method, also known as sensitivity, is the proportion of positive (negative) CGCs that are manually classified as such and correctly classified by the automated method. A higher recall results from an automated classification method missing out on fewer positive (negative) CGCs, compared to manual classification.

There is an inherent trade-off for a sentiment classification method between precision and recall as higher recall can be achieved at the price of very low precision. To provide a more balanced assessment of the performance of sentiment classification methods, the F score measure is used. The F score combines recall and precision in a single quantity as a weighted average (Cohen and Singer 1999) and is used as a single performance indicator that is high if both precision and recall are high and low if either precision or recall are low. In this paper, the F score equally weights precision and recall and corresponds to the following formula:

$$F\ score = 2 \times precision \times recall/(precision + recall)$$

The F score is bound between 0 and 1 and can be interpreted as a probability. The closer the score is to 1, the better. The practical significance of the F score is that it represents a single measure of classification performance. A high F score means that the classification method achieves both high precision and high recall.

Another performance indicator considered in this study is the accuracy of classification which is calculated as the proportion of both true positives (TP) and true negatives (TN) in comparison to false positives (FP) and false negatives (FN):

$$Accuracy = (TP + FP)/(TP + FP + TN + FN)$$

A higher accuracy indicates that the sentiment analysis approach is better able to classify positive and negative valence of CGC.

Results

The results in Table II show that lexicon-based and machine learning approaches to sentiment analysis perform very similar in terms of F scores for positive valence classification (F = 0.77 and 0.78, respectively) as well as negative valence classification (F = 0.45 and 0.47, respectively). The results directly contradict prior research regarding the performance of machine learning classification methods claimed to be more accurate than lexicon-based approaches (Chaovalit and Zhou, 2005). The results also reveal that both approaches achieve higher accuracy when classifying positive valence than negative. The lower F scores for classifying negative valence, below 0.5, is explained by the well-recognised limitations of automated sentiment analysis methods when it comes to analysing sarcasm, which is often a limitation for manual approaches too (Maynard and Greenwood, 2014). These results indicate that the existing sentiment analysis methods are appropriate for predicting positive valence and limited for predicting negative valence when applied to social media conversations.

Proposal of a combined approach to sentiment analysis

While the performances of both approaches to sentiment analysis are similar, the two approaches do differ in some classifications. Numerous methods are available to compare results of classification methods and estimate the agreement among them. One of the simplest but most effective of these procedures is to examine the intersections of the resulting classifications using UpSet plots and Venn diagrams (Lex *et al.*, 2014). UpSet plots simplify the way intersections of multiple sets can be read using bar plots and are used to compare and contrast two or more sets in terms of the relationship between them. The relationship can be the intersection, union or complement. Figures 1 and 2 illustrate the relationship between the results of machine learning classification, lexicon-based classification and manual classification.

By matching manually classified comments and the automated classifications outcome, the lexicon-based and the machine-learning approaches agree on 63.6 per cent of correctly classified positive comments and 35.3 per cent of correctly classified negative comments. Figures 1 and 2 further show the combination of both approaches significantly increases the number of consumer comments correctly classified. This indicates that one approach is complementing the other and that a combination of the approaches may produce a better outcome.

In this paper, we propose that such a combination can be as simple as using both approaches and combining their results. Thus, the motivation here is not to avoid manually labelling a training data set but rather to combine the strengths of lexicon-based and machine learning approaches for better accuracy of the results. Table II shows the results of the combined approach. The F-score for classifying positive sentiment increases substantially, scoring 0.83, but remains relatively the same for classifying negative sentiment at around 0.46 when using a combined approach. By combining the two approaches the overall performance of sentiment classification is greatly improved for classifying positive sentiment without penalizing the performance of classifying negative sentiment. This finding indicates that a combined approach is particularly valuable when marketers require sentiment analysis to accurately identify positive word of mouth.

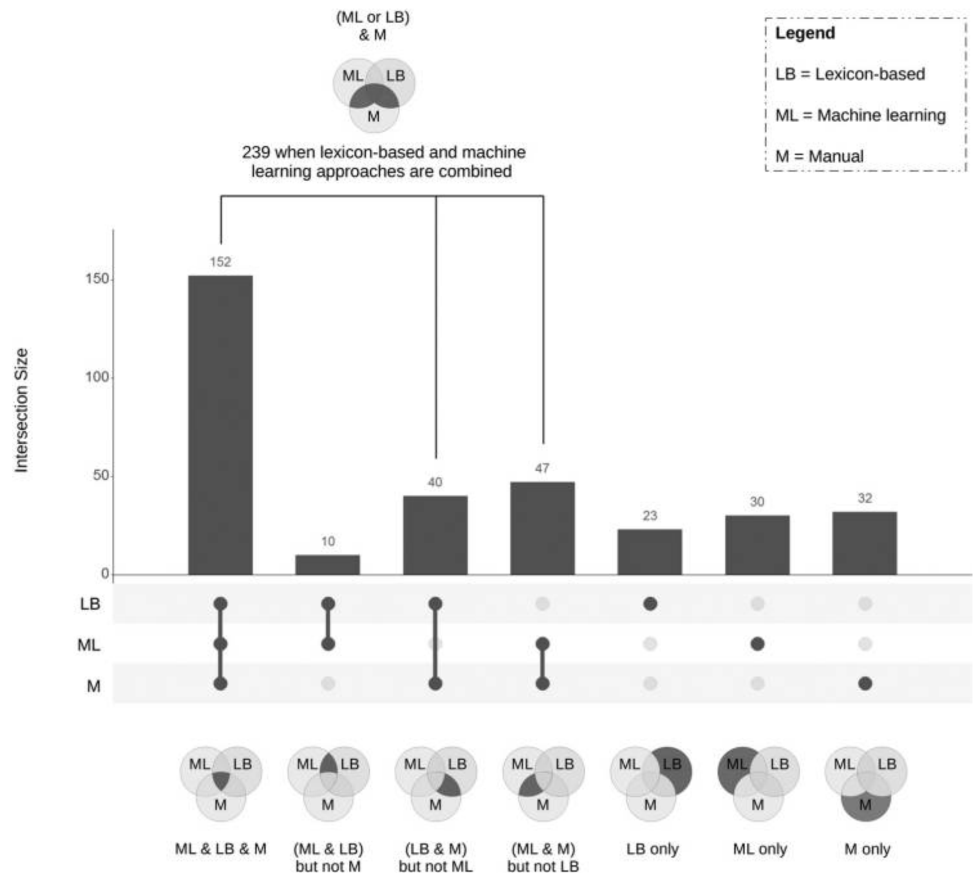
Conclusions, limitations and future research

This study makes several contributions. First, we empirically test two prominent sentiment analysis approaches, namely, lexicon-based and machine learning. The results indicate that,

Table II Evaluation of lexicon-based and machine learning approaches to sentiment analysis

Approach	Precision	Recall (Sensitivity)	F score	Accuracy	True positives (TP)	False positives (FP)	True negatives (TN)	False negatives (FN)
Evaluation of positive valence classification (total tested comments N = 425)								
Manual	1	1	1	1	271	0	154	0
Lexicon-based	0.85	0.71	0.77	0.74	192	33	121	79
Machine learning	0.83	0.73	0.78	0.74	199	40	114	72
Combined approach	0.79	0.88	0.83	0.78	239	63	91	32
Evaluation of negative valence classification (total tested comments N = 425)								
Manual	1	1	1	1	75	0	350	0
Lexicon-based	0.31	0.81	0.45	0.65	61	135	215	14
Machine learning	0.54	0.41	0.47	0.83	31	26	324	44
Combined approach	0.31	0.91	0.46	0.62	68	154	196	7

Figure 1 UpSet plot illustrating the agreement/disagreement among lexicon based, machine learning and manual sentiment analysis for the classification of positive valence

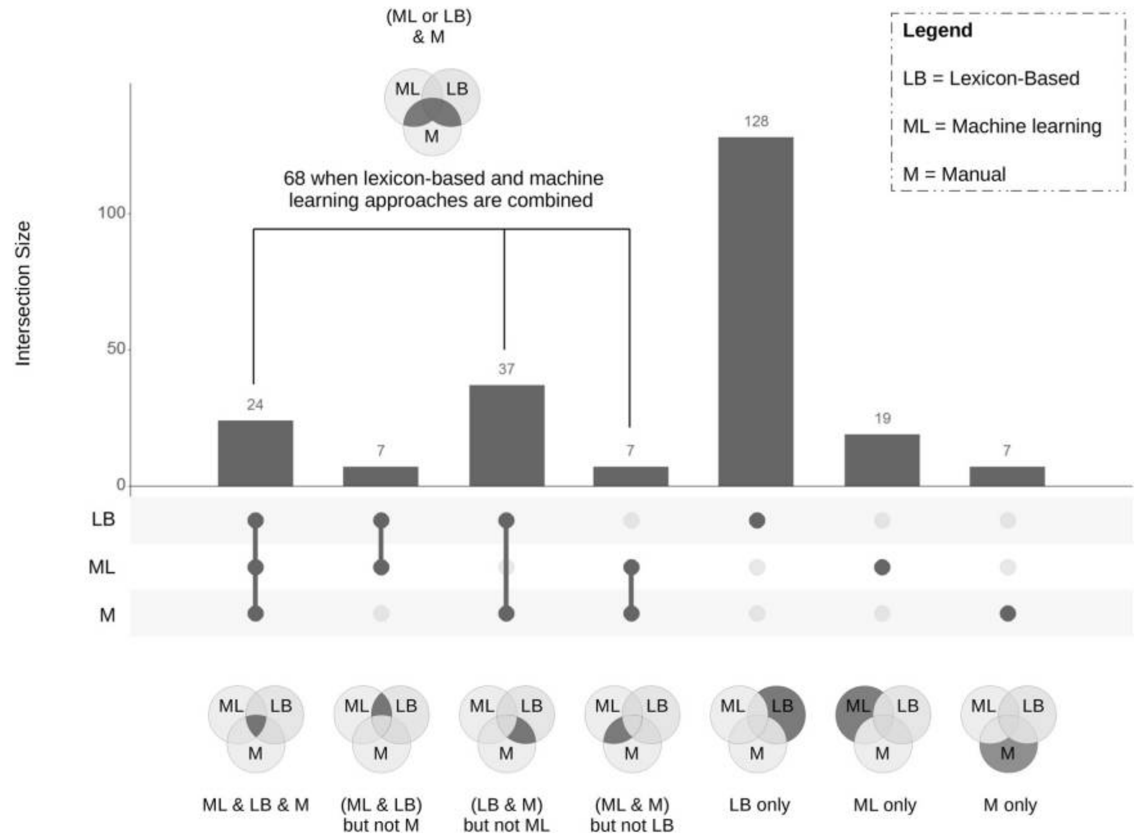


when applied on social media conversations, the two automated approaches have similar performance. Second, we demonstrate that combining the different approaches significantly improves classification performance in terms of precision and recall for positive sentiment. This finding suggests the great potential of a combined approach to gain deeper insights into positive social media conversations. Given the soaring volumes of brand-related social media conversations and the lack of guidance as to what tools are adequate to analyse such "Big Data", our study fills a gap in the literature and adds to industry best practices. Our findings form the basis of decision-making around which approach is best for marketers to analyse consumers' social media conversations and how to best combine approaches to achieve better outcome.

Sentiment analysis is only one way to explore online conversations with other analytic approaches available for knowledge discovery. For these reasons, further research is required to guide marketers on how to select and match the various text analysis approaches with the different social

Journal of Consumer Marketing
Volume 34 · Number 6 · 2017 · 480–488

Figure 2 UpSet plot illustrating the agreement/disagreement among lexicon based, machine learning and manual sentiment analysis for the classification of negative valence



media data sources to generate precise and accurate outcomes.

Among the variety of data analysis methods and techniques, the use of sentiment analysis for gauging public opinion is increasingly growing. Marketers tend to apply these methods without adequate evaluation of their effectiveness at classifying the sentiment valence of certain social media data sources, such as conversational data in the form of comments or tweets. In this paper, we have investigated the fit between the two main sentiment analysis approaches using conversational social media data consisting of Facebook consumer conversations.

Results from combining the two approaches are quite promising for positive sentiment analysis, but further research is required to improve the accuracy of negative sentiment analysis. To extend our study results, the combined approach needs to be applied to other kind of conversational data such as tweets and microblogs.

Although the fields of NLP, computational linguistics, and text analytics continue to mature, they arguably remain unable to match the ability of humans to take subtle aspects of the context into account and make fine distinctions when interpreting the content data (Conway, 2006), as empirically verified in this paper by the relatively low levels of accuracy for negative sentiments. Furthermore, this study, and most prior studies on sentiment analysis, are limited to the assessment of automated sentiment analysis applied to text only. It would be interesting to extend the study to other types of content such as images and videos using visual classification methods.

References

Bolat, E. and O'Sullivan, H. (2017), "Radicalising the marketing of higher education: learning from student-generated social media data", *Journal of Marketing Management*, Advance online publication, doi: 10.1080/0267257X.2017.1328458.

Breiman, L. (1996), "Bagging predictors", *Machine Learning*, Vol. 24 No. 2, pp. 123-140.

Brown, D., Taylor, C., Baldy, R., Edwards, G. and Oppenheimer, E. (1990), "Computers and QDA - can they help it? A report on a qualitative data analysis programme", *The Sociological Review*, Vol. 38 No. 1, pp. 134-150.

Cambria, E., Schuller, B., Xia, Y. and Havasi, C. (2013), "New avenues in opinion mining and sentiment analysis", *IEEE Intelligent Systems*, Vol. 28 No. 2, pp. 15-21.

Canhoto, A.I. and Padmanabhan, Y. (2015), "'We (don't) know how you feel' - a comparative study of automated vs manual analysis of social media conversations", *Journal of Marketing Management*, Vol. 31 Nos 9/10, pp. 1141-1157.

Chae, B.K. (2015), "Insights from hashtag# supplychain and Twitter analytics: considering Twitter and Twitter data for supply chain practice and research", *International Journal of Production Economics*, Vol. 165, pp. 247-259.

Chaovalit, P. and Zhou, L. (2005), "Movie review mining: a comparison between supervised and unsupervised classification approaches", in *System Sciences Proceedings of the 38th Annual Hawaii International Conference HICSS'05 in Hawaii, USA, IEEE, Washington, DC*, p. 112c.

Charniak, E. (1996), *Statistical Language Learning*, MIT Press, Cambridge, MA.

Chen, H. and Zimbra, D. (2010), "AI and opinion mining", *IEEE Intelligent Systems*, Vol. 25 No. 3, pp. 74-80.

Cohen, W.W. and Singer, Y. (1999), "Context-sensitive learning methods for text categorization", *ACM Transactions on Information Systems (TOIS*), Vol. 17 No. 2, pp. 141-173.

Constantinides, E., Romero, C.L. and Boria, M.A.G. (2008), "Social media: a new frontier for retailers?", in Swoboda, B., Morschett, D., Rudolph, T., Schnedlitz, P. and Schramm-Klein, H. (Eds), *European Retail Research*, Gabler Verlag, Wiesbaden, pp. 1-28.

Conway, M. (2006), "The subjective precision of computers: a methodological comparison with human coding in content analysis", *Journalism & Mass Communication Quarterly*, Vol. 83 No. 1, pp. 186-200.

Cooke, M. and Buckley, N. (2008), "Web 2.0, social networks and the future of market research", *International Journal of Market Research*, Vol. 50 No. 2, pp. 267-292.

Efron, B. and Tibshirani, R.J. (1994), *An Introduction to the Bootstrap*, CRC Press, Boca Raton, FL.

Gensler, S., Völckner, F., Liu-Thompkins, Y. and Wiertz, C. (2013), "Managing brands in the social media environment", *Journal of Interactive Marketing*, Vol. 27 No. 4, pp. 242-256.

Godbole, N., Srinivasaiah, M. and Skiena, S. (2007), "Large-scale sentiment analysis for news and blogs", in *Proceedings of the International Conference on Weblogs and Social Media ICWSM, Boulder, CO*, pp. 219-222.

Hemann, C. and Burbary, K. (2013), *Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World*, Pearson Education.

Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009), "Twitter power: tweets as electronic word of mouth", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 11, pp. 2169-2188.

Jurka, T. (2012), "MAXENT: an R package for low-memory multinomial logistic regression with support for semi-automated text classification", *The R Journal*, Vol. 4 No. 1, pp. 56-59.

Kiritchenko, S., Zhu, X. and Mohammad, S.M. (2014), "Sentiment analysis of short informal texts", *Journal of Artificial Intelligence Research*, Vol. 50, pp. 723-762.

Kivran-Swaine, F., Brody, S., Diakopoulos, N. and Naaman, M. (2012), "Of joy and gender: emotional expression in online social networks", in *Proceedings of the Conference on Computer Supported Cooperative Work Companion, ACM, New York, NY*, pp. 139-142.

Krippendorff, K. (2012), *Content Analysis: An Introduction to Its Methodology*, Sage, Thousand Oaks, CA.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. and Pfister, H. (2014), "UpSet: visualization of intersecting sets", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20 No. 12, pp. 1983-1992.

Liu, X., Zhang, S., Wei, F. and Zhou, M. (2011), "Recognizing named entities in tweets", in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, Association for Computational Linguistics, Stroudsburg, PA*, pp. 359-367.

Lusch, R.F., Vargo, S.L. and Tanniru, M. (2010), "Service, value networks and learning", *Journal of the Academy of Marketing Science*, Vol. 38 No. 1, pp. 19-31.

Maynard, D. and Greenwood, M.A. (2014), "Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis", in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 4238-4243.

Mohammad, S.M., Zhu, X., Kiritchenko, S. and Martin, J. (2015), "Sentiment, emotion, purpose, and style in electoral tweets", *Information Processing & Management*, Vol. 51 No. 4, pp. 480-499.

Mudinas, A., Zhang, D. and Levene, M. (2012), "Combining lexicon and learning based approaches for concept-level sentiment analysis", in *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York, NY*.

Murdough, C. (2009), "Social media measurement: It's not impossible", *Journal of Interactive Advertising*, Vol. 10 No. 1, pp. 94-99.

Nabi, R.L. (2007), "Emotion and persuasion: a social cognitive perspective", in Roskos-Ewoldsen, D.R. and Monahan, J.L. (Eds), *Communication and Social Cognition: Theories and Methods*, Routledge, Mahwah, NJ, pp. 377-398.

Nabi, R.L. and Oliver, M.B. (2009), *The SAGE Handbook of Media Processes and Effects*, Sage, Thousand Oaks, CA.

Pang, B., Lee, L. and Vaithyanathan, S. (2002), "Thumbs up? Sentiment classification using machine learning techniques", in *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing-Volume 10, Association for Computational Linguistics, Stroudsburg, PA*, pp. 79-86.

Pathak, X. and Pathak-Shelat, M. (2017), "Sentiment analysis of virtual brand communities for effective tribal marketing", *Journal of Research in Interactive Marketing*, Vol. 11 No. 1, pp. 16-38.

Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K. (2015), *The Development and Psychometric Properties of LIWC2015*, University of Texas at Austin, Austin, TX.

Prabowo, R. and Thelwall, M. (2009), "Sentiment analysis: a combined approach", *Journal of Informetrics*, Vol. 3 No. 2, pp. 143-157.

Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S.M., Ritter, A. and Stoyanov, V. (2015), "Semeval-2015 task 10: sentiment analysis in twitter", in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 451-463.

Sebastiani, F. (2002), "Machine learning in automated text categorization", *ACM Computing Surveys (CSUR*), Vol. 34 No. 1, pp. 1-47.

So, J.T., Grant Parsons, A. and Yap, S.F. (2013), "Corporate branding, emotional attachment and brand loyalty: the case of luxury fashion branding", *Journal of Fashion Marketing*

and Management: An International Journal, Vol. 17 No. 4, pp. 403-423.
Sommar, F. and Wielondek, M. (2015), *Combining Lexicon-and Learning-based Approaches for Improved Performance and Convenience in Sentiment Classification*, CSC, KTH, Stockholm, Sweden.
Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011), "Lexicon-based methods for sentiment analysis", *Computational Linguistics*, Vol. 37 No. 2, pp. 267-307.
Tan, S., Wang, Y. and Cheng, X. (2008), "Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples", in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY*, pp. 743-744.
Tausczik, Y.R. and Pennebaker, J.W. (2010), "The psychological meaning of words: LIWC and computerized text analysis methods", *Journal of Language and Social Psychology*, Vol. 29 No. 1, pp. 24-54.
Tirunillai, S. and Tellis, G.J. (2014), "Mining marketing meaning from online chatter: strategic brand analysis of big data using latent dirichlet allocation", *Journal of Marketing Research*, Vol. 51 No. 4, pp. 463-479.
Turney, P.D. (2002), "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Stroudsburg, PA*, pp. 417-424.
Van Rijsbergen, C.J. (1979), *Information Retrievals*, Butterworth, London.
Wang, W., Chen, L., Thirunarayan, K. and Sheth, A.P. (2012), "Harnessing twitter 'Big Data' for automatic emotion identification", in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), IEEE, Washington, DC*, pp. 587-592.

Corresponding author

Chedia Dhaoui can be contacted at: chedia.dhaoui@mq.edu.au