Sentiment Analysis for Small and Big Data

In: The SAGE Handbook of Online Research Methods

By: Mike Thelwall
Edited by: Nigel G. Fielding, Raymond M. Lee & Grant Blank
Pub. Date: 2017
Access Date: August 29, 2019
Publishing Company: SAGE Publications Ltd
City: 55 City Road
Print ISBN: 9781473918788
Online ISBN: 9781473957992
DOI: https://dx.doi.org/10.4135/9781473957992
Print pages: 344-360

Sentiment Analysis for Small and Big Data

Mike Thelwall

Introduction

There are now effective methods to automatically detect sentiment in text and these are capable of human-like levels of performance in some contexts. These programs have been used primarily to analyse online texts, both for research and commercial applications, and are valuable to help gain insights into public opinions about the topics, products and issues discussed online. Automated sentiment analysis naturally has Big Data applications because it allows huge amounts of text to be processed rapidly, enabling sentiment-related insights to be gained about issues that might not otherwise be detectable with small amounts of data.

Opinion mining is concerned with developing software to automatically extract user opinions about products or other entities from text, typically from online sources (Cambria *et al.*, 2013; Feldman, 2013; Liu, 2012; Liu and Zhang, 2012; Pang and Lee, 2008). This is also sometimes called sentiment analysis but the latter term also encompasses programs that extract sentiments from text for other purposes, such as to estimate the affective state of the text author. The remainder of this chapter uses the latter term. There are several different core tasks for sentiment analysis and they are sometimes carried out by separate systems and sometimes by a single system, with the user selecting the desired type of output. In the following, a text could be an entire document, a sentence or a part of a sentence, with or without embedded metadata.

- *Subjectivity detection*: texts are classified as containing expressions of sentiment (subjective) or not (objective).
- *Polarity detection*: texts are classified as positive or negative overall.
- *Sentiment strength detection*: texts are classified for the overall strength of positive and/or negative sentiment, or for the overall strength of sentiment and its polarity.
- *Emotion detection*: texts are classified for the predominant emotion (e.g. unhappy, angry), perhaps in addition to its strength, or the degree to which a fixed number of different emotions are evident.
- *Aspect-based sentiment analysis*: texts are dissected to identify the aspects of a product that are discussed and the sentiments expressed about these aspects.

Another variant is concept-based sentiment analysis (e.g. Poria *et al.*, 2014), which is similar to aspect-based sentiment analysis except that there is an explicit focus on using semantic resources to identify sentiment about concepts rather than about individual nouns, although the end result may be similar if similar nouns are clustered rather than analysed semantically.

Applications of sentiment analysis vary from direct, such as detecting whether sentiments towards unhealthy food vary by geographic area (Widener and Li, 2014), to indirect applications that enhance software designed for other tasks. For example, it seems reasonable to believe that a sentiment analysis capability would enhance the ability for autonomous agents and robots to interact effectively with humans (Mavridis, 2015), and sentiment analysis has already been embedded into web crawlers (Vural *et al.*, 2014) and automatic chat systems (Skowron *et al.*, 2011), but the most widespread application is to aid social media monitoring by companies and large organisations.

Sentiment analysis algorithms are typically evaluated by comparing their outputs (e.g. a decision about whether a text is positive, negative or neutral) with the judgements of human coders for the same set of at least 1,000 texts. These judgements must be made by people following precise and consistent instructions (e.g. Wiebe *et al.*, 2005) or in some cases the texts may have opinions registered by their authors, such as for reviews that are accompanied by an overall score (Turney, 2002). Depending on the precise outputs of the sentiment analysis system (e.g. sentiment polarities or strengths) a system's score might be the percentage of texts given the same ratings as those of the human coders or a metric for the correlation between the human and computer scores. If more than one sentiment analysis method is compared, then the one with the highest score compared with the human coders would be assumed to be the best. The assumption here is that the human judges' scores are essentially correct and hence can form the gold standard against which all algorithms should be compared.

This chapter reviews the main different sentiment analysis methods, including both lexical and machine learning approaches, as well as the main tasks, such as polarity detection, sentiment strength detection and fine-grained emotion detection. It also covers important related tasks, such as the need to customise software designed for one type of text before it can be applied efficiently to another and to detect the target of any sentiment expression. The chapter also reviews sentiment analyses research applications involving either Big Data or small scale samples of online texts, showing the range of current applications as well as the potential to deploy the methods to investigate a wide range of issues. Most of this research focuses on the social sciences, and on issues for which public opinion data is relevant. Some of the research also analyses the affective component of online communication within the social web in contexts such as political debates and communication between friends, when sentiment forms an important component of the interactions.

Lexical Approaches

Lexical sentiment analysis algorithms centre on a pre-defined lexical resource, such as a list or network of sentiment words. Whereas some methods exploit this resource through simple matching in text, other algorithms exploit a range of natural language processing techniques in an attempt to leverage more information from the text.

Simple Lexical Methods

A simple lexical sentiment analysis method might start with a list of positive and negative words and then either count how often they occur in a given text or apply a formula, such as a weighted sum, in order to categorise the text as positive, negative or neutral. Most algorithms go further than this, however, and include additional processing steps, such as to detect negation or to recognise emoticons. Unlike other strategies, lexical methods tend to be flexible and do not need a set of human-coded training texts for each topic area in order to work but need a human-coded lexicon instead (e.g. Tong, 2001). Such human-coded lexicons are readily available for some languages, such as ANEW for English (Nielsen, 2011).

SentiStrength (http://sentistrength.wlv.ac.uk) is a lexical sentiment strength detection program (Thelwall *et al.*, 2010). It has a manually curated list of 2,489 sentiment words and word stems. These derive from a combination of the psychology text analysis program LIWC (Tausczik and Pennebaker, 2010), the General Inquirer lexicon (Stone *et al.*, 1962) and manually identified additional terms, including neologisms and slang. The core action of SentiStrength is to process a sentence to identify all terms that are in its sentiment lexicon, looking up their polarity and strength. Each sentence is then assigned the strengths of its most positive and most negative term. For example, the text ‘tired but good day’ would score 3 out of 5 for positivity because good is in the lexicon with score 3 (1 indicates no positivity and 5 indicates very strong positivity). It would also score −2 on a scale of −1 (no negativity) to −5 (very strong negativity) for negative sentiment, so the system output would be (−2, 3). A sentence without any recognised sentiment terms would be assigned (−1, 1), indicating no positive sentiment and no negative sentiment. This method fails when typical term sentiments are modified by surrounding words and so there are additional rules for dealing with negation, questions and booster words (e.g. very). There are also rules for other ways of expressing sentiment, such as idioms, emoticons, emphatic spelling (e.g. Maaaaarieeee) and punctuation (e.g. hello!!!).

Despite the use of simple non-linguistic methods, SentiStrength has human-like accuracy on the short informal texts found in a wide range of different types of social web sites, including Twitter (Thelwall *et al.*, 2011). In other words, for some types of social web texts, its sentiment scores agree with human coder scores about as much as the human coder scores agree with each other. SentiStrength's simple approach also allows it to be fast (14,000 Tweets per second) and flexible – its lexicon can be manually customised and there are versions in many different languages.

Sentiment analysis algorithms can also exploit linguistic information by identifying the relationships between different segments of text within a sentence. For example, the word *but* in the middle of a sentence suggests that one of the two halves of the sentence is positive and the other is negative. If an algorithm can identify the polarity of one of the two segments with some certainty then it can infer that the other one has the opposite polarity, and with enough examples can also deduce new sentiment-bearing terms for a lexicon in this way (Zhang and Singh, 2014).

Natural Language Processing

Linguistic methods extend the basic lexical approach by incorporating linguistic knowledge and resources. For example, instead of using a list of affective terms, resources such as WordNetAffect (Strapparava and Valitutti, 2004) or SentiWordNet (Esuli and Sebastiani, 2006) allow terms in text to be matched to their root word form (e.g. *go, going, went* all map to the verb *go*) and reveal semantic relationships between words (e.g. *better* is weaker than *best*). This extra information can be harnessed to make a more powerful algorithm, but may make the resulting algorithm less flexible and substantially slower.

Adjectives can be given special treatment in sentiment analysis if they can be identified through linguistic methods. This is useful because they express sentiment more frequently than other types of word. This can be achieved with the natural language processing technique of Part Of Speech (POS) tagging. For this, a POS tagger application uses a set of learned heuristics to tag each word in a text with its part of speech. For example, this might convert 'beautiful shoes' to 'beautiful_JJ shoes_NNS' where JJ is the POS tag for adjective and NNS is for plural nouns. An algorithm could use this information to extract all adjectives or, if context is needed for the adjectives, specific POS patterns involving adjectives, such as all consecutive words where the first is an adjective and the second is a type of noun.

SO-CAL is a linguistic lexical algorithm for sentiment strength detection in English and Spanish (Taboada *et al.*, 2011; Taboada and Grieve, 2004). It uses dictionaries of adjectives, lemmatised (i.e. converted to a standard form) nouns, adverbs and lemmatised verbs compiled from a variety of sources, each with a human-assigned single sentiment polarity and strength integer score between −5 and +5. Some multi-word expressions are also included. SO-CAL includes rules for dealing with negation, intensifiers (e.g. extremely) and irrealis (connoting that a proposition is nonactual or nonfactual, e.g. would). SO-CAL classifies a document with the average of the sentiment expressions of its component parts, in conjunction with its additional rules mentioned earlier.

Automatic Word and Phrase Sentiment Association

Although some lexical methods exploit human constructed sentiment resources, knowledge-poor methods instead detect the sentiment associations of words and phrases automatically. This has the advantage that it can adapt to cope with domain-specific terms that are common in a set of texts from one source but are otherwise too rare to be incorporated into a general sentiment analysis program. It can also work for phrases rather than individual words. For example, a program processing a set of TV reviews might identify that the phrase 'large screen' occurred frequently and seemed to express sentiment, triggering a method to detect whether it is usually positive or negative.

Pointwise Mutual Information (PMI) is a commonly used metric for assessing whether a term or set of terms is likely to have a positive or negative connotation (Turney, 2002). It is useful when they have been automatically extracted by a heuristic that does not include polarity information. For example, a simple rule might extract all instances of adjectives followed by nouns, hoping that many will indicate sentiment (e.g. large screen, beautiful colour). The PMI formula used to assess polarity is

$$PMI(s,t) = P(s \wedge t)/P(s)P(t)$$

where P(s) and P(t) are the probabilities of s and t occurring in a text in a given corpus, respectively, and

P(s∧t) is the probability that they both occur in the same text. If s and t occur independently of each other, then P(s∧t) = P(s)P(t) but if they tend to co-occur then P(s∧t) > P(s)P(t) and so PMI(s,t) > 1. This can be useful to estimate the polarity of a new term or set of terms if s has a known polarity. For example, if PMI(‘good', ‘large screen') >1 then this suggests that ‘large screen’ tends to occur in the texts that contain ‘good’ and is perhaps more likely to be positive. More generally when comparing PMI(‘good',‘large screen') with PMI(‘bad',‘large screen'), if one is much larger than the other then ‘large screen’ is likely to be a reliable indicator of polarity. Although this is a simple example, PMI is simple, fast and flexible, with many uses in sentiment analysis.

The Turney (2002) algorithm exploits PMI by first processing a set of reviews to extract all phrases that obey any of a set of linguistic patterns (e.g. adjective followed by noun). For each phrase, PMI values are calculated for both ‘poor’ and ‘excellent', using a commercial search engine query to estimate the number of web pages matching the appropriate query. This effectively uses the web itself as a corpus and the estimated polarity of the phrase is generated by a formula based on the two PMI values. The estimated polarity of a document is the average of the polarities of all of the detected phrases.

A weakness of the Turney algorithm is that the use of the web as a corpus can cause problems with context-dependant words and phrases, such as *scary*, which is positive for horror movies.

Machine Learning

Sentiment detection can be treated as a text classification problem and tackled with generic machine learning methods. This works by converting each text into numerical vectors of features, which typically record the frequency of a list of words and short phrases (with or without linguistic classifications) in each document within a corpus. About 1,000 of these documents must have been pre-classified by human coders (i.e. the training set) and then the machine learning stage produces an algorithm that has learned how to predict the sentiment of the classified texts. There are many techniques that can be used for the machine learning stage, such as naïve Bayes, support vector machines and decision tree learning, many of which produce different types of prediction algorithms (e.g. a mathematical formula or a set of rules). The trained algorithm is then a sentiment analysis classifier that can be applied to unclassified texts to predict their sentiment (Pang *et al.*, 2002). Machine learning seems to work well for texts that are focused around a specific topic (e.g. movie reviews) but classifiers trained for one topic can perform poorly on others (Aue and Gamon, 2005).

A key stage in machine learning is the conversion of texts into feature vectors, which are sets of relevant terms or properties. Without an appropriate choice of features, no machine learning algorithm will work well. These features need to capture the essence of the way in which sentiment is expressed in a text whilst ignoring all extraneous information. This latter part is important because unnecessary information can confuse the machine learning training stage and substantially weaken the final algorithm. Most systems extract either individual words for features or all phrases with 1–3 words in conjunction with a frequency threshold to exclude rare ones. The basic approach of making a simple frequency vector for the words or phrases can be improved by giving higher weightings to those that are relatively rare in the corpus and by taking into account the length of each document (Paltoglou and Thelwall, 2010).

For social science applications, an important limitation of machine learning is that it can introduce systematic sources of bias. For example, if specific issues within a news corpus tend to attract strong negative views (e.g. Israel–Palestine conflicts), then phrases associated with the issue can be picked as effective indicators of negativity. The result would be a trained algorithm that tended to classify texts about controversial issues as negative, rather than detecting negative sentiment about them (Thelwall *et al.*, 2012).

Sarcasm Detection

The accuracy of sentiment analysis algorithms can be reduced by the presence of significant amounts of sarcasm (Thelwall *et al*, 2012). This is because sarcastic texts often contain an expression of sentiment that is intended to be interpreted with the opposite polarity (e.g. ‘I am extremely happy to be injured'), and also because sarcasm is associated with negative reviews (Filatova, 2012). In theory, the impact of sarcasm could be reduced if it is detected in text, but detection is difficult and dependent on both the topic and the language of a text (e.g. Burgers *et al.*, 2012). For example, political sarcasm in Portuguese is often accompanied by the diminutive form of the name of a politician (Carvalho *et al.*, 2009), but this linguistic style seems to occur in few other languages.

Successful sarcasm detection relies upon the sarcasm containing standard phrases or phrase patterns that are common in sarcastic texts but rare otherwise (e.g. Justo *et al.*, 2014). Such phrases are topic-dependent, which makes the construction of a general purpose sarcasm detector difficult. For example, a common type of book review sarcasm is praise for the cover rather than the content of a volume (Davidov *et al.*, 2010). However, a promising new approach has successfully detected explicit sarcasm in the sense of Tweets containing the #sarcasm hashtag. It exploits linguistic styles associated with figurative language, such as the use of rare words or unusual synonyms (Barbieri *et al.*, 2014).

Language Issues

Most of the studies reviewed in this chapter have analysed English texts. Whilst the principles are similar for most languages, it is not straightforward to start with a method that has been shown to work in one language and apply it to a different language. The first problem is that there are many more resources for English, such as lists of sentiment terms and part of speech taggers, than for any other language and some languages have very few language resources of any type. In addition, the way in which sentiment is expressed varies substantially between languages. In Chinese the phrase *not good* is equivalent to *bad* in English but in English the negating term weakens the sentiment as well as inverting it, but not in Chinese.

Perhaps the most fundamental difference between languages is that sentences in some languages, such as

Chinese, do not have markers between words. There are two main solutions to this issue: either apply a word segmenting algorithm to artificially add spaces between inferred words (Huang *et al.*, 2014) or use an n-gram method that looks for patterns of characters, irrespective of whether they are part of the same word or not (Zagibalov & Carroll, 2008). A more minor problem is that some languages glue words together to change their meaning. For example, a Turkish word can be negated by adding a negating ending to it and so a sentiment analysis algorithm may need to first separate such words and then process the sentiment negating parts of them separately (Vural *et al.*, 2013).

Additional Sentiment Analysis Tasks

Domain Transfer

Although some sentiment analysis programs are general purpose, most are designed for one type of product review. The domain transfer problem is the task of efficiently generating a sentiment analysis system for a new domain, such as reviews of a new product by re-using existing systems rather than building a completely new system (Blitzer *et al.*, 2006; Melville *et al.*, 2009). Here the term ‘domain’ is used to refer to the theme of the documents analysed. The most time-consuming part of making a new system is often the generation of a large and reliable corpus of human-coded texts for the new domain. Existing systems are likely not to perform optimally on a new domain because different features will be discussed and different ways of expressing sentiment may be used. As a result, they need to be adapted for the new (target) domain. When developing a system for a new domain, it can be important to choose the most similar domain with a classified system to start from because a classifier built from a highly dissimilar domain will not work well on the new domain (Ponomareva and Thelwall, 2012). Domain transfer approaches include training only on classified texts from the source domain that are reasonably similar to texts in the target domain, including a small amount of classified texts from the target domain in the training set, attempting to identify features (e.g. words) in the source domain that correspond to features in the target domain, and generating an ensemble of classifiers that are each trained on a different domain and combining their results (Aue and Gamon, 2005; Blitzer *et al.*, 2006).

An alternative strategy is to train a system on multiple different domains, but detect if each feature is domain-independent (Ida *et al.*, 2013). Presumably, general terms like *good* would be detected as domain-independent and more specific terms like *heavy* would be classified as domain-dependent. This approach enables a system to take advantage of additional data from other domains when training for each specific domain.

Although these strategies are all designed to deal with creating a system for a new specific domain based upon existing systems for different specific domains, a variant of the problem is to tailor a generic sentiment analysis system to be more effective on a specific domain. The general program SentiStrength can exploit a corpus of human coded texts for a specific topic to learn improved domain-specific sentiment polarities and strengths for words in its existing sentiment term lexicon. It can also learn new words to add to its lexicon by recognising those that often occur in texts that would otherwise be misclassified (Thelwall and Buckley, 2013).

Aspect-based Sentiment Analysis

Aspect-based (or feature-based) sentiment analysis is concerned with tying expressions of sentiment to specific aspects of an entity (typically a product) that are being discussed, even if multiple aspects are discussed within a single sentence. For example, if a comment reports ‘The décor was lovely but the portions were too small’ then it would be useful to pair décor with lovely and portions with too small. Here, both décor and portions are aspects of the restaurant entity that is being reviewed. An aspect-based sentiment analysis application might process a collection of reviews and then report the number of positive and negative comments about a list of aspects of the reviewed entity (Liu *et al.*, 2005). This challenging task typically includes automatically deciding which aspects are mentioned in a review and the sentiment orientation of the mention of the aspect, as well as resolving indirect references and synonyms so that different ways of mentioning the same aspect can be grouped together. As an example of an indirect reference, a review might state that a phone went for a long time before needing charging, and this could be recorded as a positive comment about the battery life aspect. Aspect-based sentiment analysis software typically involves natural language processing of the text in conjunction with pattern learning heuristics and linguistic resources in order to solve these problems. For example, association rule learning may be used to identify common connections between aspects and describing terms (Liu *et al*, 2005; Liu & Zhang, 2012).

Aspect-based sentiment analysis is most relevant for product reviews, where dense combinations of aspects and sentiments can be expected. It is less useful for microblogs, where there may not be enough space to discuss multiple aspects of a product.

Customisation for Specific Tasks

Although the texts in some sentiment analysis tasks are clearly self-contained units, such as product or movie reviews, in other cases, the texts may form a natural part of a set. In the latter case, information about the sentiment in other parts of the set may help to generate a better classification of each individual text. For example, in a dialog, it would be strange to see a sudden and isolated expression of positive or negative sentiment in an otherwise calm discussion. In response, some systems have attempted to use information about the classifications of sentiment in texts adjacent to the one being classified. The simplest approach is damping: reducing large deviations in sentiment strength on the basis that they are more likely to have been caused by classification errors than by a sudden sentiment change. This has been shown to work to some extent but the effectiveness of damping rules depends on the relationship between the posts (e.g. dialogs, monologs, multi-user interactions) and the nature of the damping to an extent that it is prohibitively time-consuming to implement in practice (Thelwall *et al.*, 2013).

A similar logical enhancement to sentiment analysis is to analyse all the comments made by a particular reviewer in order to help classify their reviews, and this is effective in many contexts in which the evidence is available (Basiri *et al.*, 2014).

Emotion Detection

Although most systems attempt to detect positive and negative sentiment, some go further by attempting to detect expressions of different types of emotion (Canales and Martínez-Barco, 2014). This is more difficult than polarity detection because it is harder to infer a fine-grained emotion unless it is explicitly described in a text. One study, for example, compared lexical and machine learning approaches for detecting the strength of anger, disgust, fear, joy, sadness and surprise (Strapparava and Mihalcea, 2008). The results suggested that the best method varies by overall objective but a linguistic lexical approach with both WordNet Affect and SentiWordNet performed well. A more detailed method can detect not just the emotion expressed but also the person that is apparently experiencing the emotion (Mohammad *et al.*, 2014).

Applications

Mining Product Reviews for Customer Opinion Information

Automatically extracting customer opinions is the main commercial application of sentiment analysis. An example of a simple but apparently useful system for product reviews is Opinion Observer, which automatically produces graphs of the main features in a set of products reviewed and the number of positive and negative comments about each feature. The potential purchaser of the products can quickly compare the graphs for the products that they are interested in and gain insights into what others believe to be their good and bad points (Liu *et al.*, 2005). This is similar to the Microsoft Pulse system that reads a collection of product reviews and creates an interactive tree map visualisation illustrating the main product-related clusters of terms, each within a rectangle proportional in size to the volume of related comments and colour-coded for sentiment. A large green rectangle containing the word ‘drive’ in a car tree map, for example, would indicate that many reviews discussed drive-related aspects of the car and they were typically negative about it. Clicking on the rectangle in the system would reveal individual drive-related reviews (Gamon *et al.*, 2005). An alternative (Google) approach is to cluster aspects of the reviews and to list the reviews in clusters (e.g. food, service, value), giving overall sentiment scores for the clusters as well as a polarity estimate for each individual review aspect (Blair-Goldensohn *et al.*, 2008).

Although there seems to be little concrete evidence of the value of consumer opinion information to organisations, it seems likely that it has wide value to large companies and for companies with products and services that people Tweet about. For example, one case study of the airline industry has shown that useful information can be extracted from social web sentiment analyses (Misopoulos *et al.*, 2014).

These ‘products’ can include services or anything else that people review or critically analyse online, such as healthcare problems in patient forums (Greaves *et al.*, 2013). Products can also be analysed by third parties with a vested interest, such as health workers analysing online discussions of tobacco-related projects with an agenda to reduce their use (Myslín *et al.*, 2013).

Other Commercial Intelligence Applications

In principle, the applications described earlier for product reviews could also be applied to any type of text containing evaluations, perhaps even if evaluation is not the primary purpose of the set of texts. One such application processes text within an organisation's internal social network in order to identify the themes discussed and their sentiments, as a management information tool (Subramanian *et al.*, 2013).

Organisations may also wish to track the flow of opinions over time for individual users or groups of users and there are some applications that can do this. OpinionFinder uses complex visualisations to illustrate changes in topics, sentiments and volume of discussions over time in order to give an understanding of how an opinion developed or changed (Wu *et al.*, 2012). This could help organisations to assess how a bad opinion about them emerged in the social web.

A Big Data style sentiment analysis application is to predict changes in prices or values of commodities, currencies or shares based upon relevant changes in sentiment extracted from the news or the social web. The belief behind this is that automatic methods may pick up small changes in market sentiment, perhaps even when they are not evident to experts. Thus, embedding sentiment analysis capabilities within online trading systems might improve their performance. There have been several attempts to build such systems, with some apparent success (Bollen *et al.*, 2011; Nassirtoussi *et al.*, 2015), but these are difficult to convincingly evaluate because a practical system would incorporate sentiment as a single component within a large range of indicators (e.g. Nassirtoussi *et al.*, 2014). The value of sentiment as an indicator is also likely to vary by market segment.

News and Blogs

News is a natural discussion topic for the social web and there have been several attempts to design sentiment analysis software for online news sites and blogs. One program detected the people that were discussed most positively and negatively in news stories and blog posts, finding substantial differences between the two (Godbole *et al.*, 2007).

Politics

Politics is a natural online discussion topic and this has been recognised by news media by monitoring sentiment in social media during elections (Wang *et al.*, 2012) or during key events, such as televised leaders' debates (e.g. Diakopoulos and Shamma, 2010). There are also some sentiment analysis programs designed specifically for political discussions (Van Atteveldt *et al.*, 2008; Young and Soroka, 2012) or with political adaptations (Vilares Calvo *et al.*, 2015). Several studies have attempted to assess political opinions or predict election outcomes using sentiment analysis in social media (Chung and Mustafaraj, 2011; O'Connor *et al.*, 2010; Tumasjan *et al.*, 2010), but this is difficult because the proportion of those online varies by political affiliation, including due to factors such as age and education level. In addition, more outspoken people are likely to be overrepresented online, and these may tend to associate with particular parties, such as those that are new or particularly radical. As a result, any serious attempt to predict election outcomes from Twitter would need to correct for a range of biases in order to be credible (Metaxas *et al.*, 2011). This is in addition to the problem of sarcasm, which makes texts that are part of political discussions particularly difficult to classify for sentiment (Bakliwal *et al.*, 2013; Thelwall *et al.*, 2012). Moreover, one study has shown that sentiment in Twitter can have little relationship with sentiment in the print news (Murthy and Petto, 2015), which casts further doubt on the value of Twitter as a reflector of public opinion. Nevertheless, the sentiment of political Tweets can give insights into the role of Twitter within political discussions (Vilares Calvo *et al.,* 2015).

A potential application of sentiment analysis would be to automatically discover the political affiliations of social media users, but this is difficult (Malouf and Mullen, 2008) and can be done by analysing the topics that they discuss without the need to harness sentiment as well (Conover *et al.*, 2011).

An interesting type of analysis for media is a minute-by-minute sentiment analysis of important political events, such as televised debates. Analysing sentiment in Twitter during the debates, for example, can point to topics within the debate that provoked the strongest reactions and the strongest positive or negative responses. Tying these results to the times when the different leaders were talking can also give insights into their performances (Diakopoulos and Shamma, 2010).

Big Data Social Web Investigations

Classic Big Data sentiment analysis applications process large volumes of text in order to identify sentiment-related patterns, even if too slight to be evident in lower volumes of data. One study analysed four million texts from blogs, Digg.com and online BBC discussion forums, looking for evidence that sentiments expressed by participants could be contagious in the sense of triggering similar sentiments from others. Partial evidence was found for this by analysing chains of consecutive texts with similar sentiments and showing mathematically that these chains were longer than if sentiment was expressed randomly (Chmiel *et al.*, 2011). This empirical evidence supports the common sense understanding that sentiments expressed in communication affect the tone of subsequent contributions. One reason why sentiment may spread in the social web is the influence of a small number of popular individuals (Bae and Lee, 2012; Zhao *et al.*, 2014).

There have been many attempts to harness sentiment in social media in order to predict changes in stock market or commodity prices on the basis that automatic methods might pick up slight changes in sentiment about a company or product before many human experts detect it. Some of these approaches appear to have had success but it seems that an effective system would incorporate sentiment as one component within a system that is reading multiple signals (Kazemian *et al.*, 2014).

Other studies have focused on the expressive style of individuals in the social web rather than on communication segments (i.e. consecutive texts) and have found that people tend to use similar levels of positivity and negativity in their social network comments to that of their friends (Bollen *et al.*, 2011; Thelwall, 2010). It is not clear whether this reflects happy people befriending each other and unhappy people befriending, happiness and sadness spreading between friends, or just friends having similar expressive styles within the social web (e.g. routinely sending cheerful messages or discussing a common interest in gothic rock).

On a huge scale, one study has used Tweets gathered from across the globe to analyse cultural and other factors that affect the relationship between sentiments expressed in Tweets and the time of day. They found, for example, that people seem to wake later at weekends and seem to express more happiness than during the week (Golder and Macy, 2011). An analysis of Facebook posts, in contrast, demonstrated a link between sentiments expressed and rainfall (Coviello *et al.*, 2014). Whilst these results are unsurprising, they show that is it now possible to conduct international studies of sentiment with millions of participants – without the need for extensive funding if the free Twitter API is used.

Tools for Sentiment Analysis

There are a number of options available for those wishing to apply sentiment analysis to their data. Most business users probably access sentiment analysis as a component within an online social media gathering and analytics service, such as Pulsar (pulsarplatform.com) or Topsy (topsy.com). These are not ideal for research because the algorithms used are typically not described by the service provider and hence operate as black box solutions, although some companies give broad information (www.lexalytics.com/technical-info/sentiment-analysis, accessed 13 July 2016). Some algorithms are also built into commercial analytics software, such as SPSS Text Analytics for Surveys and SAS Sentiment Analysis. To evaluate one of these systems, it would be useful to find out as much information as possible about how the system works and how it has been tested. A good system would presumably be kept up to date and backed by academic research, but in any case it is worth testing with at least 100 relevant texts (irrelevant texts could give misleading results) in order to discover how often the system gives reasonable answers.

Researchers wishing to know more about the sentiment analysis algorithm used, or even to create their own or modify an existing algorithm, have a number of options available. Some of the published sentiment analysis programs are available from their authors (online or via email) without charge for research, including SentiStrength (sentistrength.wlv.ac.uk: Windows and Java versions). Few seem to be open source, however, although they have an associated published article describing how they work.

Some resources are also available free online to help with building or testing sentiment analysis systems, such as the sentiment lexicons and human coded corpora of Maite Taboada (https://www.sfu.ca/～mtaboada/

research/SFU_Review_Corpus.html, accessed 13 July 2016) and Bing Liu (http://www.cs.uic.edu/～liub/FBS/sentiment-analysis.html, accessed 13 July 2016) and the linguistic resources of SentiWordNet (sentiwordnet.isti.cnr.it) and SenticNet (Cambria *et al.*, 2014). It is also possible to create sentiment analysis programs using machine learning methods with general purpose machine learning environments, such as Weka, and natural language processing toolkits, such as GATE (gate.ac.uk/sentiment). Natural language processing toolkits are available for Python (nltk.org), Java (nlp.stanford.edu/software), C++ (nlp.lsi.upc.edu/freeling), R (e.g. with the qdap package) and other languages.

Finally, it would also be reasonable to use content analysis software, such as NVivo, to help with human coding of sentiment in text, if an automatic method was not possible or desirable. Also worth a mention is QDA Miner by Provalis. It includes sentiment analysis and ships with several dictionaries that can be modified to suit the research purpose. These programs would be appropriate if there were relatively few texts to analyse or if the texts were difficult to automatically analyse due to the topic or the presence of sarcasm or figurative language.

Summary

Sentiment analysis, the automatic detection affective content in text, is a mature research area with a range of tools that can detect sentiment in different ways and for different purposes. Whilst lexical approaches tend to be generic in the sense of being designed to work across different topics and types of text, machine learning tends to need recalibrating on each different topic or text type, although they may perform better as a result. Depending on the system, the output may be an overall polarity judgement for each text, an indication of the strength of positivity and/or negativity, an indication of the strengths of various kinds of emotions, or a collection of information about the aspects discussed and their sentiments. Existing software often performs with a level of accuracy that is similar to that of humans classifying discrete texts, but is much faster and cheaper for large volumes of text.

The availability of effective sentiment analysis software has given rise to many commercial and research applications. In the commercial domain, it is now routine for companies to monitor the sentiment of important product and brand names in social media, perhaps as part of their wider social media monitoring activities. This is made possible by web intelligence companies that provide an easy web interface to access the data and data processing techniques. Nevertheless, there may be sinister overtones to some applications of sentiment analysis. It is now relatively easy for commercial and political organisations to monitor the sentiments of relevant groups of people that post text online and this may give the monitoring organisations the enhanced ability to get their message across – they may even attempt to manipulate groups more directly (Andrejevic, 2011). Moreover, this is also problematic if people are being directly or indirectly manipulated through information gained by sentiment analysis applied without their knowledge to text that they may have believed was private, and which was in any case written for another purpose (Kennedy, 2012).

Researchers have also found ways to combine social web data collection with sentiment analysis to gain insights into public opinions or reactions to news, health and financial market issues, either developing and assessing software for this or generating new understandings for the research. Despite the studies reviewed, it seems that there is still enormous potential to widen this basic approach to investigate issues of public concern that have not yet been investigated. For those opening up new research areas in this way, there are some important lessons from the research reviewed here to bear in mind. First, sentiment analysis software estimates sentiment but makes mistakes, even if it achieves human-level accuracy overall. It is likely to be particularly inaccurate for sets of texts where sarcasm is prevalent, probably including most online political discussions. Second, its performance varies by topic and text type, and so results should not be taken at face value and, where possible, evaluated and customised. Third, there may be systematic sources of bias in the results, especially if machine learning algorithms are used, and these should be checked whenever any trends are observed.

References

Andrejevic, M. (2011). ‘The work that affective economics does', *Cultural Studies*, 25(4–5), 604–20.

Aue, A. and **Gamon, M.** (2005). ‘Customizing sentiment classifiers to new domains: a case study'. In Proceedings of Recent Advances in Natural Language Processing (RANLP2005). http://research.microsoft.com/pubs/65430/new_domain_sentiment.pdf (accessed 13 July, 2016).

Bae, Y. and **Lee, H.** (2012). ‘Sentiment analysis of Twitter audiences: measuring the positive or negative influence of popular Twitterers'. *Journal of the American Society for Information Science and Technology*, 63(12), 2521–35.

Bakliwal, A., **Foster, J.**, **van der Puil, J.**, **O'Brien, R.**, **Tounsi, L.** and **Hughes, M.** (2013). ‘Sentiment analysis of political tweets: towards an accurate classifier'. In Proceedings of the Workshop on Language in Social Media (LASM 2013), Atlanta, GA, June 13. Stroudsburg, PA: Association for Computational Linguistics. pp. 49–58.

Barbieri, F., **Saggion, H.** and **Ronzano, F.** (2014). ‘Modelling sarcasm in Twitter, a novel approach'. 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2014), Baltimore, MA: ACL, pp. 50–8.

Basiri, M. E., **Ghasem-Aghaee, N.** and **Naghsh-Nilchi, A. R.** (2014). ‘Exploiting reviewers comment histories for sentiment analysis'. *Journal of Information Science*, 40(3), 313–28.

Blair-Goldensohn, S., **Hannan, K.**, **McDonald, R.**, **Neylon, T.**, **Reis, G. A.** and **Reynar, J.** (2008). ‘Building a sentiment summarizer for local service reviews'. In WWW Workshop on NLP in the Information Explosion Era. Vol. 14. https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/34368.pdf (accessed 13 July, 2016).

Blitzer, J., **McDonald, R.** and **Pereira, F.** (2006). 'Domain adaptation with structural correspondence learning'. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, pp. 120–8.

Bollen, J., **Gonçalves, B.**, **Ruan, G.** and **Mao, H.** (2011). 'Happiness is assortative in online social networks'. *Artificial Life*, 17(3), 237–51.

Bollen, J., **Mao, H.** and **Zeng, X.** (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.

Burgers, C., **Van Mulken, M.** and **Schellens, P. J.** (2012). 'Verbal irony differences in usage across written genres'. *Journal of Language and Social Psychology*, 31(3), 290–310.

Cambria, E., **Olsher, D.** and **Rajagopal, D.** (2014). 'SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis'. In Twenty-eighth AAAI Conference on Artificial Intelligence, July 27–31, Québec, Canada. http://sentic.net/senticnet-3.pdf (accessed 13 July, 2016).

Cambria, E., **Schuller, B.**, **Xia, Y.** and **Havasi, C.** (2013). 'New avenues in opinion mining and sentiment analysis'. *IEEE Intelligent Systems*, 28(2), 15–21.

Canales, L. and **Martínez-Barco, P.** (2014). 'Emotion detection from text: a survey'. In Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC 2014), pp. 1–8.

Carvalho, P., **Sarmento, L.**, **Silva, M. J.** and **De Oliveira, E.** (2009). 'Clues for detecting irony in user-generated contents: oh…!! It's so easy;-)'. In Proceedings of the 1st international CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion. New York, NY: ACM Press. pp. 53–6.

Chmiel, A., **Sienkiewicz, J.**, **Thelwall, M.**, **Paltoglou, G.**, **Buckley, K.**, **Kappas, A.** and **Hołyst, J.A.** (2011). 'Collective emotions online and their influence on community life'. *PLoS ONE*, 6(7): e22207.

Chung, J.E. and **Mustafaraj, E.** (2011). 'Can collective sentiment expressed on Twitter predict political elections?' In **W. Burgard** and **D. Roth** (eds.), Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, August 7–11. Menlo Park, CA: AAAI Press, pp. 1770–1.

Conover, M. D., **Gonçalves, B.**, **Ratkiewicz, J.**, **Flammini, A.** and **Menczer, F.** (2011). 'Predicting the political alignment of Twitter users'. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom). Los Alamitos, CA: IEEE Press, pp. 192–9.

Coviello, L., **Sohn, Y.**, **Kramer, A.D.I.**, **Marlow, C.**, **Franceschetti, M.**, **Christakis, N.A.** and **Fowler, J.H.** (2014). 'Detecting emotional contagion in massive social networks'. *PLoS ONE*, 9(3), e90315. doi:10.1371/journal.pone.0090315.

Davidov, D., **Tsur, O.** and **Rappoport, A.** (2010). ‘Semi-supervised recognition of sarcastic sentences in Twitter and Amazon'. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning. Stroudsburg, PA: Association for Computational Linguistics, pp. 107–16.

Diakopoulos, N. A. and **Shamma, D. A.** (2010). ‘Characterizing debate performance via aggregated Twitter sentiment'. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM Press, pp. 1195–8.

Esuli, A. and **Sebastiani, F.** (2006). ‘SentiWordNet: a publicly available lexical resource for opinion mining'. In Proceedings of fifth International Conference on Language Resources and Evaluation (LREC-2006), **Genoa, Italy**, Vol. 6, pp. 417–22.

Feldman, R. (2013). ‘Techniques and applications for sentiment analysis'. *Communications of the ACM*, 56(4), 82–9.

Filatova, E. (2012). ‘Irony and sarcasm: corpus generation and analysis using crowdsourcing'. In Eighth International Conference on Language Resources and Evaluation (LREC2012). Istanbul: European Language Resources Association, pp. 392–8.

Gamon, M., **Aue, A.**, **Corston-Oliver, S.** and **Ringger, E.** (2005). *‘Pulse: mining customer opinions from free text'*. In **A. Famili** (ed.) *Advances in Intelligent Data Analysis VI.* Berlin: Springer, pp. 21–32.

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 (ICWSM2007). Boulder, CO: AAAI Press. http://www.icwsm.org/papers/3–Godbole-Srinivasaiah-Skiena.pdf (accessed 13 July, 2016).

Golder, S. A. and **Macy, M. W.** (2011). ‘Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures'. *Science*, 333(6051), 1878–81.

Greaves, F., **Ramirez-Cano, D.**, **Millett, C.**, **Darzi, A.** and **Donaldson, L.** (2013). ‘Harnessing the cloud of patient experience: using social media to detect poor quality healthcare'. *BMJ Quality & Safety*, 22(3), 251–5.

Huang, M., **Ye, B.**, **Wang, Y.**, **Chen, H.**, **Cheng, J.** and **Zhu, X.** (2014). ‘New word detection for sentiment analysis'. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MA: ACL Press, pp. 531–41.

Ida, Y., **Nakamura, T.** and **Matsumoto, T.** (2013). ‘Domain-dependent/independent topic switching model for online reviews with numerical ratings'. In Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management. New York, NY: ACM Press, pp. 229–38.

Justo, R., **Corcoran, T.**, **Lukin, S. M.**, **Walker, M.** and **Torres, M. I.** (2014). ‘Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web'. *Knowledge-Based Systems,* 69(1), 124–33.

Kazemian, S., **Zhao, S.** and **Penn, G.** (2014). 'Evaluating sentiment analysis evaluation: A case study in securities trading'. 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2014), Baltimore, MA: ACL, pp. 119–27.

Kennedy, H. (2012). 'Perspectives on sentiment analysis'. *Journal of Broadcasting & Electronic Media*, 56(4), 435–50.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. New York, NY: Morgan Claypool.

Liu, B. and **Zhang, L.** (2012). *'A survey of opinion mining and sentiment analysis'*. In **C.C. Aggarwal**, **C. Zhai** (eds.), *Mining Text Data.* Berlin: Springer, pp. 415–63.

Liu, B., **Hu, M.** and **Cheng, J.** (2005). 'Opinion observer: analyzing and comparing opinions on the web'. In Proceedings of the 14th International Conference on World Wide Web. New York, NY: ACM Press, pp. 342–51.

Malouf, R. and **Mullen, T.** (2008). 'Taking sides: user classification for informal online political discourse'. *Internet Research*, 18(2), 177–90.

Mavridis, N. (2015). 'A review of verbal and non-verbal human–robot interactive communication'. *Robotics and Autonomous Systems*, 63(1), 22–35.

Melville, P., **Gryc, W.** and **Lawrence, R. D.** (2009). 'Sentiment analysis of blogs by combining lexical knowledge with text classification'. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM Press, pp. 1275–84.

Metaxas, P. T., **Mustafaraj, E.** and **Gayo-Avello, D.** (2011). 'How (not) to predict elections'. In Privacy, Security, Risk and Trust (PASSAT). Los Alamitos, CA: IEEE Press, pp. 165–71.

Misopoulos, F., **Mitic, M.**, **Kapoulas, A.** and **Karapiperis, C.** (2014). 'Uncovering customer service experiences with Twitter: the case of airline industry'. *Management Decision*, 52(4), 705–23.

Mohammad, S. M., **Zhu, X.** and **Martin, J.** (2014). 'Semantic role labeling of emotions in Tweets'. 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2014), Baltimore, MA: ACL, pp. 32–41.

Murthy, D. and **Petto, L. R.** (2015). 'Comparing print coverage and tweets in elections. A case study of the 2011–2012 US Republican primaries'. *Social Science Computer Review*, 0894439314541925.

Myslín, M., **Zhu, S. H.**, **Chapman, W.** and **Conway, M.** (2013). 'Using Twitter to examine smoking behavior and perceptions of emerging tobacco products'. *Journal of Medical Internet Research*, 15(8), e174.

Nassirtoussi, A., **Aghabozorgi, S.**, **Teh, Y.** and **Ngo, D.** (2014). 'Text mining for market prediction: a systematic review'. *Expert Systems with Applications*, 41(16), 7653–70.

Nassirtoussi, A. Aghabozorgi, S., **Wah, T.** and **Ngo, D.** (2015). 'Text mining of news-headlines for FOREX market prediction: a multi-layer dimension reduction algorithm with semantics and sentiment'. *Expert Systems with Applications*, 42(1), 306–24.

Nielsen, F. Å. (2011). 'A new ANEW: evaluation of a word list for sentiment analysis in microblogs'. In Workshop on Making Sense of Microposts: big things come in small packages. pp. 93–8. CEUR Workshop Proceedings, no. 718. http://ceur-ws.org/Vol-718/paper_16.pdf (accessed 13 July 2016).

O'Connor, B., **Balasubramanyan, R.**, **Routledge, B. R.** and **Smith, N. A.** (2010). 'From Tweets to polls: linking text sentiment to public opinion time series'. *International Conference on Web and Social Media*, 11, 122–9.

Paltoglou, G. and **Thelwall, M.** (2010). 'A study of information retrieval weighting schemes for sentiment analysis'. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, pp. 1386–95.

Pang, B. and **Lee, L.** (2008). 'Opinion mining and sentiment analysis'. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.

Pang, B., **Lee, L.** and **Vaithyanathan, S.** (2002). 'Thumbs up?: Sentiment classification using machine learning techniques'. In Proceedings of the ACL-02 conference on Empirical Methods in Natural Language Processing, Vol. 10. Stroudsburg, PA: Association for Computational Linguistics, pp. 79–86.

Ponomareva, N. and **Thelwall, M.** (2012). *'Biographies or blenders: which resource is best for cross-domain sentiment analysis?'*. In **A. Gelbukh** (ed.), *Computational Linguistics and Intelligent Text Processing*. Berlin: Springer, pp. 488–99.

Poria, S., **Cambria, E.**, **Winterstein, G.** and **Huang, G. B.** (2014). 'Sentic patterns: dependency-based rules for concept-level sentiment analysis'. *Knowledge-Based Systems*, 69(1), 45–63.

Skowron, M., **Theunis, M.**, **Rank, S.** and **Borowiec, A.** (2011). *'Effect of affective profile on communication patterns and affective expressions in interactions with a dialog system'*. In **S. D'Mello**, **A. Graesser**, **B. Schuller**, **J. Martin** (eds.), *Affective Computing and Intelligent Interaction.* Berlin: Springer, pp. 347–56.

Stone, P. J., **Bales, R. F.**, **Namenwirth, J. Z.** and **Ogilvie, D. M.** (1962). 'The General Inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information'. *Behavioral Science*, 7(4), 484–98.

Strapparava, C. and **Mihalcea, R.** (2008). 'Learning to identify emotions in text'. In Proceedings of the 2008 ACM Symposium on Applied Computing. New York, NY: ACM Press, pp. 1556–60.

Strapparava, C. and **Valitutti, A.** (2004). 'WordNet Affect: an affective extension of WordNet'. In **N. Calzolari** (ed.) Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004), Lisbon, Portugal, Vol. 4, pp. 1083–86.

Subramanian, S., **Bear, M. E.**, **Setayesh, M.** and **Horton, N.** (2013). US Patent Application 14/021,798.

Taboada, M. and **Grieve, J.** (2004). 'Analyzing appraisal automatically'. In Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report), Stanford University, CA. Menlo Park, CA: AAAI Press, pp. 158–61.

Taboada, M., **Brooke, J.**, **Tofiloski, M.**, **Voll, K.** and **Stede, M.** (2011). 'Lexicon-based methods for sentiment analysis'. *Computational Linguistics*, 37(2), 267–307.

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*, 29(1), 24–54.

Thelwall, M. (2010). 'Emotion homophily in social network site messages'. First Monday, 15(4). http://firstmonday.org/ojs/index.php/fm/article/view/2897/2483 (accessed 13 July 2016).

Thelwall, M. and **Buckley, K.** (2013). 'Topic-based sentiment analysis for the Social Web: the role of mood and issue-related words'. *Journal of the American Society for Information Science and Technology*, 64(8), 1608–17.

Thelwall, M., **Buckley, K.** and **Paltoglou, G.** (2011). 'Sentiment in Twitter events'. *Journal of the American Society for Information Science and Technology*, 62(2), 406–18.

Thelwall, M., **Buckley, K.** and **Paltoglou, G.** (2012). 'Sentiment strength detection for the social Web'. *Journal of the American Society for Information Science and Technology*, 63(1), 163–73.

Thelwall, M., **Buckley, K.**, **Paltoglou, G.**, **Cai, D.** and **Kappas, A.** (2010). 'Sentiment strength detection in short informal text'. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–58.

Thelwall, M., **Buckley, K.**, **Paltoglou, G.**, **Skowron, M.**, **Garcia, D.**, **Gobron, S.**, **Ahn, J.**, **Kappas, A.**, **Küster, D.** and **Holyst, J.A.** (2013). *'Damping sentiment analysis in online communication: discussions, monologs and dialogs'*. In **A. Gelbukh** (ed.), *CICLing 2013*, Part II, LNCS 7817. Heidelberg: Springer, pp. 1–12.

Tong, R. (2001). 'An operational system for detecting and tracking opinions in on-line discussions'. Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification. New York, NY: ACM, pp. 1–6.

Tumasjan, A., **Sprenger, T. O.**, **Sandner, P. G.** and **Welpe, I. M.** (2010). 'Predicting elections with Twitter: what 140 characters reveal about political sentiment'. In International AAAI Conference on Weblogs and Social Media (ICWSM2010), pp. 178–85. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441 (accessed 13 July 2016).

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 (ACL2002). Stroudsburg, PA: Association for Computational Linguistics, pp. 417–24.

van Atteveldt, W., **Kleinnijenhuis, J.**, **Ruigrok, N.** and **Schlobach, S.** (2008). ‘Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations'. *Journal of Information Technology & Politics*, 5(1), 73–94.

Vilares Calvo, D., **Thelwall, M.** and **Alonso, M.A.** (2015). ‘The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets'. *Journal of Information Science*, 41(6), 799–813.

Vural, A.G., **Cambazoglu, B.B.** and **Karagoz, P.** (2014). ‘Sentiment-focused web crawling'. *ACM Transactions on the Web*, 8(4), article 22.

Vural, A.G., **Cambazoglu, B.B.**, **Senkul, P.** & **Tokgoz, O.** (2013). *‘A framework for sentiment analysis in Turkish: application to polarity detection of movie reviews in Turkish'*. In *Computer and Information Sciences III*, London: Springer, pp. 437–45.

Wang, H., **Can, D.**, **Kazemzadeh, A.**, **Bar, F.** and **Narayanan, S.** (2012). ‘A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle'. In Proceedings of the ACL 2012 System Demonstrations. Stroudsburg, PA: Association for Computational Linguistics, pp. 115–20.

Widener, M. J. and **Li, W.** (2014). ‘Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US'. *Applied Geography*, 54(1), 189–97.

Wiebe, J., **Wilson, T.** and **Cardie, C.** (2005). ‘Annotating expressions of opinions and emotions in language'. *Language Resources and Evaluation*, 39(2–3), 165–210.

Wu, Y., **Liu, S.**, **Yan, K.**, **Liu, M.** and **Wu, F.** (2012). ‘OpinionFlow: visual analysis of opinion diffusion on social media'. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1763–72.

Young, L. and **Soroka, S.** (2012). ‘Affective news: the automated coding of sentiment in political texts'. *Political Communication*, *29*(2), 205–31.

Zagibalov, T. and **Carroll, J.** (2008). ‘Unsupervised classification of sentiment and objectivity in Chinese text'. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK: ICCL, pp. 1073–80.

Zhang, Z. and **Singh, M. P.** (2014). ‘ReNew: a semi-supervised framework for generating domain-specific lexicons and sentiment analysis'. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MA: ACL Press, pp. 542–51.

Zhao, K., **Yen, J.**, **Greer, G.**, **Qiu, B.**, **Mitra, P.** and **Portier, K.** (2014). ‘Finding influential users of online health communities: A new metric based on sentiment influence'. *Journal of the American Medical Informatics Association*, 21(e2), e212–18.

http://dx.doi.org/10.4135/9781473957992.n20