Journal of Web Engineering, Vol. 14, No. 1&2 (2015) 022–040

EXPLOITING EMOTICONS IN POLARITY CLASSIFICATION OF TEXT

ALEXANDER HOGENBOOM, DANIELLA BAL, FLAVIUS FRASINCAR, MALISSA BAL
Econometric Institute, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, the Netherlands
hogenboom@ese.eur.nl, daniella.bal@xs4all.nl, frasincar@ese.eur.nl, malissa.bal@xs4all.nl

FRANCISKA DE JONG
Erasmus School of History, Culture & Communication, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, the Netherlands
Department of Computer Science, Universiteit Twente
P.O. Box 217, 7500 AE Enschede, the Netherlands
f.m.g.dejong@utwente.nl

UZAY KAYMAK
Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, the Netherlands
u.kaymak@ieee.org

With people increasingly using emoticons in written text on the Web in order to express, stress, or disambiguate their sentiment, it is crucial for automated sentiment analysis tools to correctly account for such graphical cues for sentiment. We analyze how emoticons typically convey sentiment and we subsequently propose and evaluate a novel method for exploiting this with a manually created emoticon sentiment lexicon in a lexicon-based polarity classification method. We evaluate our approach on 2,080 Dutch tweets and forum messages, which all contain emoticons. We validate our findings on 10,069 English reviews of apps, some of which contain emoticons. We find that accounting for the sentiment conveyed by emoticons on a paragraph level – and, to a lesser extent, on a sentence level – significantly improves polarity classification performance. Whenever emoticons are used, their associated sentiment tends to dominate the sentiment conveyed by textual cues and forms a good proxy for the polarity of text.

Keywords: Sentiment analysis, polarity classification, emoticons, sentiment lexicon

1 Introduction

Nowadays, popular Web sites like Twitter, Blogger, and Epinions allow their users to vent opinions on just about anything through an ever-increasing amount of short messages, blog posts, or reviews. This yields a continuous flow of an overwhelming amount of data, containing traces of valuable information – people's sentiment with respect to products, brands, etcetera. As estimates indicate that one in three blog posts [34] and one in five tweets [25] discuss products or brands, the abundance of user-generated content published through the Social Web renders automated information monitoring tools crucial for today's businesses [38].

Automated sentiment analysis techniques come to answer this need [14, 28, 42]. Sentiment analysis refers to a broad area of natural language processing, computational linguistics, and text mining. Typically, the goal of sentiment analysis is to determine the polarity of natural language text. An intuitive approach involves scanning a text for cues signaling its polarity.

In face-to-face communication, sentiment can often be deduced from visual cues like frowning or smiling. However, in plain-text computer-mediated communication, such visual cues are lost. Over the years, people have embraced the usage of so-called emoticons as an alternative to face-to-face visual cues in (on-line) computer-mediated communication like virtual utterances of opinions in the Social Web. In this light, we define emoticons as visual cues used in texts to replace normal visual cues like frowning or smiling in order to express, stress, or disambiguate one's sentiment. Emoticons are typically sequences of typographical symbols such as "*:*", "=", "-", "*)*", or "*(*" and commonly represent facial expressions. Emoticons can be read either sideways, like "*:-(*" (a sad face), or normally, like "*(^_^)*" (a happy face).

Several types of automated polarity classification methods have been proposed [14, 28, 42]. Many state-of-the-art methods represent natural language text as a bag-of-words, i.e., an unordered collection of the words occurring in the text. Such an approach allows for vector representations of text, enabling the use of machine learning techniques for classifying the polarity of text. Features in such representations may be, e.g., words or parts of words. However, machine learning polarity classifiers typically require a lot of training data in order to function properly. Moreover, even though machine learning classifiers may perform very well in the domain that they have been trained on, their performance drops significantly when they are used in another domain [49]. Additionally, machine learning polarity classifiers typically give little insight into why a text is assigned a specific polarity classification.

Lexicon-based methods [6, 12, 13, 17, 19, 48] are attractive alternatives with a more robust performance across domains and texts [49]. These essentially rule-based methods keep a more linguistic view on textual data rather than abstracting away from natural language by means of vectorization. As such, deep linguistic analysis comes more naturally in lexicon-based approaches, thus allowing for intuitive ways of accounting for structural [17] or semantic [22] aspects of text in sentiment analysis. Lexicon-based methods use sentiment lexicons for retrieving the polarity of individual words and aggregate these scores – e.g., by computing a (weighted) sum or average of the individual word scores – in order to determine the text's overall polarity. A sentiment lexicon typically contains simple and compound words and their associated sentiment, possibly differentiated by Part-of-Speech (POS) and/or meaning [1].

Recent findings suggest that people's sentiment is not so much conveyed by the words in a text per se, but rather by *how* these words are used [2, 7, 17]. Emoticons could be helpful in this context, as they may for instance signal the intended sentiment of an otherwise objective statement, e.g., "*This product does not work :-(*". However, today's lexicon-based polarity classification approaches typically do not consider emoticons. Conversely, one of the first steps in most existing work is to remove many of the typographical symbols constituting emoticons, thus preventing emoticons from being detected at all. Therefore, we aim to investigate how emoticons are typically used to convey sentiment and how we can exploit this by using an additional emoticon lexicon in order to improve lexicon-based polarity classification.

We extend our existing work [20] with more details and a validation of our results on new data. Our polarity classifier detects emoticons, determines their sentiment by using an emoticon sentiment lexicon, and assigns this sentiment to the text affected by the detected emoticons. The emoticon-based information thus obtained is then combined with the sentiment conveyed by verbal cues in the remaining, unaffected text in order to help correctly classify the polarity of a document as either positive or negative.

The remainder of this paper is structured as follows. First, Section 2 elaborates on sentiment analysis and how emoticons are used in computer-mediated communication. Then, in Section 3, we analyze how emoticons are typically related to the polarity of the text they occur in and we additionally propose a method for harvesting information from emoticons when analyzing the polarity of text. The performance of our novel approach is assessed in Section 4. Last, in Section 5, we draw conclusions and propose directions for future work.

2 Related Work

In a literature survey on sentiment analysis [42], the current surge of research interest in systems that deal with opinions and sentiment is attributed to the fact that, in spite of today's users' hunger for and reliance upon on-line advice and recommendations, explicit information on user opinions is often hard to find, confusing, or overwhelming. Many sentiment analysis approaches exist, yet harvesting information from emoticons has been relatively little explored.

2.1 Sentiment analysis

The roots of sentiment analysis are in fields like natural language processing, computational linguistics, and text mining. The main objective of most sentiment analysis approaches is to extract subjective information from natural language text. Most work focuses on determining the overall polarity of words, sentences, text segments, or documents [42]. This task is commonly approached as a binary classification problem, in which a text is to be classified as either positive or negative. It can be considered as a ternary classification problem as well, by introducing a third class of neutral documents. An alternative to such polarity classification approaches is the determination of a degree of positivity or negativity of natural language text in order to produce, e.g., rankings of positive and negative documents [7, 8]. In this paper, we address the binary document-level polarity classification problem, dealing with classifying documents as either positive or negative.

As sentiment analysis tools have particularly useful applications in marketing and reputation management [5, 14, 18, 27], polarity classification tools are often evaluated on collections of reviews. Reviews typically contain people's opinions expressed in natural language, and often have an associated (numeric) score quantifying one's judgment. A widely used corpus for assessing polarity classification methods is a collection of 2,000 English movie reviews [41].

Among the popular bag-of-word approaches to polarity classification, a binary representation of text, indicating the presence of specific words, has initially proven to be an effective approach, yielding an accuracy of 87% on the English movie review collection [41]. Later research has focused on different vector representations of text, including vector representations with additional features representing semantic distinctions between words [53] or vector representations with *tf-idf*-based weights for word features [40]. Such approaches typically yield a polarity classification accuracy on the movie review data set of over 90%.

The alternative lexicon-based approaches typically exhibit lower accuracy on the movie review data set, but tend to be more robust across domains [49]. Additionally, the rule-based nature of lexicon-based polarity classification methods allows for insight into why a text is assigned a particular classification. Moreover, lexicon-based approaches can be generalized relatively easily to other languages by means of dictionaries [35]. A rather simple lexicon-based method has been shown to have an accuracy of up to 60% on the movie review data [18].

A more sophisticated lexicon-based polarity classification approach has been shown to have an average accuracy of 68% on 1,900 documents from the movie review data set [48]. A deeper linguistic analysis focusing on differentiating between rhetorical roles of text segments has recently been proven to perform comparably well too [17]. On 1,000 documents from the movie review data set, this approach yields an accuracy of 72%, which is about a 5% relative improvement over not accounting for structural aspects of content.

Even though some lexicon-based polarity classification approaches explore promising new directions of incorporating structural and semantic aspects of content [17, 22, 23], they typically fail to harvest information from potentially important cues for sentiment in today's user-generated content – emoticons. Nevertheless, emoticons have already been exploited to a limited extent, mainly for automated data annotation.

For instance, in early work, a crude distinction between a handful of positive and negative emoticons has been used to automatically generate data sets with positive and negative samples of natural language text in order to train and test polarity classification techniques [43]. These early results suggest that the sentiment conveyed by emoticons is topic- and domain-independent. These findings have been successfully applied in later work in order to automatically construct collections of positive and negative tweets [11, 39], or collections of tweets in alternative sentiment categories, such as angry and sad emotional states [56]. Combining such automatically annotated training data with manually labeled training data has been shown to yield sentiment analysis classifiers that outperform similar classifiers that have been trained on manually annotated training data only [29].

The aforementioned work on exploiting emoticons in sentiment analysis [11, 29, 39, 43, 56] has been focused on using emoticons for automatically constructing training data, rather than on interpreting the emoticons in unseen text, in order to better classify its conveyed sentiment. In some work, however, a small set of emoticons has in fact been used as features for polarity classification, in addition to more common features such as sentiment-carrying words [11, 51]. Yet, the findings of the latter work do not indicate that treating emoticons as if they are normal sentiment-carrying words yields a significant improvement over ignoring emoticons when classifying the polarity of natural language text. In this light, treating emoticons differently may seem a more viable approach, yet an alternative method that solely focuses on emoticons as cues for sentiment has been shown to yield a high performance on only a small subset of documents [16]. Provided that emoticons are, nevertheless, important cues for sentiment in today's user-generated content, the key to harvesting information from emoticons lies in understanding how they relate to a text's overall polarity.

To the best of our knowledge, existing research does not focus on investigating how emoticons affect the polarity of natural language text, nor on exploring how this phenomenon can be successfully exploited in lexicon-based polarity classification. In order to address this hiatus, we must first understand how emoticons are used in computer-mediated communication.

2.2 Emoticons

Research has demonstrated that humans are clearly influenced by the use of nonverbal cues in face-to-face communication [9, 46]. Nonverbal cues have even been shown to dominate verbal cues in face-to-face communication in case verbal and nonverbal cues are equally strong [4]. Apparently, nonverbal cues are deemed important indicators for people in order to understand the intentions and emotions of whomever they are communicating with. Translating these findings to computer-mediated communication does hence not appear to be too far-fetched, if it were not for the fact that plain-text computer-mediated communication does not leave much room for nonverbal cues.

However, users of computer-mediated communication have found their ways of overcoming the lack of personal contact by using emoticons. The first emoticon was used on September 19, 1982 by professor Scott Fahlman in a message on the computer science bulletin board of Carnegie Mellon University. In his message, Fahlman proposed to use the character sequences "*:-)*" and "*:-(*" in order to clearly distinguish jokes from more serious matters, respectively. It did not take long before the phenomenon of emoticons had spread to a much larger community. People started sending yells, hugs, and kisses by using graphical symbols formed by characters found on a typical keyboard. A decade later, emoticons had found their way into everyday computer-mediated communication and had become the paralanguage of the Web [33]. By then, 6% of the messages on electronic mailing lists [44] and 13% of UseNet newsgroup posts [55] were estimated to contain emoticons.

Thus, nonverbal cues have emerged in computer-mediated communication. It should however be noted that these nonverbal cues in computer-mediated communication are conceptually different from nonverbal cues in face-to-face communication. Real-life cues like laughing and weeping are often considered to be involuntary ways of expressing oneself in face-to-face communication, whereas the use of their respective equivalents "*:-)*" and "*:-(*" in computer-mediated communication is intentional [26]. As such, emoticons enable people to indicate subtle mood changes, to signal irony, sarcasm, and jokes, and to express, stress, or disambiguate their (intended) sentiment, perhaps even more than nonverbal cues in face-to-face communication can. Therefore, harvesting information from emoticons appears to be a viable strategy to improve the state-of-the-art of sentiment analysis. Yet, the question is not so much *whether*, but rather *how* we should account for emoticons when classifying the polarity of a document.

3 Emoticons and Polarity

In order to exploit emoticons in an automated polarity classification setting, we first need to analyze how emoticons are typically related to the polarity of the text they occur in. Insights into what parts of a text are affected by emoticons in which way are crucial for advancing the state-of-the-art of polarity classification by harvesting information from emoticons.

3.1 Emoticons as cues for polarity

In order to assess the role emoticons play in conveying the sentiment of a text, we have performed a qualitative analysis of a collection of 2,080 Dutch tweets and forum messages. We have randomly sampled this content from search results from Twitter and Google discussion groups when querying for brands like Vodafone, KLM, Kinect, etcetera.

The first hypothesis that we have evaluated on our data is the hypothesis of emoticons having a rather local effect, i.e., emoticons affecting the paragraph or sentence rather than the document in which they occur. Paragraphs typically address distinct topics or points of view, thus rendering the applicability of an emoticon in one paragraph to another unlikely. In our sample collection, upon inspection, emoticons generally have a paragraph-level effect

Table 1. Typical examples of how emoticons can be used to convey sentiment.

Sentence	How	Sentiment
I love my work :-D	Intensification	Positive
The movie was bad :-D	Negation	Positive
:-D I got a promotion	Only sentiment	Positive
-_- I love my work	Negation	Negative
The movie was bad -_-	Intensification	Negative
I got a promotion -_-	Only sentiment	Negative

for those paragraphs containing only one emoticon. In case a paragraph contains multiple emoticons, the analysis of our sample shows that an emoticon is generally more likely to affect the sentence in which it occurs.

In our sample, 84% of all emoticons are placed at the end of a paragraph, 9% are positioned somewhere in the middle of a paragraph, and 7% are used at the beginning of a paragraph. This positioning of emoticons suggests that it is typically not a single word, but rather a text segment that is affected by an emoticon. Additionally, these results imply that in case an emoticon is used in the middle of a paragraph with multiple emoticons, the emoticon is statistically more likely to be associated with the preceding text segment.

In addition to assessing *what* is affected by emoticons, we have analyzed *how* emoticons affect text as well. Our sample shows that emoticons can generally be used in three ways. First, emoticons can be used to express sentiment when sentiment is not conveyed by any clearly positive or negative words in a text segment, thus rendering the emoticons to be carrying the only sentiment in such cases. Second, emoticons can stress sentiment by intensifying the sentiment already conveyed by sentiment-carrying words. Third, emoticons can be used to disambiguate sentiment, for instance in cases where the sentiment associated with sentiment-carrying words needs to be negated. Some examples can be found in Table 1.

Table 1 clearly demonstrates how the sentiment associated with a piece of text can differ when using different emoticons, i.e., the happy emoticon "*:-D*" and the "-_-" emoticon indicating extreme boredom or disagreement, irrespective of the position of the emoticons. The sentiment carried by an emoticon is independent from its embedding text, rendering word sense disambiguation techniques [37] not useful for emoticons. Thus, the sentiment of emoticons appears to be dominating the sentiment carried by verbal cues in sentences, if any.

In some cases, this may be a crucial property that can be exploited by automated sentiment analysis approaches. For instance, when an emoticon is the only cue in a sentence conveying sentiment, we are typically dealing with a phenomenon that we refer to as factual sentiment. For example, the sentence "*I got a promotion*" does nothing more than stating the fact that one got promoted. However, getting a promotion is usually linked to a positive emotion like happiness or pride. Therefore, human interpreters could be inclined to acknowledge the implied sentiment and thus consider the factual statement to be a positive statement. However, this requires an understanding of context and involves incorporating real-world knowledge into the process of sentiment analysis. This is a cumbersome task for machines. In this light, emoticons can be valuable cues for deriving an author's intended sentiment.

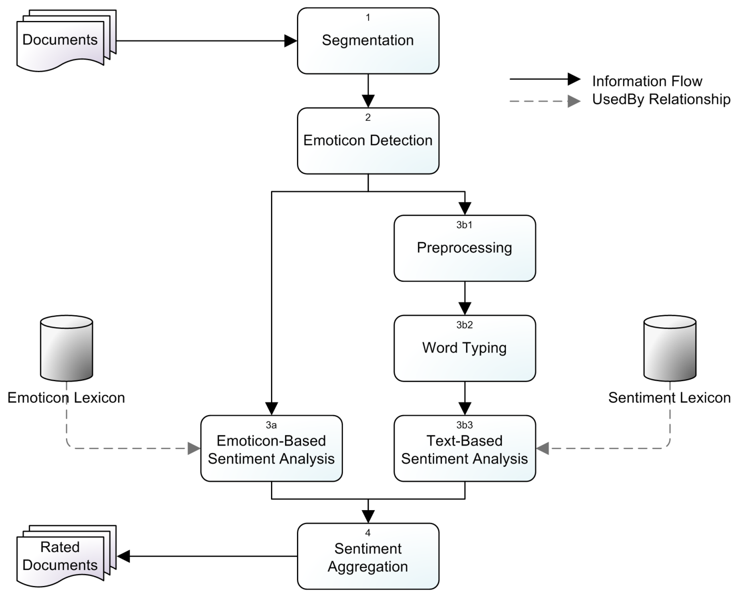

Fig. 1. Overview of our sentiment analysis framework.

3.2 Framework

We propose a framework for automated document-level polarity classification, which takes into account the information conveyed by emoticons. This framework detects emoticons, determines their sentiment, and assigns the associated sentiment to the text affected by the detected emoticons. The emoticon-based information thus obtained is then combined with the sentiment conveyed by verbal cues in the remaining unaffected text, in order to help correctly classify the polarity of a document as either positive or negative. In order to accomplish this, we build upon existing work [2]. Our framework, depicted in Figure 1, is a pipeline in which each component fulfills a specific task in analyzing the sentiment of a document. Here, a document is a coherent piece of text that can be as small as a one-line tweet or as big as a news article, review, blog, or forum message with multiple paragraphs.

First, we load a document to be analyzed for sentiment. Then, the document is split into text segments, which may be either paragraphs or sentences (step 1). Sentiment analysis is subsequently initially performed on segment level, after which the segment-level sentiment analysis results are combined.

Each text segment is checked for the presence of emoticons (step 2). To this end, we propose to use an emoticon sentiment lexicon, which we define as a list of character sequences – emoticons – and their associated sentiment scores. These emoticons may be organized into emoticon synsets, which we define as groups of emoticons denoting the same emotion. Table 2 shows examples of such emoticon synsets, representing emotions like happiness and sadness. When checking a text segment for the presence of emoticons, we compare each word in the text segment with the entries in the emoticon sentiment lexicon. Here, we consider words to be character sequences, separated by whitespace characters. If a word in a text segment matches a character sequence in the emoticon sentiment lexicon, the segment is rated for

Table 2. Typical examples of emoticon synsets.

Emoticon synset	Emoticons
Happiness	:-D, =D, xD, (^_^)
Sadness	:-(, =(
Crying	:'(, ='(, (;_;)
Boredom	-_-, -.-, (>_<)
Love	<3, (L)
Embarrassment	:-$, =$, >///<

sentiment based on the sentiment imposed onto the text by its emoticons (step 3a). Else, the segment is analyzed for the sentiment conveyed by its sentiment-carrying words (step 3b1–3).

In case a text segment is analyzed based on the emoticons it contains (step 3a), the segment is assigned a sentiment score equal to the total (summed) sentiment associated with all of its emoticons, as derived from the emoticon sentiment lexicon. Sentiment scores of sentiment-carrying words (if any) are ignored in this process, as the sentiment of emoticons tends to dominate the sentiment carried by verbal cues (see Section 3.1).

In order to analyze a text segment for the sentiment conveyed by its sentiment-carrying words (step 3b1–3), it is first preprocessed by removing diacritics and other special characters (step 3b1) and identifying each word's POS, lemma, and its purpose in the text, i.e., sentiment-carrying term, modifying term, or irrelevant term (step 3b2). Following existing work [2], we consider modifying terms to change the sentiment of corresponding sentiment-carrying word(s). We assume negations to change the sentiment sign and amplifiers to multiply the sentiment of the affected sentiment words with an appropriate factor. After determining the word types, the text segment is rated for its conveyed sentiment by means of a lexicon-based sentiment scoring method that essentially computes the sentiment of the text segment as the total sentiment score of all (modified) sentiment-carrying words in the segment (step 3b3).

As such, the sentiment score ζ_{s_i} of the i-th segment s_i of document d can be computed as either a function of the sentiment scores ζ_{e_j} of each emoticon e_j in segment s_i, or as a function of the sentiment scores ζ_{t_j} of each sentiment-carrying word t_j and the weight w_{t_j} of its modifying term (if any, else, this weight defaults to 1), i.e.,

$$\zeta_{s_i} = \begin{cases} \sum_{j=1}^{V_i} \zeta_{e_j} & \text{if } V_i > 0, \\ \sum_{j=1}^{T_i} \left(\zeta_{t_j} \cdot w_{t_j}\right) & \text{else,} \end{cases} \tag{1}$$

with V_i the number of visual cues for sentiment in segment s_i and T_i the number of sentiment-carrying textual cues (i.e., combinations of sentiment-carrying words and their modifiers) in the segment. In (1), ζ_{e_j} and ζ_{t_j} are real numbers ranging from -1 (negative) to 1 (positive).

After determining the sentiment conveyed by each individual text segment, all text segments are recombined into a single document. Note that a document can have both segments with and without emoticons. The sentiment score ζ_d of a document d is then calculated as the average over all segment-level sentiment scores, i.e.,

$$\zeta_d = \frac{\sum_{i=1}^{p} \zeta_{s_i}}{\sum_{i=1}^{p} \left(V_i + (a_i \cdot T_i)\right)}, \tag{2}$$

with p the number of segments of document d and a_i indicating whether a full sentiment analysis needs to be performed on the textual cues of text segment s_i (1) or not (0), i.e.,

$$a_i = \begin{cases} 0 & \text{if } V_i > 0, \\ 1 & \text{else.} \end{cases} \tag{3}$$

A negative document-level sentiment score thus computed typically indicates a negative document (-1), whereas other scores yield a positive classification (1). As sentiment scores can to a certain extent be used to distinguish between universal classes of intended sentiment [21], the classification c_d of document d is defined as a function of its sentiment score ζ_d, i.e.,

$$c_d = \begin{cases} 1 & \text{if } \zeta_d \geq 0, \\ -1 & \text{else.} \end{cases} \tag{4}$$

4 Polarity Classification by Exploiting Emoticons

By means of a set of experiments, we evaluate our novel method of polarity classification of natural language text by exploiting emoticons. The experimental setup is detailed in Section 4.1, whereas we discuss our results and some caveats in Sections 4.2, 4.3, and 4.4.

4.1 Experimental setup

We evaluate our method on a collection of Dutch tweets and forum posts and we validate our findings on a collection of documents in another domain and language, i.e., English app reviews. Section 4.1.1 describes these data sets. Details on the implementation of our framework and the construction of our emoticon sentiment lexicon are provided in Sections 4.1.2 and 4.1.3, respectively. Last, in Section 4.1.4, we present our evaluation methodology.

4.1.1 Data

Our Dutch collection consists of 2,080 Dutch documents, i.e., 1,040 tweets and 1,040 forum messages. We have randomly sampled these messages from search results from Twitter and Google discussion groups when querying for brands like Vodafone, KLM, and Kinect, while making sure to select only those texts that contain emoticons. Three human annotators have manually annotated these documents for their associated polarity, i.e., positive or negative, until they reached agreement. The resulting data set consists of 1,067 positive documents and 1,013 negative documents. Emoticons occur in all documents in this data set.

The 10,069 English app reviews in our validation set have been crawled from Apple's App Store for the United Kingdom. The reviewed apps in our corpus range from, among others, the Dropbox, Gmail, and WhatsApp Messenger apps to the TomTom Europe, Bloomberg, and Pocket Whip apps. Each review was annotated by its author for the associated sentiment by means of a star rating, ranging from one star (very negative) to five stars (very positive). We have converted these ratings into binary polarity classes by assigning a negative classification to reviews with one, two, or three stars, and a positive classification to reviews with four or five stars. We have thus obtained 7,017 positive and 3,052 negative English app reviews. By applying our emoticon detection method described in Section 3.2, we have detected emoticons in a subset of 655 English app reviews, i.e., in 527 positive and 128 negative documents.

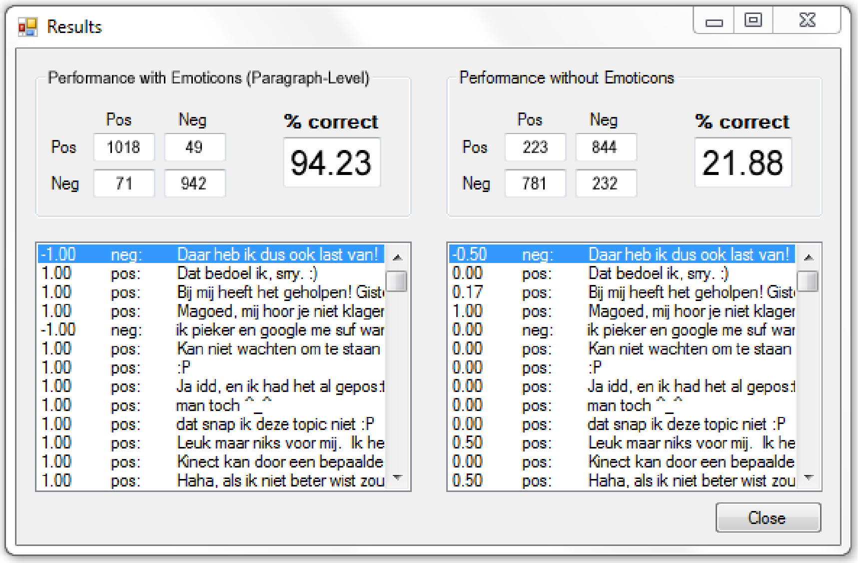

Fig. 2. Graphical user interface facilitating comparison of results.

4.1.2 Implementation

In a C# implementation of our framework for polarity classification of Dutch documents, we look for empty lines or lines starting with an indentation in order to split a document into paragraphs. When splitting a document into sentences, we look for punctuation marks, such as ".", "*!*", and "*?*", as well as for emoticons, as most emoticons are placed at the end of a text segment (see Section 3.1). We utilize a proprietary maximum-entropy based POS tagger for Dutch and a proprietary sentiment lexicon for Dutch words, both of which have been provided to us by Teezir (`http://www.teezir.com`). This proprietary sentiment lexicon enables us to retrieve both the sentiment scores of sentiment-carrying words and the values for their associated modifiers, i.e., negators or amplifiers, if any. The design of the graphical user interface of our software, depicted in Figure 2 facilitates the comparison between sentiment analysis with and without taking into account the information conveyed by emoticons.

In a similar, Java-based implementation of our framework for polarity classification of English documents, we detect paragraphs by making use of empty lines in a document, as we assume these empty lines to be used to separate two paragraphs from one another. This is a valid assumption for our collection of English app reviews. Furthermore, we employ the Stanford Tokenizer [30] for identifying sentences and words in the identified paragraphs. For POS tagging and lemmatization of words, we use the OpenNLP [3] POS tagger and the Java WordNet Library (JWNL) API [52], respectively. We use the identified lemma and POS of each word in order to retrieve its associated sentiment score from the SentiWordNet 3.0 [1] sentiment lexicon. Following recent findings [24], we account for negation by inverting the polarity of the two words following a negation keyword that is listed in a negation lexicon [24]. Last, we account for amplification by means of an existing amplification lexicon, listing amplification keywords and their effect on the sentiment conveyed by the first succeeding word [48].

4.1.3 Emoticon sentiment lexicon

One of the key elements in our novel emoticon-based polarity classification framework is the emoticon sentiment lexicon. Several lists of emoticons are readily available [10, 32, 50, 36, 15, 45, 31, 54]. We propose to combine these eight existing lists into one large emoticon sentiment lexicon. In this process, we leave out duplicate entries. Additionally, we leave out character representations of body parts and representations of objects, as the latter two types of emoticons do not carry any sentiment.

This process yields a list of 477 emoticons representing facial expressions or body poses like thumbs up. Three human annotators have manually rated the emoticons in our lexicon for their associated sentiment with a numeric score in the interval $[-1, 1]$, which we have discretized a priori into -1.0 (negative), 0.0 (neutral), and 1.0 (positive), in order for the scores of sentiment-carrying emoticons to reflect the dominance of emoticons over textual cues. As the sentiment scores are assumed to be real numbers (see Section 3.2), the numeric sentiment score of each individual emoticon has subsequently been determined as the score closest to the average of the annotators' scores for that emoticon, thus assigning equal weights to the annotators' opinions. For about 88% of the considered emoticons, our three annotators assigned identical scores to the respective emoticons. The lexicon thus generated[a] is utilized in two implementations of our framework – one for polarity classification of Dutch documents, the other for polarity classification of English documents.

4.1.4 Evaluation

The implementation of our proposed polarity classification framework allows us to perform experiments in order to compare the performance of several configurations of our framework. First, we consider an absolute baseline of not accounting for the information conveyed by emoticons (BASELINE), thus essentially reducing our analysis to an existing lexicon-based document-level polarity classification method [2]. Then, as a first alternative, we consider an approach in which the sentiment conveyed by emoticons is assumed to affect the surrounding text on a sentence level (EMO.S). Last, we consider accounting for the sentiment conveyed by emoticons on a paragraph level when classifying the polarity of a piece of text (EMO.P).

We assess the performance of each method in terms of precision, recall, and F_1-score for positive and negative documents separately, as well as the overall accuracy and macro-level F_1-score. Precision is the proportion of the positively (negatively) classified documents which have an actual classification of positive (negative), whereas recall is the proportion of the actual positive (negative) documents which are also classified as such. The F_1-score is the harmonic mean of precision and recall. The macro-level F_1-score is the average of the F_1-scores of the positive and negative documents, weighted for their respective relative frequencies. Accuracy is the proportion of correctly classified documents.

In order to get a clear view on the impact of accounting for the sentiment conveyed by emoticons in sentiment analysis, we assess the statistical significance of the observed performance differences by means of a paired two-sample one-tailed t-test. To this end, we randomly split our data sets into ten equally sized subsets, on which we assess the performance of our considered methods. The mean performance measures over these subsets can then be compared by means of the t-test.

[a] Available on-line at `http://people.few.eur.nl/hogenboom/files/EmoticonSentimentLexicon.zip`

Table 3. Performance measures of all of our considered approaches on our collection of Dutch tweets and forum posts. The best performance is printed in bold for each performance measure.

	Positive			Negative			Overall	
Method	Precision	Recall	F_1	Precision	Recall	F_1	Accuracy	F_1
BASELINE	0.222	0.209	0.215	0.216	0.229	0.222	0.219	0.219
EMO.S	0.670	0.650	0.660	0.680	0.590	0.632	0.590	0.646
EMO.P	**0.935**	**0.954**	**0.944**	**0.951**	**0.930**	**0.940**	**0.942**	**0.942**

Table 4. Relative differences of the overall accuracy of our considered approaches, benchmarked against one another on our collection of Dutch tweets and forum posts. Performance differences marked with * are statistically significant at $p < 0.05$, those marked with ** are significant at $p < 0.01$, and those marked with *** are significant at $p < 0.001$.

Benchmark	BASELINE	EMO.S	EMO.P
BASELINE	0.000	1.697***	3.308***
EMO.S	-0.629***	0.000	0.597***
EMO.P	-0.768***	-0.374***	0.000

Table 5. Relative differences of the macro-level F_1-score of our considered approaches, benchmarked against one another on our collection of Dutch tweets and forum posts. Performance differences marked with * are statistically significant at $p < 0.05$, those marked with ** are significant at $p < 0.01$, and those marked with *** are significant at $p < 0.001$.

Benchmark	BASELINE	EMO.S	EMO.P
BASELINE	0.000	1.955***	3.310***
EMO.S	-0.662***	0.000	0.458***
EMO.P	-0.768***	-0.314***	0.000

4.2 Experimental results on Dutch tweets and forum posts

On our collection of Dutch tweets and forum posts, our considered polarity classification approaches exhibit clear differences in terms of performance, as demonstrated in Tables 3, 4, and 5. The absolute baseline of not accounting for the sentiment conveyed by emoticons (BASELINE) is outperformed by both considered methods of harvesting information from emoticons for the sentiment analysis process. Overall, sentence-level accounting for emoticon sentiment (EMO.S) yields a statistically significant increase in accuracy and macro-level F_1 from 22% to 59% and from 22% to 65%, respectively. Assuming the sentiment conveyed by emoticons to affect the surrounding text on a paragraph level (EMO.P) significantly increases both overall polarity classification accuracy and macro-level F_1 even further to 94%.

The performance differences between our novel EMO.S and EMO.P methods suggest that, in our set of Dutch tweets and forum posts, the scope of influence of the sentiment conveyed by emoticons is not always limited to the surrounding text on a sentence level, but may extend to the surrounding text on a paragraph level as well. In general, assuming a paragraph-level influence of emoticons yields significantly better polarity classification performance than assuming a sentence-level influence. Nevertheless, both approaches work significantly better than the BASELINE approach, which assumes no influence of emoticons at all.

The comparably weak performance of our BASELINE method suggests that in our Dutch corpus, emoticons do not often stress sentiment that is already conveyed by sentiment-carrying words. Conversely, the authors of our tweets and forum posts mostly use emoticons to express or disambiguate their sentiment. This holds for positive texts, as well as for negative ones.

Table 6. Performance measures of all of our considered approaches on our collection of English app reviews. The best performance is printed in bold for each performance measure.

	Positive			Negative			Overall	
Method	Precision	Recall	F_1	Precision	Recall	F_1	Accuracy	F_1
Baseline	0.788	0.856	0.821	0.587	0.469	0.521	0.739	0.730
Emo.S	0.793	0.859	0.825	0.600	0.484	0.536	0.746	0.737
Emo.P	**0.794**	**0.860**	**0.826**	**0.602**	**0.486**	**0.538**	**0.747**	**0.738**

Table 7. Relative differences of the overall accuracy of our considered approaches, benchmarked against one another on our collection of English app reviews. Performance differences marked with * are statistically significant at $p < 0.05$, those marked with ** are significant at $p < 0.01$, and those marked with *** are significant at $p < 0.001$.

Benchmark	Baseline	Emo.S	Emo.P
Baseline	0.000	0.009***	0.010***
Emo.S	-0.009***	0.000	0.001*
Emo.P	-0.010***	-0.001*	0.000

Table 8. Relative differences of the macro-level F_1-score of our considered approaches, benchmarked against one another on our collection of English app reviews. Performance differences marked with * are statistically significant at $p < 0.05$, those marked with ** are significant at $p < 0.01$, and those marked with *** are significant at $p < 0.001$.

Benchmark	Baseline	Emo.S	Emo.P
Baseline	0.000	0.010***	0.011***
Emo.S	-0.010***	0.000	0.001*
Emo.P	-0.011***	-0.001*	0.000

As such, in our Dutch texts, emoticons are crucial proxies for people's sentiment, as they often capture sentiment that cannot typically be inferred from the sentiment-carrying words used in our texts. This confirms that accounting for the sentiment conveyed by emoticons is a viable strategy when performing sentiment analysis of text.

4.3 Validation on English app reviews

In order to validate our findings presented in Section 4.2, we have assessed the performance of our considered polarity classification approaches on a collection of documents in another language, covering another domain. Tables 6, 7, and 8 present the performance of our methods on this collection of English app reviews, some of which contain emoticons.

Accounting for emoticons in the polarity classification process has a small, yet significant effect on the polarity classification performance when considering all 10,069 documents in this corpus. The overall accuracy significantly increases with about 1% from 74% for the Baseline approach to 75% for both considered emoticon-guided polarity classification methods. Similarly, the macro-level F_1-score exhibits a significant increase of approximately 1% from 73% for the Baseline method to 74% for our novel Emo.S and Emo.P approaches. The differences between the latter emoticon-guided polarity classification approaches are very small, yet statistically significant, on the full data set.

The observed performance improvements of our emoticon-guided approaches, compared to the emoticon-ignoring Baseline method, are mainly driven by improved polarity classification performance on the negative documents in our collection of English app reviews.

Table 9. Performance measures of all of our considered approaches on English app reviews that contain emoticons. The best performance is printed in bold for each performance measure.

	Positive			Negative			Overall	
Method	Precision	Recall	F_1	Precision	Recall	F_1	Accuracy	F_1
BASELINE	0.867	0.863	0.865	0.446	0.453	0.450	0.783	0.784
EMO.S	0.952	0.903	0.927	0.671	0.813	0.735	0.885	0.889
EMO.P	**0.964**	**0.911**	**0.937**	**0.701**	**0.859**	**0.772**	**0.901**	**0.904**

Table 10. Relative differences of the overall accuracy of our considered approaches, benchmarked against one another on English app reviews that contain emoticons. Performance differences marked with * are statistically significant at $p < 0.05$, those marked with ** are significant at $p < 0.01$, and those marked with *** are significant at $p < 0.001$.

Benchmark	BASELINE	EMO.S	EMO.P
BASELINE	0.000	0.131***	0.150***
EMO.S	-0.116***	0.000	0.017*
EMO.P	-0.131***	-0.017*	0.000

Table 11. Relative differences of the macro-level F_1-score of our considered approaches, benchmarked against one another on English app reviews that contain emoticons. Performance differences marked with * are statistically significant at $p < 0.05$, those marked with ** are significant at $p < 0.01$, and those marked with *** are significant at $p < 0.001$.

Benchmark	BASELINE	EMO.S	EMO.P
BASELINE	0.000	0.135***	0.154***
EMO.S	-0.119***	0.000	0.017*
EMO.P	-0.133***	-0.017*	0.000

Whereas the performance of our BASELINE method may be thwarted by people's tendency of writing negative texts with rather positive words [7, 17, 49], the EMO.S and EMO.P methods compensate for this possible bias towards positivity by harvesting crucial information from emoticons. These results support our underlying assumption of the sentiment conveyed by nonverbal cues (i.e., emoticons) dominating the sentiment conveyed by verbal cues – especially for the negative app reviews in our corpus, emoticons appear to play a crucial role in expressing or disambiguating an author's sentiment.

As only about 7% of the documents in our collection of English app reviews contain emoticons, the nevertheless significant differences in terms of polarity classification performance of our proposed emoticon-guided methods are rather small on the full data set. Tables 9, 10, and 11 demonstrate more apparent differences in performance on the 655 app reviews that contain emoticons.

The performance of our considered approaches on the documents containing emoticons in our collection of English app reviews exhibits a pattern that is similar to our findings on our collection of Dutch tweets and forum posts containing emoticons, as presented in Section 4.2. Our novel polarity classification method that accounts for emoticons on a sentence level (EMO.S) significantly outperforms the BASELINE method with about 13% (in relative terms), with an overall accuracy and a macro-level F_1-score increasing from about 78% to approximately 89%. Accounting for emoticons on a paragraph level (EMO.P) yields a further, statistically significant, relative increase of the overall polarity classification performance with about 2%, with an overall accuracy and a macro-level F_1-score amounting to more than 90%.

Consequently, as is the case for our Dutch corpus, assuming a paragraph-level influence of emoticons is to be preferred over assuming a sentence-level influence in our collection of English app reviews. Additionally, both emoticon-guided polarity classification approaches significantly outperform our baseline of not accounting for emoticons at all.

Interestingly, in contrast with its observed weak performance on our collection of Dutch tweets and forum posts containing emoticons, the BASELINE method performs rather well on our English app reviews that contain emoticons. This suggests that in most of the latter documents, emoticons convey sentiment that is conveyed by the sentiment-carrying words in these documents as well. As such, emoticons form an equally good proxy for an author's sentiment as the sentiment-carrying words in such app reviews. Nevertheless, emoticons play a crucial role in a subset of our English app reviews containing emoticons, where the main purpose of these emoticons is to express or disambiguate an author's intended sentiment. This mainly holds for most of the negative reviews, where the authors have a tendency of using rather positive sentiment-carrying words and negative emoticons in order to convey their negative sentiment. Properly accounting for the sentiment conveyed by emoticons in such cases yields a significantly improved overall polarity classification performance, thus confirming that our proposed method of accounting for the sentiment conveyed by emoticons is not only a viable strategy in our initial corpus of Dutch tweets and forum posts, but in our validation corpus as well.

4.4 Caveats

Experiments in competitions for sentiment analysis, such as the SemEval 2007 Task 14 on Affective Text [47], have shown how difficult it is to extract the valence (i.e., sentiment) of natural language text for both supervised and unsupervised sentiment analysis approaches, which currently lag behind the performance of the inter-annotator agreement for valence. In this light, our results clearly indicate that considering emoticons when analyzing the sentiment conveyed by natural language text appears to be a fruitful addition to the state-of-the-art of (lexicon-based) sentiment analysis. Our results suggest that whenever emoticons are used, these visual cues play an important, if not crucial role in conveying an author's sentiment. The sentiment conveyed by emoticons tends to dominate the sentiment conveyed by verbal cues in both of our considered corpora. As such, emoticons have proven to be helpful indicators of intended sentiment.

However, some issues still remain to be solved. One source of polarity classification errors lies in the interpretation of natural language text by human readers and the preference of these readers for certain aspects of a text over others. Consider, for example, the message "*Interesting product =D Just not for me... =/*". This message would receive a sentiment score of 0 when using our framework, as the emoticons cancel each other out in this particular piece of text. However, in the annotation process of our collection of Dutch tweets and forum messages, our three human annotators initially did not typically agree on the polarity of fragments like this as a whole. All three human interpreters turned out to deem one part of such a fragment to be more important for conveying the overall sentiment than the other part, even though they initially did not agree on which part was crucial for the polarity of the fragment. Conversely, for our framework, each part of a text contributes equally to conveying the overall sentiment of the text.

Another source of errors can be nicely illustrated when analyzing the polarity conveyed by our English app reviews. The reviews in our corpus often start with a description of the app. These descriptions may already contain sentiment-carrying words, whereas the writer is not yet expressing his or her own opinion at that stage of the review. Apparently, aspects other than sentiment-carrying words and emoticons, such as their positioning or rhetorical role [17], may be worthwhile exploiting in sentiment analysis.

5 Conclusions

As people increasingly use emoticons in their utterances of opinions through the Social Web, it is of paramount importance for automated polarity classification tools to correctly interpret these graphical cues for sentiment. Our key contribution lies in our analysis of the role that emoticons typically play in conveying a text's overall sentiment, as well as in the proposal and evaluation of our novel method for exploiting emoticons in lexicon-based polarity classification.

Whereas emoticons have until now been considered to be used in a way similar to how textual cues for sentiment are used [51], the qualitative analysis presented in our current paper demonstrates that the sentiment associated with emoticons typically dominates the sentiment conveyed by textual cues in a text segment. The results of our analysis indicate that people typically use emoticons in natural language text in order to express, stress, or disambiguate their sentiment in particular text segments, thus rendering them potentially better local proxies for people's intended overall sentiment than textual cues.

On a collection of Dutch tweets and forum messages, as well as on another collection of English app reviews, we find that accounting for the sentiment conveyed by emoticons on a paragraph level – and, to a lesser extent, on a sentence level – significantly improves the performance of a lexicon-based polarity classifier. Our findings suggest that whenever emoticons are used, their associated sentiment dominates the sentiment conveyed by textual cues and forms a good proxy for the polarity of text.

As our results are very promising, we envisage several directions for future work. First, we would like to further explore and exploit the interplay of emoticons and textual cues for sentiment, for instance in cases when emoticons are used to intensify sentiment that is already conveyed by the text. Another possible direction for future research would be exploiting structural and semantic aspects of text, e.g., the rhetorical roles of text segments, in order to be able to differentiate between important and less important text segments in emoticon-guided polarity classification.

Acknowledgments

The authors of this paper are partially supported by the Dutch national program COMMIT. Special thanks go to Teezir (`http://www.teezir.com`) for their fruitful discussions and for supplying us with Dutch data and sentiment analysis software components for this research.

References

1. S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *7th Conference on International Language Resources and Evaluation (LREC 2010)*, pages 2200–2204. European Language Resources Association, 2010.

2. D. Bal, M. Bal, A. van Bunningen, A. Hogenboom, F. Hogenboom, and F. Frasincar. Sentiment Analysis with a Multilingual Pipeline. In *12th International Conference on Web Information System Engineering (WISE 2011)*, volume 6997 of *Lecture Notes in Computer Science*, pages 129–142. Springer, 2011.
3. J. Baldridge and T. Morton. OpenNLP, 2004. Available online, `http://opennlp.sourceforge.net/`.
4. J. Burgoon, D. Buller, and W. Woodall. *Nonverbal Communication: The Unspoken Dialogue*. McGraw-Hill, 2nd edition, 1996.
5. P. Casoto, A. Dattolo, and C. Tasso. Sentiment Classification for the Italian Language: A Case Study on Movie Reviews. *Journal of Internet Technology*, 9(4):365–373, 2008.
6. C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sagoff, and V. Subrahmanian. OASYS: An Opinion Analysis System. In *AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs (CAAW 2006)*, pages 21–26. Association for the Advancement of Artificial Intelligence, 2006.
7. J. Chenlo, A. Hogenboom, and D. Losada. Sentiment-Based Ranking of Blog Posts using Rhetorical Structure Theory. In *18th International Conference on Applications of Natural Language to Information Systems (NLDB 2013)*, volume 7934 of *Lecture Notes in Computer Science*, pages 13–24. Springer, 2013.
8. J. Chenlo and D. Losada. Effective and Efficient Polarity Estimation in Blogs Based on Sentence-Level Evidence. In *20th ACM Conference on Information and Knowledge Management (CIKM 2011)*, pages 365–374. Association for Computing Machinery, 2011.
9. T. Childers and M. Houston. Conditions for a Picture-Superiority Effect on Consumer Memory. *Journal of Consumer Research*, 11(2):643–654, 1984.
10. ComputerUser. Emoticons, 2013. Available online, `http://www.computeruser.com/resources/dictionary/emoticons.html`.
11. D. Davidov, O. Tsur, and A. Rappoport. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. In *23rd International Conference on Computational Linguistics: Posters (COLING 2010)*, pages 241–249. Association for Computational Linguistics, 2010.
12. A. Devitt and K. Ahmad. Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. In *45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*, pages 984–991. Association for Computational Linguistics, 2007.
13. X. Ding, B. Lu, and P. Yu. A Holistic Lexicon-Based Approach to Opinion Mining. In *1st ACM International Conference on Web Search and Web Data Mining (WSDM 2008)*, pages 231–240. Association for Computing Machinery, 2008.
14. R. Feldman. Techniques and Applications for Sentiment Analysis. *Communications of the ACM*, 56(4):82–89, 2013.
15. P. Gil. Emoticons and Smileys 101, 2013. Available online, `http://netforbeginners.about.com/cs/netiquette101/a/bl_emoticons101.htm`.
16. P. Goncalves, M. Araujo, F. Benevenuto, and M. Cha. Comparing and Combining Sentiment Analysis Methods. In *1st ACM Conference on Online Social Networks (COSN 2013)*, pages 27–38. Association for Computing Machinery, 2013.
17. B. Heerschop, F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong. Polarity Analysis of Texts using Discourse Structure. In *20th ACM Conference on Information and Knowledge Management (CIKM 2011)*, pages 1061–1070. Association for Computing Machinery, 2011.
18. B. Heerschop, A. Hogenboom, and F. Frasincar. Sentiment Lexicon Creation from Lexical Resources. In *14th International Conference on Business Information Systems (BIS 2011)*, volume 87 of *Lecture Notes in Business Information Processing*, pages 185–196. Springer, 2011.
19. B. Heerschop, P. van Iterson, A. Hogenboom, F. Frasincar, and U. Kaymak. Analyzing Sentiment in a Large Set of Web Data while Accounting for Negation. In *7th Atlantic Web Intelligence Conference (AWIC 2011)*, pages 195–205. Springer, 2011.
20. A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak. Exploiting Emoticons

in Sentiment Analysis. In *28th Symposium on Applied Computing (SAC 2013)*, pages 703–710. Association for Computing Machinery, 2013.

21. A. Hogenboom, M. Bal, F. Frasincar, D. Bal, U. Kaymak, and F. de Jong. Lexicon-Based Sentiment Analysis by Mapping Conveyed Sentiment to Intended Sentiment. *International Journal of Web Engineering and Technology*, 9(1), 2014. To Appear.
22. A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, and F. de Jong. Multi-Lingual Support for Lexicon-Based Sentiment Analysis Guided by Semantics. *Decision Support Systems*, Online First (DOI: 10.1016/j.dss.2014.03.004), 2014.
23. A. Hogenboom, F. Hogenboom, U. Kaymak, P. Wouters, and F. de Jong. Mining Economic Sentiment using Argumentation Structures. In *7th International Workshop on Web Information Systems Modeling (WISM 2010) at the 29th International Conference on Conceptual Modeling (ER 2010)*, volume 6413 of *Lecture Notes in Computer Science*, pages 200–209. Springer, 2010.
24. A. Hogenboom, P. van Iterson, B. Heerschop, F. Frasincar, and U. Kaymak. Determining Negation Scope and Strength in Sentiment Analysis. In *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011)*, pages 2589–2594. IEEE, 2011.
25. B. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, 2009.
26. A. Kendon. On Gesture: Its Complementary Relationship with Speech. In *Nonverbal Communication*. Lawrence Erlbaum Associates, 1987.
27. S. Kim and E. Hovy. Automatic Identification of Pro and Con Reasons in Online Reviews. In *21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL 2006)*, pages 483–490. Association for the Advancement of Artificial Intelligence, 2006.
28. B. Liu. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
29. K. Liu, W. Li, and M. Guo. Emoticon Smoothed Language Models for Twitter Sentiment Analysis. In *26th AAAI Conference on Artificial Intelligence (AAAI 2012)*, pages 1678–1684. Association for the Advancement of Artificial Intelligence, 2012.
30. C. Manning, T. Grow, T. Grenager, J. Finkel, and J. Bauer. Stanford Tokenizer, 2010. Available online, `http://nlp.stanford.edu/software/tokenizer.shtml`.
31. T. Marks. Recommended Emoticons for Email Communication, 2004. Available online, `http://www.windweaver.com/emoticon.htm`.
32. J. Marshall. The Canonical Smiley (and 1-Line Symbol) List, 2003. Available online, `http://www.astro.umd.edu/~marshall/smileys.html`.
33. L. Marvin. Spoof, Spam, Lurk, and Lag: The Aesthetics of Text-Based Virtual Realities. *Journal of Computer-Mediated Communication*, 1(2), 1995.
34. P. Melville, V. Sindhwani, and R. Lawrence. Social Media Analytics: Channeling the Power of the Blogosphere for Marketing Insight. In *1st Workshop on Information in Networks (WIN 2009)*, 2009.
35. R. Mihalcea, C. Banea, and J. Wiebe. Learning Multilingual Subjective Language via Cross-Lingual Projections. In *45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, pages 976–983. Association for Computational Linguistics, 2007.
36. Msgweb. List of Emoticons in MSN Messenger, 2006. Available online, `http://www.msgweb.nl/en/MSN_Images/Emoticon_list/`.
37. R. Navigli. Word Sense Disambiguation: A Survey. *ACM Computing Surveys*, 41(2):1–69, 2009.
38. B. Ojokoh and O. Kayode. A Feature-Opinion Extraction Approach to Opinion Mining. *Journal of Web Engineering*, 11(1):51–63, 2012.
39. A. Pak and P. Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *7th Conference on International Language Resources and Evaluation (LREC 2010)*, pages 1320–1326. European Language Resources Association, 2010.
40. G. Paltoglou and M. Thelwall. A Study of Information Retrieval Weighting Schemes for Sentiment

Analysis. In *48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pages 1386–1395. Association for Computational Linguistics, 2010.
41. B. Pang and L. Lee. A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts. In *42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, pages 271–280. Association for Computational Linguistics, 2004.
42. B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1):1–135, 2008.
43. J. Read. Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification. In *Student Research Workshop at the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pages 43–48. Association for Computational Linguistics, 2005.
44. L. Rezabek and J. Cochenour. Visual Cues in Computer-Mediated Communication: Supplementing Text with Emoticons. *Journal of Visual Literacy*, 18(2):201–215, 1998.
45. Sharpened. Text-Based Emoticons, 2013. Available online, `http://www.sharpened.net/emoticons/`.
46. R. Shepard. Recognition Memory for Words, Sentences, and Pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1):156–163, 1967.
47. C. Strappavara and R. Mihalcea. SemEval-2007 Task 14: Affective Text. In *4th International Workshop on Semantic Evaluations (SemEval 2007)*, pages 70–74. Association for Computational Linguistics, 2007.
48. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2):267–307, 2011.
49. M. Taboada, K. Voll, and J. Brooke. Extracting Sentiment as a Function of Discourse Structure and Topicality. Technical Report 20, Simon Fraser University, 2008. Available online, `http://www.cs.sfu.ca/research/publications/techreports/#2008`.
50. M. Thelwall, K. Buckley, and G. Paltoglou. SentiStrength, 2011. Available online, `http://sentistrength.wlv.ac.uk/`.
51. M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.
52. B. Walenz and J. Didion. OpenNLP, 2008. Available online, `http://jwordnet.sourceforge.net/`.
53. C. Whitelaw, N. Garg, and S. Argamon. Using Appraisal Groups for Sentiment Analysis. In *14th ACM International Conference on Information and Knowledge Management (CIKM 2005)*, pages 625–631. Association for Computing Machinery, 2005.
54. Wikipedia. List of Emoticons, 2013. Available online, `http://en.wikipedia.org/wiki/List_of_emoticons/`.
55. D. Witmer and S. Katzman. On-Line Smiles: Does Gender Make a Difference in the Use of Graphic Accents? *Journal of Computer-Mediated Communication*, 2(4), 1997.
56. J. Zhao, L. Dong, J. Wu, and K. Xu. MoodLens: An Emoticon-Based Sentiment Analysis System for Chinese Tweets. In *18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012)*, pages 1528–1531. Association for Computing Machinery, 2012.