Exploiting Emoticons in Sentiment Analysis

Alexander Hogenboom[1]
hogenboom@ese.eur.nl

Daniella Bal[1]
daniella.bal@xs4all.nl

Flavius Frasincar[1]
frasincar@ese.eur.nl

Malissa Bal[1]
malissa.bal@xs4all.nl

Franciska de Jong[1,2]
f.m.g.dejong@utwente.nl

Uzay Kaymak[3]
u.kaymak@ieee.org

[1]Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands
[2]Universiteit Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands
[3]Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands

ABSTRACT

As people increasingly use emoticons in text 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 demonstrate how we can exploit this by using a novel, manually created emoticon sentiment lexicon in order to improve a state-of-the-art lexicon-based sentiment classification method. We evaluate our approach on 2,080 Dutch tweets and forum messages, which all contain emoticons and have been manually annotated for sentiment. On this corpus, paragraph-level accounting for sentiment implied by emoticons significantly improves sentiment classification accuracy. This indicates that whenever emoticons are used, their associated sentiment dominates the sentiment conveyed by textual cues and forms a good proxy for intended sentiment.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Linguistic processing*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Language parsing and understanding*

General Terms

Algorithms, experimentation, performance

Keywords

Sentiment analysis, emoticons, sentiment lexicon

1. INTRODUCTION

Today's Web enables users to produce an ever-growing amount of utterances of opinions. People can write blogs and reviews, post messages on discussion forums, and publish whatever crosses their minds on Twitter in a trice. This phenomenon 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 recent estimates indicate that one in three blog posts [18] and one in five tweets [14] discuss products or brands, the abundance of user-generated content published through such social media renders automated information monitoring tools crucial for today's businesses.

Sentiment analysis comes to answer this need. Sentiment analysis refers to a broad area of natural language processing, computational linguistics, and text mining. Typically, the goal is to determine the polarity of natural language texts. An intuitive approach would involve scanning a text for cues signaling its polarity.

In face-to-face communication, sentiment can often be deduced from visual cues like 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 computer-mediated communication like virtual utterances of opinions. In this light, we define emoticons as visual cues used in texts to replace normal visual cues like smiling to express, stress, or disambiguate one's sentiment. Emoticons are typically made up 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).

In recent years, several approaches to sentiment analysis of natural language text have been proposed. Many state-of-the-art approaches represent text as a bag of words, i.e., an unordered collection of the words occurring in a 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 a different domain [34]. In this light, alternative lexicon-based methods have gained (renewed) attention in recent research [4, 7, 8, 9, 11, 33], not in the least because they have been shown to have a more robust performance across domains and texts [34]. These methods tend to keep a more linguistic view on textual data rather than abstracting away from natural language by means of vectorization.

SAC'13 March 18-22, 2013, Coimbra, Portugal.

As such, deep linguistic analysis comes more naturally in lexicon-based approaches, thus allowing for intuitive ways of accounting for structural or semantic aspects of text in sentiment analysis [9].

Lexicon-based sentiment analysis approaches use sentiment lexicons for retrieving the polarity of individual words and aggregate these scores in order to determine the text's 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].

However, today's lexicon-based approaches typically do not consider emoticons. Conversely, one of the first steps in most existing work is to remove many of the typographical symbols typically constituting emoticons, thus preventing emoticons from being detected at all. Yet, state-of-the-art sentiment analysis approaches may be ignoring important information, as an emoticon may for instance signal the intended sentiment of an otherwise objective statement, e.g., "*This product does not work :-(*". Therefore, we aim to investigate how emoticons are typically used to convey sentiment and how we can exploit this in order to improve the state-of-the-art of lexicon-based sentiment analysis.

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 sentiment of the text they occur in and we additionally propose a method for harvesting information from emoticons when analyzing the sentiment of natural language 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 recent literature survey on sentiment analysis [26], the current surge of research interest in systems that deal with opinions and sentiment is attributed to the fact that, despite 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

As sentiment analysis tools have particularly useful applications in marketing and reputation management [10], sentiment analysis tools are often evaluated on collections of reviews, which typically contain people's opinions expressed in natural language, often along with an associated (numeric) score quantifying one's judgment. In this light, a widely used corpus for assessing sentiment analysis approaches is a collection of 2,000 English movie reviews, annotated for sentiment [25].

Among the popular bag-of-word approaches, a binary representation of text, indicating the presence of specific words, has initially proven to be an effective approach, yielding an accuracy of 87.2% on the movie review data [25]. Later research has focused on different vector representations of text, including vector representations with additional features representing semantic distinctions between words [36] or vector representations with *tf-idf*-based weights for word features [24]. Such approaches typically yield an accuracy on the movie review data set of over 90.0%.

The alternative lexicon-based approaches typically exhibit lower accuracy on the movie review data set, but tend to be more robust across domains [34]. Also, lexicon-based approaches can be generalized relatively easily to other languages by using dictionaries [19]. A rather simple lexicon-based sentiment analysis framework has been shown to have an accuracy up to 59.5% on the full movie review data set [10]. A more sophisticated lexicon-based sentiment analysis approach has been shown to have an average accuracy of 68.0% on 1,900 documents from the movie review data set [33]. A deeper linguistic analysis focusing on differentiating between rhetorical roles of text segments has recently been proven to perform comparably well too [9]. On 1,000 documents from the movie review data set, this approach yields an accuracy of 72.0%, which is a 4.5% improvement over not accounting for structural aspects of content.

Even though recent lexicon-based sentiment analysis approaches explore promising new directions of incorporating structural and semantic aspects of content [9, 12], 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 [27]. These early results suggest that the polarity information conveyed by emoticons is topic- and domain-independent. These findings have been successfully applied in later work in order to automatically construct sets of positive and negative tweets [23].

In more recent research, a small set of emoticons has been used as features for polarity classification [35]. However, the results 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. 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 sentiment.

To the best of our knowledge, existing research however does not focus on investigating how emoticons affect the sentiment of natural language text, nor on exploring how this phenomenon can be exploited in lexicon-based sentiment analysis. In order to be able to address this hiatus, we need to 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 [5, 30]. Nonverbal cues have even been shown to dominate verbal cues in face-to-face communication in case both types of cues are equally strong [3]. Apparently, nonverbal cues are deemed important indicators for people in order to understand the intentions and emotions of whoever they are communicating with. Translating these findings to computer-mediated communication does hence not seem 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 ways to overcome 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 "*:-)*" and "*:-(*" to 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 [17]. By then, 6.1% of the messages on electronic mailing lists [28] and 13.2% of UseNet newsgroup posts [38] were estimated to contain emoticons.

Thus, nonverbal cues have emerged in computer-mediated communication. These cues are however conceptually different from nonverbal cues in face-to-face communication – cues like laughing and weeping are often referred to as involuntary ways of expressing oneself in face-to-face communication, whereas the use of their respective equivalents "*:-)*" and "*:-(*" in computer-mediated communication is intentional [15]. 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 analyzing a text for sentiment.

3. EMOTICONS AND SENTIMENT

In order to exploit emoticons in automated sentiment analysis, we first need to analyze how emoticons are typically related to the sentiment 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 sentiment analysis by harvesting information from emoticons.

3.1 Emoticons as Cues for Sentiment

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.

First, we hypothesize that emoticons have a rather local effect, i.e., they affect a paragraph or a sentence. Paragraphs typically address different points of view for a single topic or different topics, thus rendering the applicability of an emoticon in one paragraph to another paragraph rather unlikely. In our sample collection, upon inspection, emoticons generally have a paragraph-level effect for paragraphs containing only one emoticon. When a paragraph contains multiple emoticons, our sample shows that an emoticon is generally more likely to affect the sentence in which it occurs.

Interestingly, in our sample, 84.0% of all emoticons are placed at the end of a paragraph, 9.0% are positioned somewhere in the middle of a paragraph, and 7.0% 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.

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

Rather than only looking into *what* is affected by emoticons, we have also assessed *how* emoticons affect text. 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 the sentence 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 shows that the sentiment associated with a sentence 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 [21] not useful for emoticons. As such, 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 which 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 typically be inclined to acknowledge the implied sentiment and thus consider the factual statement to be a positive statement. This however requires an understanding of context and involves incorporating real-world knowledge into the process of sentiment analysis. For machines, this is a cumbersome task. In this light, emoticons can be valuable cues for deriving an author's intended sentiment.

3.2 Framework

We propose a novel framework for automated sentiment analysis, which takes into account the information conveyed by emoticons. The goal of this framework is to detect emoticons, determine their sentiment, and assign the associated sentiment to the affected text in order to correctly classify the polarity of natural language text as either positive or

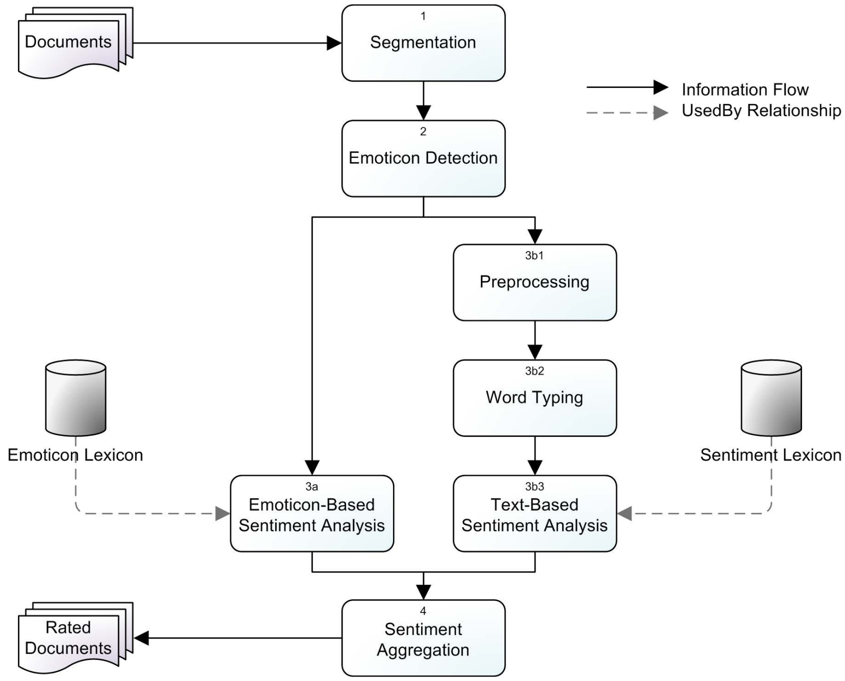

Figure 1: Overview of our sentiment analysis framework.

negative. In order to accomplish this, we build upon existing work [2]. Our framework, depicted in Figure 1, is essentially a pipeline in which each component fulfills a specific task in analyzing the sentiment of an arbitrary document. Here, a document is a piece of natural language text which can be as small as a one-line tweet or as big as a news article, blog, or forum message with multiple paragraphs, as long as it is one coherent piece of text.

First, we load a document in order for it to be analyzed for sentiment. Then, the document is split into text segments, which may be either paragraphs or sentences (step 1). When splitting a document into paragraphs, we look for empty lines or lines starting with an indentation. 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). Sentiment analysis is subsequently initially performed on segment level, after which the segment-level results are combined.

Each text segment is checked for the presence of emoticons (step 2). To this end, we propose an emoticon sentiment lexicon, which we define as a list of character sequences, representing 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. When checking a text segment for the presence of emoticons, we compare each word in the segment with 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 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 average sentiment associated with its emoticons, as derived from the emoticon sentiment lexicon. Sentiment scores of sentiment-carrying words (if any) are ignored in this process, as our analysis presented in Section 3.1 indicates that the sentiment of emoticons tends to dominate the sentiment carried by verbal cues.

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 and its purpose in the text, i.e., sentiment-carrying or modifying term (step 3b2). Following existing work [2], we consider modifying terms to change the sentiment of corresponding word(s) – negations change the sentiment sign and amplifiers increase the sentiment of the affected sentiment words. After determining the word types, the text segment is rated for its conveyed sentiment by means of a lexicon-based sentiment scoring method [2] that essentially computes the sentiment of the text segment as the average sentiment score of all sentiment-carrying words in the segment (step 3b3).

As such, the sentiment score $\text{sent}(s_i)$ of the i-th segment s_i of document d can be computed as a function of the sentiment scores of either each emoticon e_{ij} in segment s_i or each sentiment-carrying word w_{ij} and its modifier m_{ij}, (if any, else this modifier defaults to 1), i.e.,

$$\text{sent}(s_i) = \begin{cases} \frac{\sum_{j=1}^{v_i} \text{sent}(e_{ij})}{v_i} & \text{if } v_i > 0, \\ \frac{\sum_{j=1}^{t_i} (\text{sent}(w_{ij}) \cdot \text{sent}(m_{ij}))}{t_i} & \text{else,} \end{cases} \tag{1}$$

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, if any) in the segment.

Table 2: Typical examples of emoticon synsets.

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

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 document sentiment score is then calculated as a weighted average of all segment-level sentiment scores, where the weights correspond with the relative proportions of the number of sentiment-carrying words or emoticons (whichever is applicable) in each respective segment (step 4). As such, the sentiment score $\text{sent}(d)$ of a document d is calculated as

$$\text{sent}(d) = \frac{\sum_{i=1}^{p} \left(\text{sent}(s_i) \cdot (v_i + (a_i \cdot t_i))\right)}{\sum_{i=1}^{p} \left(v_i + (a_i \cdot t_i)\right)}, \tag{2}$$

with p the number of partitions of document d and a_i a Boolean variable 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.,

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

Thus, the document's sentiment score is returned. A negative score typically indicates a negative document (-1), whereas other scores yield a positive classification (1). The classification $\text{class}(d)$ of document d is therefore defined as a function of its sentiment score $\text{sent}(d)$, i.e.,

$$\text{class}(d) = \begin{cases} 1 & \text{if sent}(d) \geq 0, \\ -1 & \text{else.} \end{cases} \tag{4}$$

4. POLARITY CLASSIFICATION BY EXPLOITING EMOTICONS

Our novel method of classifying natural language text in terms of its polarity by exploiting emoticons is evaluated by means of a set of experiments. For our current purpose, we focus on a test collection of Dutch documents. This test collection consists of 2,080 Dutch tweets and forum messages (1,067 positive documents and 1,013 negative documents), which have been manually annotated for sentiment by three human annotators until they reached agreement. We have randomly sampled these messages from search results from Twitter and Google discussion groups when querying for the brands Vodafone, KLM, Kinect, etcetera. Emoticons occur in all of our considered documents.

4.1 Experimental Setup

One of the key elements in our novel framework is the emoticon sentiment lexicon. Several lists of emoticons are readily available [6, 13, 16, 20, 22, 29, 32, 37]. We propose to combine these eight existing lists into one large lexicon, while leaving out duplicate entries, 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 574 emoticons representing facial expressions or body poses like thumbs up. We have let three human annotators manually rate the emoticons in our lexicon for their associated sentiment. The annotators were allowed to assign ratings of -1.0 (negative), -0.5, 0.0, 0.5, and 1.0 (positive). The sentiment score of each individual emoticon has subsequently been determined as the score closest to the average of the annotators' scores for that particular emoticon. In 87.5% of all cases, our three annotators assigned identical scores to the respective emoticons.

The sentiment lexicon thus generated is utilized in the C# implementation of our framework. In our implementation, 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`). Our implementation can perform both paragraph-level and sentence-level sentiment analysis and the design of its graphical user interface, depicted in Figure 2 facilitates the comparison between sentiment analysis with and without taking into account the information conveyed by emoticons.

The implementation of our proposed framework allows us to perform a set of experiments in order to compare the performance of several configurations of our sentiment analysis framework. First, as an absolute baseline, we assess the performance of our framework when not accounting for the information conveyed by emoticons, thus essentially reducing the functionality of our pipeline to that of a state-of-the-art lexicon-based document-level sentiment analysis approach [2]. Then, as a first alternative approach, we consider a sentiment analysis approach in which the sentiment conveyed by emoticons affects the surrounding text on a sentence level. Last, we consider accounting for the sentiment conveyed by emoticons on a paragraph level when analyzing the sentiment of a piece of natural language text.

In order to get a clear view on the impact of accounting for the sentiment conveyed by emoticons in sentiment analysis, we compare the performance of our considered sentiment analysis approaches on our test collection, in which each document contains at least one emoticon. In our comparisons, 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 of 208 documents, 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.

4.2 Experimental Results

Our considered sentiment analysis approaches exhibit clear differences in terms of performance, as demonstrated in Table 3. This table reports precision, recall, and F_1 measure for positive and negative documents containing emoticons separately, as well as the accuracy and macro-level F_1 measure over this set of documents as a whole. 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 measure is the harmonic mean of precision and recall. The macro-level F_1 measure is the average of the F_1 scores of the positive and negative documents. Accuracy is the proportion of correctly classified documents.

Table 3: Experimental results for all approaches on a set of documents containing emoticons.

	Positive			Negative			Overall	
Method	Precision	Recall	F_1	Precision	Recall	F_1	Accuracy	Macro F_1
Baseline	0.21	0.22	0.22	0.23	0.22	0.22	0.22	0.22
Sentence-level	0.65	0.67	0.66	0.59	0.68	0.63	0.59	0.65
Paragraph-level	**0.95**	**0.93**	**0.94**	**0.93**	**0.95**	**0.94**	**0.94**	**0.94**

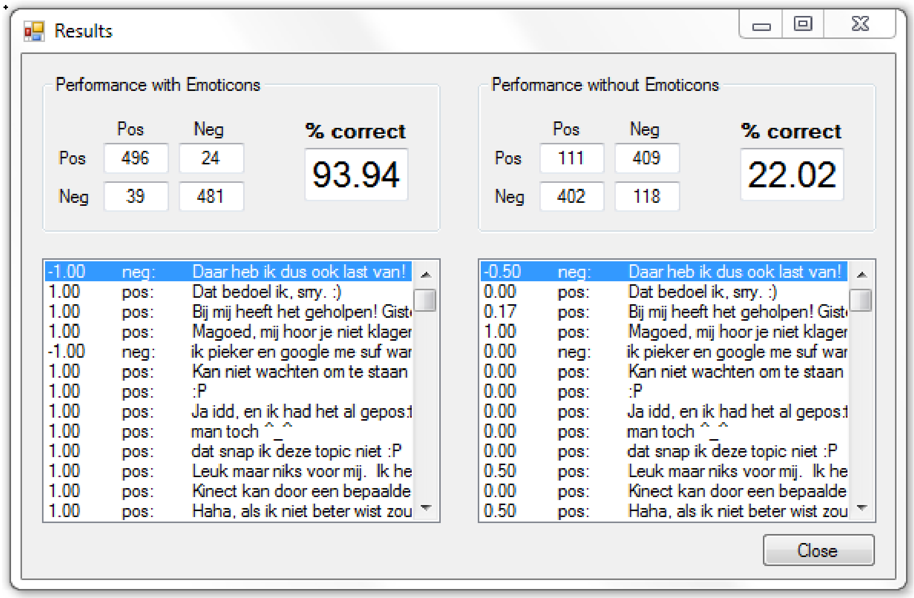

Figure 2: Graphical user interface facilitating comparison of results.

Table 3 clearly shows that on a set of documents containing emoticons, the absolute baseline of not accounting for the information conveyed by emoticons is outperformed by both considered methods of harvesting information from emoticons for the sentiment analysis process. Overall, sentence-level accounting for emoticon sentiment yields an 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 increases both overall polarity classification accuracy and macro-level F_1 even further to 94%. All reported differences in performance are statistically significant at a significance level $p < 0.001$.

Experiments in recent competitions for sentiment analysis, such as the SemEval 2007 Task 14 on Affective Text [31], have shown how difficult it is to extract the valence (sentiment) of text for both supervised and unsupervised 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 sentiment on 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 a crucial role in conveying an author's sentiment.

However, some issues still remain to be solved. One source of polarity classification errors lies in the interpretation of human readers and their preference for certain aspects of a text over others. For instance, the fragment "*The weather is bad :(. I want sunshine!! :)*" 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, before reaching agreement, two out of three annotators initially rated the fragment as positive, whereas one annotator classified the text as carrying negative sentiment. All three human interpreters turned out to deem one part of the 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 movie reviews. The reviews in our corpus often start with a summary of the plot of a movie. Often, these summaries 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, may be worthwhile exploiting in sentiment analysis.

5. CONCLUSIONS

As people increasingly use emoticons in their virtual utterances of opinions, it is of paramount importance for automated sentiment analysis tools to correctly interpret these graphical cues for sentiment. The key contribution of our work lies in our analysis of the role that emoticons typically play in conveying a text's overall sentiment and how we can exploit this in a lexicon-based sentiment analysis method.

Whereas emoticons have until now been considered to be used in a way similar to how textual cues for sentiment are used [35], 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.

In order to validate these findings, we have assessed the performance of a lexicon-based sentiment analysis approach accounting for the sentiment conveyed by emoticons on a collection of 2,080 Dutch tweets and forum messages, with each document containing one or more emoticons. As a baseline, we have considered a similar lexicon-based sentiment analysis approach without support for emoticons. On our data set, accounting for the sentiment implied by emoticons rather than by the textual cues on a paragraph level significantly improves overall document polarity classification accuracy from 22% to 94%, whereas applying our method on a sentence level yields an accuracy of 59%.

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 text, for instance in cases when emoticons are used to intensify sentiment that is already conveyed by the text. Another possible direction for future research includes applying our results in a multilingual context and thus investigating how robust our approach is across languages. Additionally, future research could be focused on other collections of texts in order to verify our findings in, e.g., specific case studies. Last, we would like to exploit structural and semantic aspects of text in order to identify important and less important text spans in emoticon-based sentiment analysis.

6. ACKNOWLEDGMENTS

We would like to thank Teezir (`http://www.teezir.com`) for their technical support, fruitful discussions, and for supplying us with data for this research. The authors of this paper are partially supported by the Dutch national program COMMIT.

7. 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. Burgoon, D. Buller, and W. Woodall. *Nonverbal Communication: The Unspoken Dialogue*. McGraw-Hill, 2nd edition, 1996.

[4] 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. AAAI Press, 2006.

[5] T. Childers and M. Houston. Conditions for a Picture-Superiority Effect on Consumer Memory. *Journal of Consumer Research*, 11(2):643–654, 1984.

[6] ComputerUser. Emoticons, 2011. Available online, `http://www.computeruser.com/resources/dictionary/emoticons.html`.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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 29th International Conference on Conceptual Modeling (ER 2010)*, volume 6413 of *Lecture Notes in Computer Science*, pages 200–209. Springer, 2010.

[13] J. Marshall. The Canonical Smiley (and 1-Line Symbol) List, 2003. Available online, `http://www.astro.umd.edu/~marshall/smileys.html`.

[14] 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.

[15] A. Kendon. On Gesture: Its Complementary Relationship with Speech. In *Nonverbal Communication*. Lawrence Erlbaum Associates, 1987.

[16] M. Thelwall and K. Buckley and G. Paltoglou. SentiStrength, 2011. Available online, `http://sentistrength.wlv.ac.uk/`.

[17] L. Marvin. Spoof, Spam, Lurk, and Lag: The Aesthetics of Text-Based Virtual Realities. *Journal of Computer-Mediated Communication*, 1(2), 1995.

[18] 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.

[19] 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.

[20] Msgweb. List of Emoticons in MSN Messenger, 2006. Available online, `http://www.msgweb.nl/en/MSN_Images/Emoticon_list/`.

[21] R. Navigli. Word Sense Disambiguation: A Survey. *ACM Computing Surveys*, 41(2):1–69, 2009.

[22] P. Gil. Emoticons and Smileys 101, 2011. Available online, `http://netforbeginners.about.com/cs/netiquette101/a/bl_emoticons101.htm`.

[23] 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.

[24] 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.

[25] 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.

[26] B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1):1–135, 2008.

[27] 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.

[28] L. Rezabek and J. Cochenour. Visual Cues in Computer-Mediated Communication: Supplementing Text with Emoticons. *Journal of Visual Literacy*, 18(2):201–215, 1998.

[29] Sharpened. Text-Based Emoticons, 2011. Available online, `http://www.sharpened.net/emoticons/`.

[30] R. Shepard. Recognition Memory for Words, Sentences, and Pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1):156–163, 1967.

[31] 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.

[32] T. Marks. Recommended Emoticons for Email Communication, 2004. Available online, `http://www.windweaver.com/emoticon.htm`.

[33] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2):267–307, 2011.

[34] 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`.

[35] 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.

[36] 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.

[37] Wikipedia. List of Emoticons, 2011. Available online, `http://en.wikipedia.org/wiki/List_of_emoticons/`.

[38] 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.