Sarcasm Detection Using Sentiment Flow Shifts

Elena Filatova
City University of New York
efilatova@citytech.cuny.edu

Abstract

One of the most frequently cited sarcasm realizations is the use of positive sentiment within negative context. We propose a novel approach towards modeling a sentiment context of a document via the sequence of sentiment labels assigned to its sentences. We demonstrate that the sentiment flow shifts (from negative to positive and from positive to negative) can be used as reliable classification features for the task of sarcasm detection. Our classifier achieves the F_1-measure of 0.7 for all reviews, going up to 0.9 for the reviews with high star ratings (positive reviews), which are the reviews that are materially affected by the presence of sarcasm in the text.

Introduction

Verbal irony or sarcasm has been studied by psychologists, linguists, and computer scientists for different types of text: speech, fiction, Twitter messages, Internet dialog, product reviews, etc. Sentiment is widely used as a classification feature for the detection of whether a text snippet or a document is sarcastic or not. The popularity of this feature can be explained by the fact that it is agreed that in many cases sarcasm is manifested in a document via a text snippet with positive sentiment applied to a negative situation. Given that the notion of sarcasm (or verbal irony, or irony for that matter) does not have a formal definition except that in the case of sarcasm/irony a nonsalient interpretation has the priority over a salient one, positive utterance within a negative context is a reliable feature to use (Riloff et al. 2013).

Other features (textual and non-textual) used for the task of identifying sarcastic text are: emoticons (Gonzalez-Ibáñez, Muresan, and Wacholder 2011), heavy punctuation (Carvalho et al. 2009), hashtags (Wang et al. 2015), quotation marks (Carvalho et al. 2009), positive interjections (Gonzalez-Ibáñez, Muresan, and Wacholder 2011), lexical N-gram cues associated with sarcasm (Davidov, Tsur, and Rappoport 2010), lists of positive and negative words (Gonzalez-Ibáñez, Muresan, and Wacholder 2011), etc. It must be noted that the above features are designed to predict sarcasm in short messages. In this work we demonstrate that these features do not work well for long documents. This means that other features should be devised for detecting sarcasm on a document level.

Recently the necessity of looking beyond the text snippets and into the context that surrounds the possibly sarcastic text utterance got a lot of attention. Researchers investigate the effect of context on sarcasm and design features to capture the global context within which sarcasm appears. Wallace et al. (2015) work on comments from Reddit threads about politics. Wang et al. (2015) work with Twitter messages and analyze these messages as a part of a larger Twitter thread. In both cases, the context is derived using lexical and non-lexical features of the surrounding messages and the information about the overall polarity of the thread (e.g., whether the Reddit thread is a part of the conversation among conservatives or not). The generated context has a certain sentiment that is used for the task of sarcasm detection.

In our work we rely on the importance of context for sarcasm detection. Our approach to contexualization is based on the common belief that a sarcastic document contains a passage which, when taken out of context and analyzed as a stand-alone sentence with the priority of the salient meaning over non-salient one, can be classified as positive but within a given (typically negative) context becomes the holder of sarcasm. For example, the following sentence marked with a positive sentiment label[1] while being a part of an overall negative ($1\star$) review of a Bill Clinton biography documentary signals the presence of sarcasm in the review[2].

> This dvd is great if you think that Gennifer Flowers, Paula Jones and Monica Lewinsky were the highlights of the Clinton administration.

However, sarcasm can be observed in overall positive ($5\star$) reviews as well. For example, in a positive ($5\star$) review about a movie, the following sentence marked as negative is a good signal of sarcasm being present in the review.

> I believe this film was secretly banned from Oscar consideration due to the fact the committee felt it would be unfair to the other nominees.

[1] All sentiment labels presented in this paper are obtained using the Stanford Sentiment Analysis tool (Socher et al. 2013) with the 5-point sentiment scale: very negative (-2), negative (-1), neutral (0), positive (+1), very positive (+2). The Stanford Sentiment Analysis tool sentence sentiment prediction accuracy is 85.4%

[2] All examples presented in this paper are from existing Amazon product reviews. We preserve the original orthography, punctuation, and capitalization

We assume that many of the sarcastic Amazon product reviews that have a $5\star$ rating give this high score to push to the extreme the reviewer's sarcastic attitude to the product. In such cases, a whole review is a positive passage within negative context (sarcastic attitude of a review author to a product). For example, many Amazon reviews for products such as herpes plush dolls, uranium ore, etc. have $5\star$ rating.

To summarize, we propose a novel approach for context modelling via sentence sentiment flow. Our results demonstrate that sentiment flow shifts could be used as reliable classification features for detecting sarcasm, especially for the reviews with high star rating. The contribution of this paper is three-fold:

- we propose a novel approach towards document analysis and context modeling, where we use the sentence sentiment label sequence within a document to capture the document sentiment context and the shifts in sentiment flow – to capture the sentiment context switches;
- we successfully apply the sentiment flow shifts as classification features to build high performance classifiers for the task of identifying documents containing sarcasm, and improve the state-of-the-art F_1-measure by 0.2;
- based on the experimental results and corpus analysis we provide new evidence of the correlation between sarcasm and sentiment and start the discussion about the difference of the sarcasm nature in the documents with overall positive and documents with overall negative sentiment.

Related Work

As mentioned in Introduction there exists a variety of corpora and features used for the task of sarcasm detection.

Corpora

Many of the current research projects, irrespectively of the initial document length, classify only short text snippets as sarcastic or non-sarcastic. Twitter messages that are used in many experiments (Davidov, Tsur, and Rappoport 2010; Gonzalez-Ibáñez, Muresan, and Wacholder 2011; Riloff et al. 2013; Bouazizi and Ohtsuki 2015; Rajadesingan, Zafarani, and Liu 2015) are 140 character long. For the fiction literature experiment (Kreuz and Caucci 2007) only the passages with the direct speech described by the author with the "said sarcastically" phrase are used. While dealing with Amazon reviews, Davidov et al. (2010) analyze parts of the reviews rather than complete reviews. For the Internet dialog (Lukin and Walker 2013; Wallace, Choe, and Charniak 2015), one of the passages in a thread/dialog is classified as being sarcastic or not.

Buschmeier et al. (2014) classify complete Amazon reviews. They do not propose features for macro-level sarcasm detection. Rather, they treat an Amazon review as a single text snippet and extrapolate the features discovered for short texts towards classifying complete Amazon product reviews.

In speech, typically, a short phrase is classified as being said sarcastically or not. Rakov and Rosenberg (2013) use a set of acoustic a pragmatic features on the word and sentence level to classify the phrase "Sure, I Did The Right Thing" into sarcastic and regular. In Tepperman et al. (2006), the phrase "yeah right" is classified whether it is used sarcastically or not according to the "prosodic, spectral, and contextual cues." The contextual cues used in this work are: laughter, whether the phrase comes in the beginning or at the end of the speaker's turn, etc.

Features for Sarcasm Detection in Text

The presence in the text snippets of emoticons, interjections (e.g., ah, oh, yeah) and punctuations (e.g., !, ?) is found to be a good predictor of whether a text snippet is sarcastic or not (Kreuz and Caucci 2007; Carvalho et al. 2009; Gonzalez-Ibáñez, Muresan, and Wacholder 2011). Bootstrapping has been used for learning N-gram lexical cues for predicting sarcasm (Davidov, Tsur, and Rappoport 2010; Lukin and Walker 2013). The presence of positive words and phrases in an overall negative context has been used for detecting sarcasm. These words and phrases are obtained from dictionaries (Pennebaker, Francis, and Booth 2001) as in the work by Gonzalez-Ibáñez et al. (2011), mined using bootstrapping (Davidov, Tsur, and Rappoport 2010; Riloff et al. 2013), or generated using a parse–based lexicon generation algorithm (Bharti, Babu, and Jena 2015).

Context

Given that sarcasm is often associated with the use of positive sentiment within negative context, the idea of automatically capturing the context switch between negative and positive sentiments is a powerful one (Riloff et al. 2013). However, the notion of context within a short message (Twitter) has a peculiar nature as there is not much of text to rely on. Joshi et al. (2015) who work with Twitter messages model the context using a set of sentiment-bearing verb and noun phrases (implicit incongruity), and word-based polarity shifts (explicit incongruity).

The attempts to use behavioral and psychological studies to construct a behavioral modeling framework tuned for detecting sarcasm (Rajadesingan, Zafarani, and Liu 2015) apply the above lexical features but also attempt to capture the the mood of the user by studying the sentiment expressed in her past tweets.

Many recent projects on sarcasm detection, while working on short text snippets like Twitter messages or Reddit comments go beyond one message and propose several ways of taking into account the context surrounding the message that is classified as sarcastic or regular. Wallace et al. (2015) suggest to consider a broad context around a text passage that should be classified as either sarcastic or not. Their approach towards contextualization is based on the assumption that different groups of people have different views on the same issue and thus, "the statement 'I really am proud of Obama' is likely to have been intended ironically if it was posted to a forum frequented by political conservatives."In this work, a global context is defined by the origin of the thread and is used to identify whether a certain passage from this thread is sarcastic or not.

Other context modeling approaches for learning sarcasm on Twitter include extra-linguistic information such as, the

topic of the surrounding messages, historical sentiment profile of the message author, the intended audience, the messages with the same hashtag (Bamman and Smith 2015; Wang et al. 2015).

Context is important not only for sarcasm detection, but for sentiment analysis in general. Miller et al. (2011) trace the flow of sentiment through the blogosphere and demonstrate "that the sentiment of a blog post is affected not only by the sentiment of its immediate parent, but also by its position within a cascade and this cascade's characteristics."

Contextualization for Sarcasm Detection

In our work we analyze Amazon product reviews and identify whether a review is sarcastic or not. For us a review is sarcastic if it contains a sarcastic text snippet or if the entire review is written in a sarcastic manner. One common assumption about sarcasm is that it is frequently manifested via a positive sentiment within a negative context. We demonstrate that such situations can be automatically discovered through the sentiment flow shifts.

To capture the sentiment flow within an Amazon review we tag all the review sentences with sentiment labels. To identify the cases of shifts in sentiment flow we analyze the sentences with positive sentiment labels surrounded by the sentences with negative sentiment labels and the sentences with negative sentiment labels surrounded by the sentences with positive sentiment labels. In our work the sentiment labels are generated **automatically** (Socher et al. 2013). Table 1 contains snippets from several sarcastic Amazon reviews from the corpus collected by Filatova (2012) together with a sentiment label for every sentence.

Examples 1 – 3 in Table 1 are extracted from the reviews with low star rating (negative reviews). In these examples sentences with positive sentiment labels are surrounded by the sentences with negative sentiment labels. We believe, the sentences with positive sentiment labels in these examples are good predictors of the presence of sarcasm in the review and, possibly, are bearers of sarcasm. The positive sentences in Example 1 demonstrate that the review author is not a fan of Emily Brontë's "Wuthering Heights." The positive sentence in Example 2 discusses the *real contribution* of the reviewed movie. The sentence with positive sentiment in Example 3 infers that the movie is boring.

Examples 4 – 5 in Table 1 are extracted from the reviews with high star rating (positive reviews). Example 4 demonstrates the situations where the sentiment flow shift is manifested via the presence of a sentence with negative sentiment label surrounded by the sentences with positive sentiment labels. Example 5 demonstrates the situations where a sentence with positive sentiment label is surrounded by the sentences with negative sentiment labels. In contrast to Examples 1 – 3, the sentiment flow shifts in Examples 4 – 5 do not reveal the bearer of the sarcasm. Rather, one can say that the whole review rather than a part of it is sarcastic.

Thus, the examples in Table 1 demonstrate: sarcasm in negative reviews is of different nature than the sarcasm in positive reviews. Despite this difference in the nature of sarcasm, we believe that in both cases sarcasm is manifested via using a positive sentiment in a negative context. For negative reviews, the global (negative) context can be explicitly deduced from the review text. For positive reviews, however, the global (negative) context is not expressed directly through textual features but rather can be deduced using extra-textual features. However, the presence of sentiment flow shifts in positive reviews still can be a signal of the review being sarcastic.

Sarcasm Classification Experiment

Experiment Corpus

For our classification experiment we start with the corpus of sarcastic and regular Amazon product reviews collected by Filatova (2012). This corpus has been used for the tasks of sarcasm detection (Buschmeier, Cimiano, and Klinger 2014) and sentiment analysis (Otmakhova and Shin 2015). In both cases it is noted that the corpus is imbalanced. According to Table 2, the corpus imbalance is two-directional:

- The distribution of reviews according to the start rating;
- Within every star rating, the number of regular reviews and the number of sarcastic reviews differ.

The corpus imbalance is indirectly exploited by Buschmeier et al. (2014). Their classifier achieves an F_1-measure of 0.74 "by using the star-rating feature together with bag-of-words and specific features with a logistic regression approach." Interestingly, while using only the review star rating as the classification feature, all the five classification methods used in the experiment achieve F_1-measure equal to 0.717. Thus, the star rating feature alone is a good predictor of whether a review is sarcastic or not.

According to Table 2, the original corpus has:

- 289 negative sarcastic reviews (262 1$\star$ and 27 2$\star$ reviews)
- 81 negative regular reviews (64 1$\star$ and 17 2$\star$ reviews)
- 128 positive sarcastic reviews (14 4$\star$ and 114 5$\star$ reviews)
- 701 positive regular reviews (96 4$\star$ and 605 5$\star$ reviews)

Analyzing only positive reviews and classifying all of them as regular provides the accuracy of: 701/(701+128) = 0.846. Analyzing only negative reviews and classifying all of them as sarcastic provides the accuracy of: 289/(289+81) = 0.781.

Thus, we believe that the classification result (F_1 = 0.74) does not capture the performance of the sarcasm classifier; rather, it captures the nature of the corpus where most regular reviews are positive and most sarcastic reviews are negative. Needless to say, this is a peculiarity of the original dataset, and we cannot generalize a sarcasm classifier by exploiting feature correlations peculiar to the dataset.

To avoid artificially boosting our performance by leveraging (implicitly or explicitly) the review $\star$ ratings, we use a balanced version of the original corpus. It is suggested (Batista, Prati, and Monard 2004) that oversampling for the underrepresented class can be used to combat the imbalance issue. Thus, we randomly oversample the minority class for the 1$\star$, 2$\star$, 4$\star$ and 5$\star$ reviews. We omit from further consideration the 3$\star$ reviews and divide the corpus into positive (1$\star$ and 2$\star$) and negative (4$\star$ and 5$\star$) reviews. Thus,

Book: Wuthering Heights, by Emily Brontë	
Example 1: Complete 1⋆ review	
Sent. Tag	*Sentence*
-1:	Well as if the Nineteenth Century weren't bad enough we now have Penguin shoving it down our maws every other day with another re-issue of some tepid "classic."
+1:	Miss Bronte has done it again and wielded her magic pen as a wand and cast her net of sleep on the unsuspecting reading public of America.
-1:	The only consolation the non-preteen girl reader can get form this sack of slumber is the final realisation that "wuthering" is British slang for "your eyelids are getting heavy, why don't you just nod off?"
-1:	I really have to say to Miss Bronte that I did not find Garfield's antics convincing in the least.
Book: The Bush Tragedy, by Jacob Weisberg	
Example 2: Snippet from a 1⋆ review	
Sent. Tag	*Sentence*
-1:	That is the tragedy – that a great nation could be hoodwinked into voting for someone not even qualified for local dog catcher.
-1:	As such the story is a terrible farce.
+1:	The only real contribution is a good assessment of the much over exaggerated role of Dick Cheney in the formulation of Bush policy.
-1:	Cheney remains an evil presence but hardly the power behind the throne.
Movie DVD: The Twilight Saga: New Moon, by Chris Weitz (Director)	
Example 3: Snippet from a 1⋆ review	
Sent. Tag	*Sentence*
-1:	By the end of the movie, it seemed like there was a lot of suspense, but not the right kind of suspense.
+1:	I found myself with enough time between the character's lines to figure out what I'm doing for the rest of the week/month.
-1:	A lot of times the pauses in the dialog were filled by grunts, sighs, and heavy breathing, which is more than a little uncomfortable.
Movie DVD: Alone in the Dark, by Uwe Boll (Director)	
Example 4: Snippet from a 5⋆ review	
Sent. Tag	*Sentence*
+2:	This is another of visionary director Uwe Bolls brilliantly forged masterpieces.
-1:	I believe this film was secretly banned from Oscar consideration due to the fact the committee felt it would be unfair to the other nominees.
+2:	This sweeping epic tells the story of several nameless teens who travel to Isle del Muerte, Canada, to party and are swept into a tale of zombies, deceit, suspense, and moral choices that push the boundaries of the human psyche.
+1:	Uwe Boll deserves an award for his contributions; he is truly a pioneer of German tax loopholes, and will continue to inspire future generations with his brilliance.
+1:	I salute thee, Mr. Boll.
Book: How to Avoid Huge Ships, by John W. Trimmer	
Example 5: Complete 5⋆ review	
Sent. Tag	*Sentence*
-1:	This is a must-read for anyone who encounter huge ships daily and do not want to get run over by them.
+1:	I found this book extremely helpful.
-1:	To this day, I have never been run over by a single huge ship!!

Table 1: Examples of sentiment flow shift.

		Number of reviews with				
		1⋆	2⋆	3⋆	4⋆	5⋆
sarcastic	437	262	27	20	14	114
regular	817	64	17	39	96	605

Table 2: Distribution of reviews by star rating.

we end up with the corpus of 1,980 reviews: 578 negative reviews (289 sarcastic and 289 regular) and 1402 positive reviews (701 sarcastic and 701 regular). By applying oversampling we address the imbalance between regular and sarcastic reviews. The corpus still lacks the positive / negative reviews balance. For this work this balance is not crucial.

Baseline

The only work that uses the Amazon product reviews corpus collected by Filatova (2012) is the work by Buschmeier et al. (2014), where the features that are devised for sarcasm detection on short messages (mainly Twitter) are extrapolated to be used for documents of various lengths. Given the discussion above, the results of Buschmeier et al. (2014) are equivalent to a random classifier baseline.

For our classification experiment, we equalize the number of sarcastic and regular reviews. Given the updated oversampled balanced corpus and the fact that the prior work does not generate better-than-random results, rather, it captures the high correlation between sarcasm and sentiment, we believe that the accuracy of 0.5 is a fair baseline.

Very Negative – Positive	Very Positive – Negative
Very Negative – Very Positive	Very Positive – Very Negative
Negative – Positive	Positive – Negative
Negative – Very Positive	Positive – Very Negative

Table 3: Eight Classification Features.

Corpus Pre-Processing

We use the Stanford Sentiment Analysis tool (Socher et al. 2013) with the 5-point sentiment scale (very negative (-2), negative (-1), neutral (0), positive (+1), very positive (+2)) to assign sentiment labels to all the sentences in all the Amazon product reviews in the oversampled balanced corpus.

Classification Features

We hypothesize that the sentiment flow can be used for sentiment contextualization and the sentiment flow shifts – to predict the presence of sarcasm in text. To test our hypothesis we create a set of bi-gram features that capture the positive to negative and negative to positive sentiment shifts.

Every sentence in our corpus is marked with a sentiment label on a 5-point scale. All in all, there are 25 different sentiment shift bi-grams that can be generated using 5 sentiment labels. Out of these 25 bi-grams we keep only those that capture the positive to negative or negative to positive shifts. Thus, we end up with eight bi-gram features for our experiment (Table 3). The value of every feature is computed as the ratio of how many times we encounter this shift within the Amazon product review under analysis over the combined number of all eight shifts for this Amazon review.

It must be noted that sentiment flow can be modelled not only via sentiment shifts but also using Isotonic Conditional Random Fields (Mao and Lebanon 2006). This approach towards sentiment flow has been successfully used for polarity classification (Wachsmuth, Kiesel, and Stein 2015).

Experiment Settings

Given the nature of sentiment flow shift features, we propose to use a sequence classification method.[3] We use a feature based sequence classification, where a sequence of labels is transformed into a feature vector and then conventional classification methods are applied to the obtained feature set.

Here are the settings for our classification experiment:

- tool: the open-source machine learning library *scikit-learn* (Pedregosa et al. 2011) for Python;
- classifiers: 7 different classification techniques (Table 4);
- features: 8 sentiment shift bi-grams (Table 3);
- training/testing: 10-fold cross validation.

Though our experiment does not have a star rating and consequently the overall sentiment of the document as one of its classification features, we believe that the correlation between sarcasm and sentiment is very high. Thus, we apply these classification settings to three corpora, each of which has an equal number of sarcastic and regular reviews:

[3]For more on sequence classification see Xing et al. (2010).

	Classifier	F_1	Rec.	Prec.
All Reviews	k-NN	0.719	0.767	0.683
	SVM (linear)	**0.789**	**1.000**	**0.652**
	SVM (RBF)	0.788	0.999	0.651
	Decision Tree	0.698	0.705	0.721
	Random Forest	**0.789**	**1.000**	**0.651**
	AdaBoost	**0.724**	**0.717**	**0.734**
	Logistic Regr.	0.788	0.996	0.651
Positive Reviews	k-NN	0.846	0.847	0.848
	SVM (linear)	**0.916**	**1.000**	**0.846**
	SVM (RBF)	**0.916**	**1.000**	**0.846**
	Decision Tree	0.905	0.966	0.849
	Random Forest	**0.916**	**1.000**	**0.846**
	AdaBoost	0.904	0.971	0.845
	Logistic Regr.	**0.916**	**1.000**	**0.846**
Negative Reviews	k-NN	0.469	0.445	0.552
	SVM (linear)	0.670	1.000	0.581
	SVM (RBF)	0.522	0.512	0.581
	Decision Tree	0.606	0.691	0.534
	Random Forest	0.618	0.660	0.596
	AdaBoost	0.607	0.651	0.571
	Logistic Regr.	0.489	0.531	0.479

Table 4: Experimental Results for the Balanced Dataset. The baseline classifier has recall/precision equal to 0.5

1. balanced corpus of 1,980 Amazon product reviews (contains $1\star$, $2\star$, $4\star$, $5\star$ reviews);
2. balanced corpus of positive 1402 Amazon product reviews (contains $4\star$, $5\star$ reviews);
3. balanced corpus of negative 578 Amazon product reviews (contains $1\star$, $2\star$ reviews).

The results for our classification experiments are summarized in Table 4.

Results Analysis

According to the results presented in Table 4, for the first experiment with the corpus of both positive and negative reviews, the best F_1-measure is equal to 0.789, which is higher than the random baseline. For the best precision value (0.734), the F_1-measure drops to 0.724.

However, most interesting result of this experiment comes from the separation of the corpus into positive and negative corpora. The classification of the negative reviews into sarcastic and regular is done almost at random. While the classification of the positive reviews given the sentiment shift features is very reliable: we achieve perfect recall, and very high precision (0.846) and F_1-measure (0.916) for four out of seven classifiers used in the experiment.

These experimental results support our hypothesis that the nature of sarcasm in negative Amazon reviews is different from the nature of sarcasm in positive Amazon reviews. Let us examine again the examples presented in Table 1. In Examples 1 – 3, one can notice that the sentence with a positive sentiment label surrounded by the sentences with negative sentiment labels is the bearer of sarcasm in the document. While in Examples 4 – 5, it is difficult to pinpoint

the bearer of sarcasm, rather, one can notice that the whole review rather than a part of it is sarcastic.

According to our observation, negative reviews are overall genuine and only some of the review sentences (marked by the shift in the sentiment flow) can be identified as sarcastic; while positive reviews are not genuine (i.e., they require non-salience-based interpretation) with the complete review text being sarcastic. We believe that in both cases, the statement that sarcasm is manifested via positive statement within negative context holds. For negative reviews, negative context is manifested in text as such reviews consist mainly of truly negative information about the product. While, in positive reviews, negative context comes from the attitude of the review author to the product, it is not manifested by salient interpretation of text, rather it can be described only using extra-textual features. Looking into the positive sarcastic examples in the corpus used in our experiments, we observe that many of these reviews describe such products as: plush herpes dolls, uranium ore, etc.

Conclusion

We introduce a novel approach for sentiment context modeling for the task of sarcasm detection. We use the sequence of sentiment labels assigned to the document sentences for context modeling. To capture the presence of sarcasm in a document we use sentiment flow shifts as the predictors of context switch from negative to positive and from positive to negative. We demonstrate that these sentiment flow shifts can be used as a reliable classification features for the task of sarcasm detection. Given the substantial difference in the sarcasm classification performance on positive ($4\star$ and $5\star$) and negative ($1\star$ and $2\star$) reviews, we assume that the nature of sarcasm in positive reviews is different than the one in negative reviews.

References

Bamman, D., and Smith, N. A. 2015. Contextualized sarcasm detection on twitter. In *Proceedings of ICWSM*.

Batista, G. E.; Prati, R. C.; and Monard, M. C. 2004. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.* 6(1):20–29.

Bharti, S. K.; Babu, K. S.; and Jena, S. K. 2015. Parsing-based sarcasm sentiment recognition in Twitter data. In *Proceedings of ASONAM*.

Bouazizi, M., and Ohtsuki, T. 2015. Sarcasm detection in Twitter: "All your products are incredibly amazing!!!" – Are they really? In *Proceedings of GLOBECOM*.

Buschmeier, K.; Cimiano, P.; and Klinger, R. 2014. An impact analysis of features in a classification approach to irony detection in product reviews. In *Proceedings of WASSA*.

Carvalho, P.; Sarmento, L.; Silva, M. J.; and de Oliveira, E. 2009. Clues for detecting irony in user-generated contents: Oh...!! it's "so easy" ;-). In *Proceedings of TSA*, 53–56.

Davidov, D.; Tsur, O.; and Rappoport, A. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of CoNLL*.

Filatova, E. 2012. Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In *Proceedings of LREC*.

Gonzalez-Ibáñez, R.; Muresan, S.; and Wacholder, N. 2011. Identifying sarcasm in twitter: A closer look. In *Proceedings of ACL/HLT*, 581–586.

Joshi, A.; Sharma, V.; and Bhattacharyya, P. 2015. Harnessing context incongruity for sarcasm detection. In *Proceedings of ACL*, 757–762.

Kreuz, R., and Caucci, G. 2007. Lexical influences on the perception of sarcasm. In *Proceedings of the NAACL Workshop on Computational Approaches to Figurative Language*.

Lukin, S., and Walker, M. 2013. Really? well. apparently bootstrapping improves the performance of sarcasm and nastiness classifiers for online dialogue. In *LASM*.

Mao, Y., and Lebanon, G. 2006. Isotonic conditional random fields and local sentiment flow. In *Proceedings of NIPS*.

Miller, M.; Sathi, C.; Wiesenthal, D.; Leskovec, J.; and Potts, C. 2011. Sentiment flow through hyperlink networks. In *Proceedings of ICWSM*.

Otmakhova, Y., and Shin, H. 2015. Do we really need lexical information? towards a top-down approach to sentiment analysis of product reviews. In *Proceedings of NAACL/HLT*.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. *Linguistic Inquiry and Word Count: LIWC*. Mahwah, NJ: Erlbaum Publishers.

Rajadesingan, A.; Zafarani, R.; and Liu, H. 2015. Sarcasm detection on twitter: A behavioral modeling approach. In *Proceedings of WSDM*, 97–106.

Rakov, R., and Rosenberg, A. 2013. "Sure, I Did The Right Thing": A System for Sarcasm Detection in Speech. In *Proceedings of INTERSPEECH*, 842–846.

Riloff, E.; Qadir, A.; Surve, P.; Silva, L. D.; Gilbert, N.; and Huang, R. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of EMNLP*.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D.; Ng, A. Y.; and Potts, C. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of EMNLP*, 1631–1642. Stroudsburg, PA: Association for Computational Linguistics.

Tepperman, J.; Traum, D.; and Narayanan, S. S. 2006. "yeah right": Sarcasm recognition for spoken dialogue systems. In *Proceedings of INTERSPEECH*, 1838–1841.

Wachsmuth, H.; Kiesel, J.; and Stein, B. 2015. Sentiment flow – a general model of web review argumentation. In *Proceedings of EMNLP*, 601–611. Lisbon, Portugal: Association for Computational Linguistics.

Wallace, B. C.; Choe, D. K.; and Charniak, E. 2015. Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment. In *ACL*.

Wang, Z.; Wu, Z.; Wang, R.; and Ren, Y. 2015. Twitter sarcasm detection exploiting a context-based model. In *Proceedings of WISE*.

Xing, Z.; Pei, J.; and Keogh, E. 2010. A brief survey on sequence classification. *SIGKDD Expl. Newsl.* 12(1):40–48.