Automatic Detection of Double Meaning in Texts from the Social Networks

Todor V. Tsonkov
Faculty of Mathematics and Informatics
Sofia University
todort@fmi.uni-sofia.bg

Ivan Koychev
Faculty of Mathematics and Informatics
Sofia University
koychev@fmi.uni-sofia.bg

Abstract

The paper presents a method for automatic detection of double meaning of texts in English from the social networks. For the purposes of this paper we define double meaning as one of irony, sarcasm and satire. We proposed nine rules selected from a pool of twenty. We defined six features and evaluated their predictive accuracy. Further we compared the accuracy of three different classifiers - Naive Bayes, k-Nearest Neighbours and Support Vector Machine. We also studied the predictive accuracy of all words and bi-terms. We test the algorithms above against opinions from the social network: sample opinions from the social networks Facebook, Twitter and Google+. These opinions were extracted via HTTP requests using one of the hashtags #sarcasm, #irony or #satire and we select 3000 opinions for each of the tests.

1 Introduction

The automatic detection of double meaning presents a big challenge to the field of opinion mining as standard algorithms don't produce expected results. In this study we present an approach for automatic detection of double meaning that achieves improvement in existing results on texts from social networks.

Mining the opinions from the social networks become very widespread in sentiment analysis. Most of those networks provide public APIs, which allow streams of posts to be captured and continuously analyzed for public opinions on particular topic [Bif10].

In this paper we won't make any difference between irony, sarcasm and satire. The reason for this is that usually the users in the social networks don't make a

clear difference between the three terms. We have conducted an experiment to test this hypothesis. The results provide strong evidence that there's no difference in the texts meaning when it was tagged with one of those three double meaning terms.

In the current paper we propose a new method for determining double meaning.

The contributions of the current work are:

1. Creation of heuristics rules for double meaning detection and providing templates for easy modification, testing and changing of the rules.
2. Collection of corpus with double meaning opinions.
3. Suggestion of set of derived features and testing their accuracy on the double meaning corpus. Using features selection algorithm.
4. Using classifiers and machine learning approaches for automatic determining of double meaning in texts.
5. Providing experimental evidences that users in the social networks don't make a difference between irony, sarcasm and satire.
6. Testing accuracy of words and bi-terms against double meaning.

2 Related work

Although double meaning detection has had significant research in the field of psychology the task of automatically detecting of it on social networks has received considerable attention in recent years. The task is very similar to traditional NLP sentiment analysis. [Has12] use a supervised Markov model, part of speech, and dependency patterns to identify polarities in threads posted to Usenet discussion posts. [Has12], [Bar14] investigate definitions of irony, sarcasm. Verbal irony has been defined in several ways over the years but there is no consensual agreement on the definition. The standard definition is considered "saying the opposite of what you mean" [Qin53] where the opposition of literal and intended meanings is very clear. Grice believes that irony is a rhetorical figure that violates the maxim of quality: "Do not say what you

believe to be false". Irony is also defined [Gio12] as any form of negation with no negation markers (as most of the ironic utterances are affirmative, and ironic speakers use indirect negation).

There are also a few computational models that detect sarcasm on Twitter and Amazon [Dav10], [Gon11], [Lie13], but even if one may argue that sarcasm and irony are the same linguistic phenomena, the latter is more similar to mocking.

[Gon11] use an approach of looking at lexical features for identification of sarcasm. In addition, they also look at pragmatic features, such as establishing common ground between speaker and hearer and emoticons.

Others have designated sentiment scores for news, stories and blog posts based on algorithmically generated lexicons of positive and negative words. [God07] demonstrate experimentally that despite frequent occurrences of irregular speech patterns in tweets, twitter can provide a useful corpus for sentiment analysis. The diversity of Twitter users makes this corpus especially valuable for instance to track misleading political memes. Along with many advantages using Twitter as a corpus for sentiment analysis it has a unique challenge as the posts are less than 140 symbols. This means they can often contain unusual grammar and unconventional words and symbols.

3   Using Heuristic Rules

We use nine heuristic rules to detect whether the opinion is really satirical, ironical or sarcastic. We check if specific assumptions about an opinion can be successful in determining whether it really fits one of irony, satire or sarcasm. Table 1 shows a detailed explanation of these rules. The usage of double meaning (containing one of the hashtags #irony, #sarcasm and #satire) is rare in other opinions and when we don't have any specific knowledge about the topic that is presented the task becomes impossible to resolve. The nine rules are selected from a pool of 20 rules based on best accuracy against a test set of 200 random opinions. We are using four main features to describe a rule:

1. A description of the rule in natural language – it is used in order to be clear what we have expressed by the template.

2. A template that describes it. Each rule can be represented as a regular expression and thus easily more rules can be created and added and described in a formal language. One of the innovations of this paper is easily automating rules addition.
3. Checking if a rule is language dependent or not. This is important as some of the rules can be easily changed for a different language.
4. Determining which double meaning needs to be tested against – sarcasm, satire or irony. We will test the accuracy of each rule against each double meaning in the experiments part.

Each rule can be added and modified easily and thus more rules can be added or existing rules modified and we provide a template for the purposes of formal description of new rules.

A formal description of the rules can be found in the tables below. We divide them on language dependent and language independent.

Table 1. Language Dependent Rules for double meaning detection.

Description	Template
1. Contains at least 3 positive or 3 negative words (only positive and negative words not combined).	[+word] AND [+word] AND [+word] OR [-word] AND[-word] AND [-word]
2. Contains a positive word and a negative word	+word and –word
3. Detection of opposite meaning words from a containing list	[word1] and [word2]
4. Contains a positive word and a negative emotion.	[+word] and [-emoticon]
5 Gap between common and rare words from a list	Common – Rare

Table 2. Language Independent Rules for double meaning detection.

Description	Template
6. Contains at least two of exclamation/question mark or	([!] or[?] or […]) AND ([!] or [?] or […])
7. Using at least three consecutive adjectives	+word and –word
8. Using a word in capital letters	[WORD]
9. Contains "" or ''	['] OR [""]

4 Design of ad hoc Features

We employ five extra features aiming to improve the accuracy of the rule based approach.

We use some of the ideas from [Bar12] and our research on the topic and have modified the features we take from their research so that they match the purposes of our research.

A broader description and an explanation why these features have been selected follows:

1. The difference between words containing positive sentiment and negative sentiment is indicative for double meaning as shown in [Vea10] and by research made by the authors of this paper. We investigate the number of positive and negative words as well. They have been selected from [Com15] lists and are language specific.

2. The number of punctuation signs, emoticons and links is important because as shown in [Van14] the average number of emoticons in ironic tweets is higher than in not ironic. Besides, more than one punctuation sign may mean double meaning which is the purpose of this paper. Opinions with links usually express the opinion of the author about the link which often can be ironical, satiric or sarcastic. Out of 100 randomly selected opinions for the social networks that are ironical, satirical or sarcastic, 65 contain links. This may be due to the nature of the social networks opinions where many links are shared.

3. The number of adjectives usually shows a specific relation towards the topic and more often than not expresses an opinion rather than just stating facts.

The length of the opinion is important because as shown in [Dav10] the longer the sentence the more probably it is for it to be ironical or sarcastic. We use the sum of the length + number of adjectives multiplied by 10.

4. We define intensity score as described in [Dip12]. We measure the intensity of the adverbs and adjectives and we calculate the intensity score as the sum of the intensity of both adverbs and adjectives.

5. The gap between common and rare words measures how unique the opinion is in comparison to others. We define common words as one of the most common 2000 words as defined from here: http://www.talkenglish.com/Vocabulary/Top-2000-Vocabulary.aspx. The other words are considered rare. The gap is calculated as the difference between the common words and the rare words divided over all the words in the sentence.

Table 3. Features Description

Feature Number	Description
1.	Number of words containing positive sentiment and negative sentiment
2.	Number of emoticons, punctuation signs and links
3.	Number of adjectives and length of the opinion
4.	Intensity score
5.	Frequency (gap between common and rare words).

The performance of the features can be found in the experiments part below.

5 Experiments

We test against 3000 opinions from Twitter and Facebook containing one of the hashtags #sarcasm, #satire or #irony. 80% of the opinions are from Twitter; and 20% from Facebook. All of the opinions are in English. If one opinion matches more than one rule it will be counted in any of the rules it matches. We filter out all sentences that are not in English. We also

consider every opinion that matches these tags is really double meaning. We test each opinion against each of the double meanings – irony, sarcasm and satire and each has 1000 opinions. For the experiment described in this paper we use Facebook and Twitter social networks and we implement a tool written in the Java language for extracting the posts (via HTTP requests and responses). Each post matches only one of the hashtags "#sarcasm", #irony or #satire and is manually reviewed form the authors of the paper.

We don't include opinions that match more than one of the hashtags in order to make sure the users specifies only one of the double meanings.

Example posts extracted from the research include:

1. Really delighted that I took the fantasy league captaincy off Diego Costa today #sarcasm
2. I know you "protestors" won't believe this, but the cops apparently just shot a white guy. #sarcasm
3. Because it is the gun that kills, not the humans holding it. #sarcasm
4. I'm SO glad racism is dead in this country. Whew! *wiping brow*#SARCASM
5. BLOODY IMMIGRANTS DRAINING THE STATE....... #irony
6. Black Donor's Sperm Mistakenly Sent To Neo-Nazi Couple"
Well. That's some serious #Irony. #KarmaIsWonderful
7. This just behooves me to share. The irony that the conference is happening to fight islamophobia while islamophobes rally outside. Note the facility is used for "several cultural and religious-based groups and events" according to the Garland ISD spokesman. #misinformation #iamamuslim #irony #equality?
8. The Borowitz Report: Queen Elizabeth II took to the airwaves to inform the people of Scotland that she "graciously and wholeheartedly" accepted their apology. #satire

The experiment design can be described in the following formal way:
For each post from the social networks do the following:
1. Create a thread that processes each rule as described:
2. For each rule described as a regular expression by the pattern above
3. Check if the post matches the rule.

4. If the post matches the rule increase the counter of the current rule
5. Add the opinion to the list of opinions of the current rule
For each list of opinions do:
If the rule is correctly classified increase the counter of correctly classified rules.
For each rule calculate the success rate as:
Correctly classified rules that are matching this rule divided all rules classified as matching this rule.

Table 4. Rules Doube Meaning Accuracy

Rule #	Irony	Sarcasm	Satire
1	78%	73%	76%
2	71%	69%	75%
3	80%	77%	81%
4	64%	66%	65%
5	84%	80%	83%
6	63%	67%	65%
7	58%	59%	62%
8	59%	55%	55%
9	61%	60%	58%
1-9	77.25%	75.75%	77.5%

The results above show why the task of the automatic detection of double meaning in texts is so tough. Without knowing very specific details it is usually difficult to create exact rules when there's no specific information about the topic and the authors of the opinions. From the results in the last row of table 4, we can conclude that there's no difference between irony, satire and sarcasm in texts from the social networks as the differences in accuracy are small.

In Table 6 we show the accuracy in terms of correctly predicted double meaning in each feature described above.

Table 6. Features Accuracy Results

Feature	Accuracy

1.	74%
2.	70%
3.	63%
4.	59%
5.	67%

The accuracy in determining double meaning of each of the terms above show that using terms only it's tough to achieve high accuracy.

We compare with the rule based approach and the feature methodology algorithm with other approaches. We train and test three different classifiers – Naïve Bayes, k-Nearest Neighbours and Support Vector Machine.

We select 4000 opinions from Facebook and Twitter social networks and test what is the precision of each classifier as described in part 4. Out of these opinions 2000 contain the tags #irony, #satire or #irony and the other 2000 are randomly selected from the social networks.

In table 7 we investigate the accuracy of features selection. We investigate all combinations of two features as if we use three of more features we have too few opinions that match them to have a meaningful statistics

Table 7. Features Selection

Feature	Accuracy
1-2.	82%
1-4	77%
2-4	73%
1-5	77%
1-3	76%

The precision is in terms of opinions that can be classified as ironical, sarcastic or satirical.

We use five features as described in section 5 Feature Methodology as a dimension and thus create a five dimensional space. The distance used is normalized Manhattan distance. We have decided to use the Manhattan distance as it's provided in WEKA and we use it as defined in [Gio95]

We use the classifiers implemented in the open source product WEKA to test the classifiers. The results are provided in Table 8 below.

Table 8. Classifiers results

Classifier	Accuracy
Naive Bayes	79%
K-NN	78%
Support Vector Machine	81%

Besides these features we test against all the words and bi-words to determine what the relevance of each is. We have selected the words and bi-words and check what their accuracy is against double meaning and included all the words and bi-words that have been used in the sample opinions. For the purposes of this research we won't make a difference between irony, sarcasm and satire. We have tested against double meaning and the results can be seen in Table 8 below and are measuring the precision against two classes of opinions – double meaning and direct meaning (not double meaning). We have tested against all the words that can be found at least five times in the opinions. This is done in order to have a sufficient number of occurrences of each term in order to have meaningful statistics.

Table 9 Words or Bi-word double meaning accuracy

Word/Bi-word	Accuracy
She/He/It	52%
He/she/it has	57%
And	40%
Finest	66%
Have	40%
His	45%
Wish	60%
Smart	77%
Congratulations	80%
He is	42%
His	64%

The results of in Table 9 show that using terms only is not a good way to achieve high accuracy in

predicting double meaning. However terms like "congratulations" and "smart" are very good predictor for double meaning. Those results can further be used for weighting features.

6   Conclusion

In this article we have described and proposed three different ways for determining automatically double meaning in English texts. Firs we proposed 9 heuristic rules for detecting double meaning. The rules have been tested by the authors of this paper and have shown unique details about double meaning based on manually reviewing sample opinions from the social networks and have been selected from a pool of twenty based on the best accuracy against the opinions.

Adding features and the three classifiers described above shows an improvement in the accuracy in comparison with the rule based selection. The classifiers use features described in the feature methodology parts. The features have been invented and tested by the authors of this paper based on previous research made on this topic and our investigation of what needs to be added for the specifics of the topic.

In order to verify the results are correct one should filter out opinions written in other languages and include only opinions in English. Future work can include improvement of the rules in order to include more cases in English that contain statements with double meaning. The ideas in this article can be developed in other languages as well. We can add more rules and classifiers. Another idea for improvement is to test against other languages as there might be some language specific rules and features. Also more social networks and more judges could be added in order to determine whether the opinion really contains a double meaning or not.

7   References

[Tso13] Tsonkov, T., and Koychev I.- Detecting Irony in texts from the social networks: the Bulgarian language case (2013)

[Pak13] Pak and Paroubek–"Twitter as a Corpus for Sentiment Analysis and Opinion Mining", 2010

[God07] Godbole et al – "Large-Scale Sentiment Analysis for News and Blogs" 2007

[Has12] Hassan et al – "Semantic Sentiment Analysis of Twitter" 2012

[Pot11] Potts, C. 2011. Developing adjective scales from usersupplied textual metadata. NSF Workshop on Restructuring Adjectives in WordNet. Arlington,VA

[Bar14] Francesco Barbieri and Horacio Saggion, 2014 1- Automatic Detection of Irony and Humour in Twitter

[Vea10] Veale, T., and Hao, Y. 2010b. An ironic fist in a velvet glove: Creative mis-representation in the construction of ironic similes. Minds and Machines 20(4):635–650

[Bif05] Albert Bifet and Eibe Frank 2005- Sentiment Knowledge Discovery in Twitter Streaming Data

[Dav10] Davidov, D.; Tsur, O.; and Rappoport, A. 2010. Semisupervised recognition of sarcastic sentences in twitter and amazon. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, 107–116. Association for Computational Linguistics.

[Dip12] Dipankar D. and Sivaji B. 2012, Identifying Emotional Expressions, Intensities and Sentence level Emotion Tags using a Supervised Framework*

[Com15] Common Opposites - Antonyms Vocabulary Word List http://www.enchantedlearning.com/wordlist/opposites.shtml

[Top20] Top 2000 Vocabulary Words http://www.talkenglish.com/Vocabulary/Top-2000-Vocabulary.aspx

[Kis15] https://blog.kissmetrics.com/facebook-statistics/

[Van14] Vanin, A.; de Freitas L.; Viera R.; Bochernistan M. 2014 Some clues on Irony Detection of Tweets

[Gio95] Rachel Giora. 1995. On irony and negation. Discourse processes, 19(2):239–264.

[Qun53] Quintilien and Harold Edgeworth Butler. 1953. The Institutio Oratoria of Quintilian. With an English Translation by HE Butler. W. Heinemann.

[Gon11] Roberto Gonzalez-Ibanez, Smaranda Muresan, and ˜ Nina Wacholder. 2011. Identifying sarcasm in twitter: A closer look.

In ACL (Short Papers), pages 581–586. Citeseer.

[Lie13] Christine Liebrecht, Florian Kunneman, and Antal van den Bosch. 2013. The perfect solution for detecting sarcasm in tweets# not. WASSA 2013, page 29.