An Effective Approach for Sarcasm Detection in Text Data for Sentimental Analysis

Adarsh M J[1], Dr. Pushpa Ravikumar[2]

Research Scholar, Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, Karnataka, India
Professor and Head, Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, Karnataka, India
*Corresponding author E-mail: adarshmj@aitckm.in

Abstract

The stream of Sentiment Analysis has become very popular today helping people and corporate to analyze the orientation of sentiments towards particular product and people. The Sentiment Analysis will not be complete without analyzing the Sarcasm or Irony in statements. Sarcasm is the art of saying something opposite to the original meaning in a sentence. Most of the times Sarcasm in sentences makes it more Negative than positive. In this paper, an approach is adopted to identify the sarcasm in sentences using Sentiwordnet. A set of core popular sarcastic sentences are considered and scores are calculated. The scores points out at the sarcasm in sentences which most of the time is negative. The polarities of sentences are also calculated and also the sentences are checked for sarcasm scores in Sarcasm detector tool.

Keywords: Polarity, Sarcasm, Score, Sentiment

1. Introducion

In the current world of Data Analysis, identifying information in the available data has opened many avenues in Research particularly towards Machine Learning and artificial Intelligence. One such approach of identifying information for useful causes is sentimental Analysis. Detection of sentiments in a given text or a sentence helps in coming out with good decisions which really helps both corporate and common man. As, today is the era of E-Commerce, the analysis of sentiments in data has gained importance.

Sarcasm is an art of saying something opposite to what one originally mean. Sarcasm is identified as ironic or satirical wit that is used to insult, mock or amuse by not directly mentioning the words. Identifying sarcasm is an important aspect of sentimental analysis as many positive sentiments may end up with the negative intent and negative sentiments may come up with positive intent.

In social media like twitter or facebook, when people write sarcastic sentences, they append the sentence with a hashtag(#). It is meant that, a sentence with a sarcasm hashtag is identified as a sarcastic sentence. The greater challenge is to identify the exact sentiment in a sarcastic sentence.

The sentences with sarcasm start with a positive notion and ends with a negative one or vice versa. We consider some popular sarcastic sentences which have both negative and positive intent. We try to bring out the scores for each sentence saying whether it is sarcastic or not. We also check for essence of sarcasm in sentences through sarcasm detector tool.

2. Related Work

Ellen Riloff et.al [1] in their work tried to identify sarcasm in sentences by identifying positive sentiment phrases and negative situation phrases. They used a Bootstrapping algorithm to identify the phrases. The phrases considered were limited to specific syntactic structures and were not with contrasting phrases with constrained context.

Mondher Bouazizi and Tomoaki ohtsuki [2] in their paper proposed a method to detect sarcasm in twitter that makes use of the different components of the tweet. They proposed 4 sets of features like sentiment related, punctuation related, Lexical related and pattern related that cover different types of sarcasm they identified and used it to classify tweets as sarcastic and Non-sarcastic.

Mondher Bouazizi and Tomoaki ohtsuki [3] in their subsequent work proposed a minimal set of features like non textual and textual that classified tweets irrespective of topic. Non Textual features considered entities like hash tag and textual features considered entities like positive and negative words.

Santosh Kumar Bharti et.al [4] in their paper proposed two approaches to detect sarcasm in the text of Twitter data. They proposed Parsing based Lexicon generation algorithm for sarcasm detection and also detection of sarcasm based on occurrence of Interjection word. They used F score to evaluate the results.

Santosh Kumar Bharti et.al [5] in their continued work proposed a Hadoop frame work that captures real time tweets and processes it with a set of algorithms which identifies sarcastic sentiments. They observed that the elapse time for analyzing and processing under Hadoop based framework significantly outperforms conventional methods.

Ashwin Rajadesigan [6] in his work explored the possibility of using behavior traits intrinsic to users of sarcasm to detect

sarcastic tweets. He considered physiological, behavioral sciences and observations on twitter users. He developed computational features to model the manifestations of these forms using user's profile information and tweets.

Aditya Joshi et.al **[7]** made a survey on the works of automatic sarcasm detection considering three approaches. Semi-supervized pattern extraction to identify implicit sentiment, use of hash-tag based supervision and incorporation of context beyond target text.

Mathieu Cliché **[8]** developed a tool based on n-grams and with an observation that sarcastic tweets may be more negative than non-sarcastic tweets. He divided tweets into multiple parts and used two sentiment analyzers based on Sentiwordnet and Text blob. The Sarcasm detector gives scores from negative to positive with sarcastic sentences giving positive scores

3. Methodology

The Sentiwordnet **[8]** dictionary provides scores for positive and Negative words in English Language. Textblob **[9]** package provides sentences with objectivity and subjectivity. Usually Sarcastic sentences consist of Powerful Positive words to describe Negative scenarios. For Example: *Absolutely adore it, when my cab is late.* Where, absolutely, adore are positive words (adjectives) describing the Negative situation.

The sentences with Sarcasm contain Positive words following Negative words and Negative words following Positive words. The Sentiwordnet provides sentiment oriented words with polarities. It is done through Parts of Speech (POS) Tagging. Each word is assigned a value based on its parts of speech.

Once we, have a sentence, its polarity of words are calculated, then the average of those polarities are found and we can call that average as score. If the score is less than 0.1 then we can term that sentence as sarcastic and the sentences with score greater than 0.1 can be termed as non sarcastic sentences as most of the sarcastic sentences have negative orientation with more negative words.

We can summarize it in form of following algorithm.

Input: Sentence
Output: Classification: Sarcastic or Non-Sarcastic sentence
1.   For sentence
2.   Calculate Polarity of words
3.   Score= Average of Polarity
4.   End for
5.   If score < 0.1
6.   Sarcastic sentence
7.   Else if score > 0.1
8.   Non-Sarcastic sentence
9.   Else Neutral sentence
10. End if

The Textblob tool provides us with polarity and subjectivity for sentences. The Polarity values ranges from -1.0 to +1.0 and subjectivity values from 0.0 to 1.0. The polarity directs towards Positiveness and Negativeness of a sentence and Subjectivity guides towards fairness.

While applying Textblob for Polarity, the sentence may be termed as Positive or Negative but will not have a say about sarcasm. Figure 1 shows the process of sarcasm detection with inputting the sentences, Polarity of words calculation with scores and classification of sarcastic and non sarcastic sentences.

The sentences are subjected to POS tagging and the corresponding Parts of speech of the sentiment oriented words are found. The Polarities of the sentiment oriented words are found using Sentiwordnet for each sentence. The average of the scores is calculated for each sentence. The average of scores comes less than 0.1 for most of the sentences.

In this paper we have considered the most popular daily use sarcastic sentences and have found the average Polarity scores using Sentiwordnet. Table 1 shows some of the sarcastic sentences we have considered and the scores obtained for each sentences. The sarcastic sentences will have scores
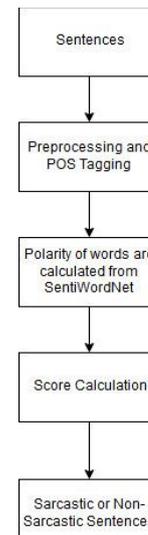

Fig 1: Sarcasm detection on Text Data

Table 1: Sarcastic sentences and Scores

Sl.No	Sentence	score
1	*I hate when I buy a bag of air and there's chips in it*	-0.13333333
2	*my grandad always sounds so ill when i speak to him on the phone*	-0.075
3	*when i speak politely to people they understand wrong so ill be rude*	-0.0987606
4	*Light travels faster than sound. This is why some people appear bright until they speak.*	-0.00379807
5	*It's okay if you don't like me. Not everyone has good taste*	0.04166666
6	*Mirrors can't talk, lucky for you they can't laugh either*	0.066
7	*If had a dollar for every smart thing you say. I'll be poor*	0.0418081
8	*Are you always so stupid or is today a special ocassion?*	0.0425101
9	*I'm sorry I hurt your feelings when I called you stupid. I really thought you already knew*	0.06317
10	*Sometimes I need what only you can provide: your absence*	0.0166666

Figure 2 shows graph which depicts the scores which dangle between positive and negative values with most of the scores falling below the value 0.1 drawing a conclusion for the presence of sarcasm in the sentences or showing the essence of negativity in sentences overpowering the Positive contents in them. The Presence of intense negative words like kill, accident etc gives negative scores.
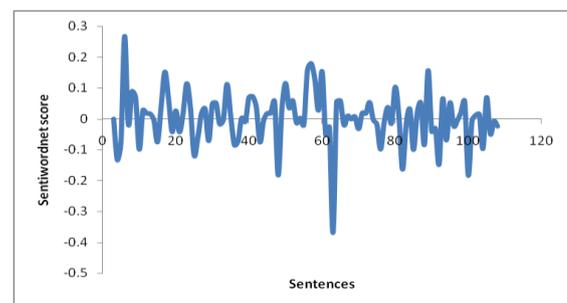

Fig 2: Sentiwordnet score for sentences

The Polarity and subjectivity of the sentences were found using The Textblob. The Polarity of sentences varied from positive to negative depending on the impact of positive and negative

sentimented words in the sentences. The Polarities will not show light towards the presence of sarcasm in sentences.

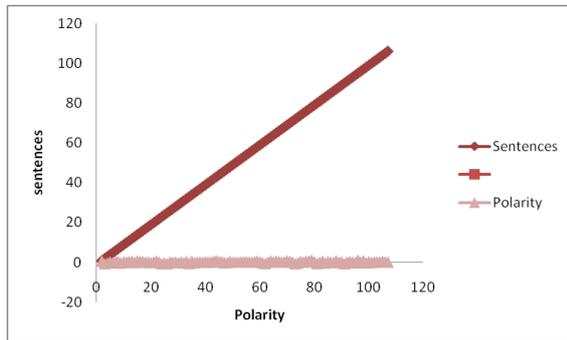Figure 3 shows the Polarity values of each sentences with values ranging from -0.09 to 1.0.
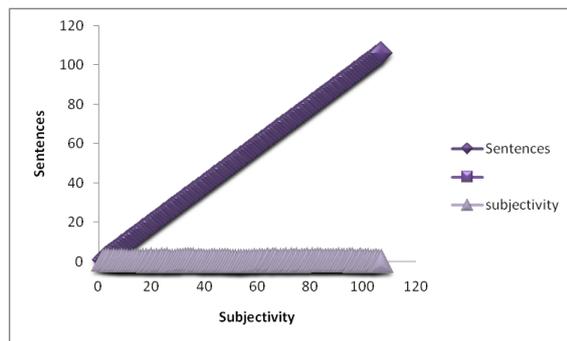

Fig 3: Polarity using Textblob

Fig 4: Subjectivity using Textblob

Figure 4 shows the subjectivity values of each sentences with values ranging from -0.06 to 1.0. The subjectivity and polarity values for particular sentences may not give appropriate information for sentimental analysis in all cases.

A tool for sarcasm detection is developed by Mathieu Cliché **[7]**, which takes sentences as input and gives out scores ranging from negative values to positive values. Negative values convey a message that the inputted sentence is not sarcastic where as a positive value depicted sarcasm in a given sentence. All the sentences were given as input to the sarcasm detector tool and values were found. The values spread from positive to negative with bulk of them being negative even though the inputted sentences were sarcastic. The tool fails at some junctures to detect sarcasm in sentences by providing negative values.
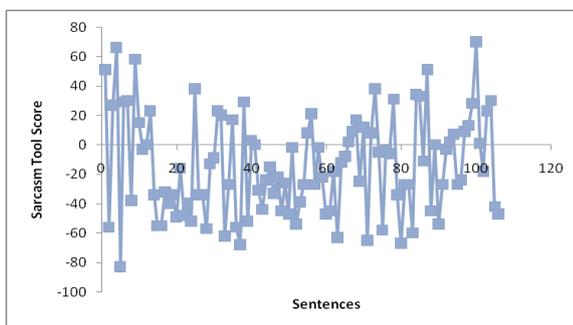

Fig 5: Scores from sarcasm detector tool

Figure 5 shows the values obtained from sarcasm tool which ranges between -56 to +70. According to the tool the positive scores depicts sarcasm and Negative scores represents absence of sarcasm or normal sentences which is not true in our case as all the sentences have some essence of sarcasm in them.

Detection of sarcasm is not simple, as, the text containing sarcasm consists of many factors. Different people come with different opinions to come out with sarcasm. But it is difficult to say or comment on the appropriate method.

4. Conclusion

In this paper an effective approach of detecting sarcasm in sentences is proposed by considering most popular sarcastic sentences. The scores were calculated based on the polarities of sentimented words using Sentiwordnet. Majority of the scores suggested the presence of sarcasm in sentences. The polarity and subjectivity of the sentences were found using the Textblob and also the scores were determined from the sarcasm detector tool. The sarcasm in most of the sentences was not determined in the tool. We have not considered the position of various parts of speech words depicting the sentiments, hence it can be considered for the future study.

References

[1] Ellen riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, Ruihong Hunag, " Sarcasm as Contrast between a Positive sentiment and Negative Situation", Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP, 18-21 October 2013, Seattle, USA.

[2] Mondher Bouazizi, Tomoaki ohtsuki, " Sarcasm Detection in Twitter-All your products are incredibly amazing, are they really ?", IEEE Global Communications Conference (GLOBECOM), 6-10 Dec. 2015, San Diego, CA, USA.

[3] Mondher Bouazizi, Tomoaki ohtsuki, " Opinion Mining in twitter-How to make use of sarcasm to enhance sentiment analysis", IEEE/ACM International conference on Advances in Social Networks and Mining(ASONAM), 25-28 August 2015, Paris, France.

[4] Santosh Kumar Bharti, Korra Sathya Babu, Sanjay Kumar Jena, "Parsing based sarcasm sentiment recognition in Twitter Data", IEEE/ACM International conference on Advances in Social Networks and Mining(ASONAM)), 25-28 August 2015, Paris, France.

[5] S K Bharti, B Vachha, R K Pradhan, K S Babu, S K Jena, " Sarcastic sentiment detection in tweets streamed in real time: A BigData approach", Elsevier, 2016.

[6] Ashwin Rajadesingan, "Detecting Sarcasm on Twitter: A Behaviour Modeling approach", Arizona state University, 2014.

[7] Aditya Joshi, pushpak Bhattacharyya, Mark j Carman, "Automatic Sarcasm detection: A survey", ACM computing Surveys, 2017.

[8] http://www.thesarcasmdetector.com/

[9] http://sentiwordnet.isti.cnr.it/

[10] https://textblob.readthedocs.io/en/dev/