Sarcasm Detection of Twitter Data using Sentiment Analysis

Ms. Shiva Nandini
Assistant Professor/CSE
Aiswarya A, Reshmi Anantapalli, Ria Singh, Sri Haripriya N
Computer Science Department, SRM Institute of Science and Technology
Corresponding authors: Ms. Shiva Nandini

Abstract— Sentiment Analysis is the process of calculating and predicting whether a piece of writing is positive, negative or neutral. It is used to derive the speculation or judgement of the writer, usually an internet user.

Emotions such as happiness, sadness, anger, cautiousness, etc. can be easily classified as positive, negative, or neutral. But sarcasm, being a refined form of irony, might be usually used for various schemes, such as criticism or travesty. However, it is hard even for humans to perceive. Therefore, recognizing sarcastic statements can be very helpful to improve automatic sentiment analysis of the data collected from micro blogging websites or social networks.

The paper proposes a pattern-based approach to detect sarcasm on Twitter. It proposes four sets of features that cover the various types of sarcasm we defined. It uses those to segregate tweets as sarcastic and non-sarcastic by making use of three sets of data, namely training set, optimization set and testing set.

I. INTRODUCTION

Sarcasm being a sharp and bitter expression or remark, mainly known as a taunt can be a very tricky emotion for one to analyse. And for getting this emotion analysed was a whole new challenge till long time.

As we know emotions such as happiness, sadness, anger, cautiousness, etc. can be easily classified as positive, negative, or neutral. But how could one classify the sarcasm that might be used for various purposes, such as criticism or mockery. Therefore, recognizing sarcastic proclamations can be very useful to improve automatic sentiment analysis of the information collected from micro blogging websites or social networks.

Twitter became one of the enormous web destinations for people to express their opinions. Throughout the previous years, Twitter content continued to increase, thus constituting a typical example of the so-called big data. Today, according to its official website,1 Twitter has more than 288 million operative users, and more than 500 million tweets are shared every day

The paper has a pattern-based approach to analyse sarcasm on Twitter. Which proposes four sets of features that cover the various types of sarcasm defined. By using those to classify tweets, which are shared, as sarcastic and non-sarcastic by making use of three sets of data, namely training set, optimization set and testing set.

II. EXISTING SYSTEM

For a set of tweets, the intention is to classify each tweet depending on whether it is sarcastic or not. Hence for each tweet, the extraction of a set of feature, referred as a training set and uses machine learning algorithms to perform the classification was used.

To collect sarcastic tweets, it queries the API for tweets that incorporate the hashtag ''#sarcasm''. In total, collection of 58609 tweets with the hashtag ''#sarcasm'', which was cleaned up by removing the noisy and irrelevant ones alone. As for non-sarcastic tweets, they collected tweets dealing with different topics and made sure they have some emotional content.

The whole study revolved around 3 data sets. First set contains 6000 tweets, half of them are sarcastic, and the other half are not. The tweets on this data set are manually checked and segregated depending on their level of sarcasm from 1 (highly non-sarcastic) to 6 (highly sarcastic).For the second one no manual check is done, which makes it a very noisy data set. The third one, all tweets are manually monitored and classified as sarcastic and non-sarcastic. This set will serve as a test set, and will be used to evaluate.

III. PROPOSED SYSTEM

A randomly collected set of tweets is taken, each one of them depending on whether it is sarcastic or not. Therefore, from each tweet, we extract a set of features, refer to a training set and use machine learning algorithms to perform the classification.
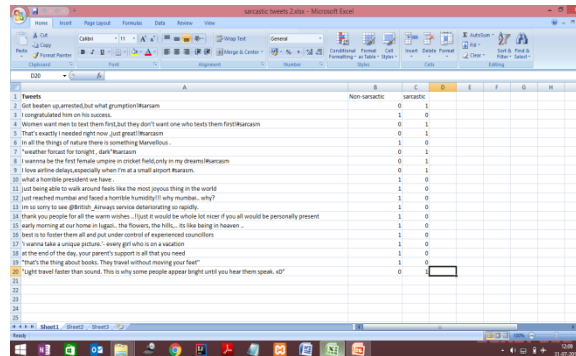

Figure 1: Sample of collected Tweets from Twitter

The features are extracted in a way that makes use of different components of the tweet, and covers different types of sarcasm. The set of tweets on which we run your experiments is checked and annotated manually. The proposed system can also detect sarcasm based on the emoticons used in the tweet.
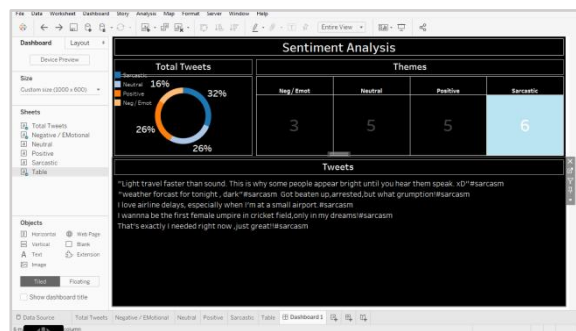

Figure 2: Sentiment Analysis of Sample tweets
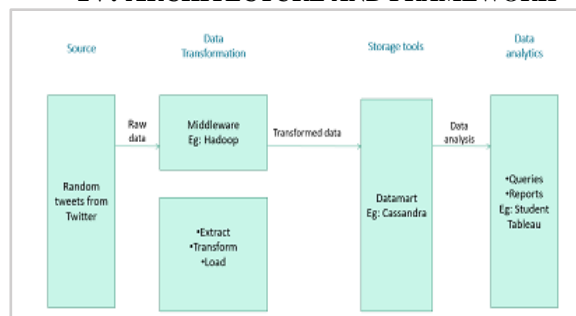
IV. ARCHITECTURE AND FRAMEWORK

Figure 3: Architecture of Proposed System

V. CONCLUSION

The paper gives as an insight about the sarcasm detection system which was implemented successfully. The approach on the sarcasm detection using the twitter data was executed properly and that lead to the desired results.

REFERENCES

[1]. Mondher Bouazizi and TomoakiOtsuki," A Pattern-Based Approach for Sarcasm Detection on Twitter", in Vol 4, pg. no 5477-5488, September 28 2016.

[2]. PyaePhyo Thu and Nwe New ICIS 2017, May 24-26, 2017," Impact Analysis of Emotion in Figurative Language", in IEEE ICIS, pg. no 209-214, May 24-26, 2017.

[3]. Parvesh Kumar Singh and Mohd Shahid Husain," Methodological Study of Opinion Mining and Sentiment Analysis Techniques", in International Journal on Soft Computing (IJSC) Vol. 5, pg. no 11-21, No. 1, February 2014.

[4]. Mondher Bouazizi and TomoakiOtsuki," A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", in vol 5, pg. no 20617-20639, august 18 ,2017.

[5]. Bing Liu," Sentiment Analysis: A Multi-Faceted Problem", in IEEE Intelligent Systems, January 2010.

[6]. Diana Maynard and Mark A Greenwood, "Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis", from pg. no 4238-4243 , In vol 4, July 2016.