

# A Survey on Sentiment Analysis with the help of Sarcasm

Akash Gupta<sup>1</sup>, Arjun Chauhan<sup>2</sup>, Brijesh kumar Jaiswal<sup>3</sup>, Prof. Janhavi Sangoi<sup>4</sup>

Department of Computer Engineering, Mumbai University, India

**Abstract**— various social networking sites has become an emerging platform to express the opinion on various issues. Every day thousands and lakhs of people express their views on twitter on various things. This huge publicly available data can be used for research and a variety of applications by using sentiment analysis. Since, twitter has a limitation of 140 characters per tweet, people are using short forms and the use of slang has also been increased. Moreover, the presence of sarcasm makes the task even more challenging. Sarcasm is a way when a person conveys implicit information, usually the opposite of what is said, within the message he transmits. This paper reviews and compares various algorithms, Naïve Bayes, Decision Tree, Random Forest, Logistic Regression and SVM.

**Keywords**---Sentiment Analysis, Sarcasm, Machine Learning.

## I. INTRODUCTION

In the recent years with the rapid development of the internet, more and more people are inclined to move towards the internet to express their views. Opinion mining is used to know what people think and feel about various products and services in social media platforms. Millions of users share opinion on different aspects of life every day. Due to this growth, media organizations and companies are increasingly seeking way to mine information. User expressions and their impact on organizations are two important characteristics.

Nowadays, thousands and lakhs of people post their reviews and opinions on social media. This huge publicly available data can be used for research and a variety of applications [1]. This can be possible with the help of sentiment analysis, because of which big companies can use that data and figure out whether the people like their product or not, similarly the producers and the film makers will also get to know how positively and how negatively people have reacted on their films, etc. So, sentiment analysis is an important field and has got its applications in various fields like business analytics, etc.

## II. LITERATURE SURVEY

### 2.1 Sentiment Analysis for Sarcasm Detection on Streaming Short Text Data [1].

In this paper sentiment analysis is done by taking sarcasm into consideration. Here, the proposed methodology is classified into three major steps they are: 1. Data pre-processing, 2. Data preparation and 3. Sarcasm Detection.

1. In data pre-processing, i. The hashtags are identified and removed, ii. Emoji dictionary is used to map emoji's to their meaning in the dictionary, iii. Slang dictionary is used to map slangs used in the tweet to their meaning in the dictionary.

2. In data preparation, i. Word tokenization is done, ii. POS tagging is done, iii. Stemming and Lemmitation is done, IV. In feature identification: The major features considered a. Blob polarity, b. Blob subjective, c. Capitalization, d. Positive sentiment, e. Negative sentiment.

### 2.2 Combining Naïve Bayes and Adjective Analysis for Sentiment Detection on Twitter [2].

In the proposed approach, Naive Bayes and adjective analysis is used for the sentiment analysis of Twitter data. In first step, the data is extracted from twitter, here node-xl tool is used for data extraction from Twitter, just enter the search query in it and it will fetch the results. In second step, the data is preprocessed by removing the stop words, URLs, user references, etc. In third step, the pre-processed data is fed as an input to the Naïve Bayes classifier which classifies the tweets as positive and negative, it also results in falsely polarized tweets, called as ambiguous tweets. Now in step four, these tweets are fed as an input to the adjective classifier which classifies the tweets as positive, negative and undetermined.

### 2.3 SentiReviews: Sentiment analysis based on Text and Emotions [3].

Sentiment Analysis is a study of people's attitude, emotions and opinions to classify whether user is "sad", "Angry", "happy" or "joyful" by positive, negative or neutral sentiments. It can be applied on various forms of data such as emoticons, text, images, audio, and video. People post their views, opinions and emotions on social networking sites such as Facebook and Twitter with wide use of emoticons.

Sentiment analysis can be done based on two approaches: Lexicon Based Approach: Lexicon based sentiment approaches use lexicons for calculating polarity of individual words and aggregate their scores to determine overall polarity of text. Machine Learning: In this approach the machine or computer learns the sentiment on regular bases and polarities are assign. It is branch of computer science that provides computers the propensity to learn without program`mming them explicitly.

### 2.4 Improvement sarcasm Analysis using NLP and corpus based Approach [4].

The tweets are classified based on sarcastic or non-sarcastic. The tweets are collected from twitter API.

Twitter data contains various types of data as text or (using @<user>) tags called hash tags. Such as #understand, #beautiful, so that is necessary to process data before applying feature extraction. It includes cleaning, normalization, transformation, Instance selection, feature extraction and selection etc.

In this first action words are classify then compare with corpus of sarcasm data using semantic analysis. This will give score of sarcasm for the given tweet, after comparing the score we polarized the tweets into sarcastic and non-sarcastic. For detection of sarcasm on twitter various methods are used as NLP classifier, Tokenization, and POS tagging. This shows a good result on real time database of twitter.

### 2.5 Real Time Sentiment Analysis of Tweets Using Naive Bayes [5].

Sentiment is done in several stages the database used is twitter API the recent twits are taken in consideration. The steps are has follows 1 the selection of training data, which is based on which type of sentiment, we have to perform. 2 The pre-processing of trained data which includes removing of irrelevant data like URL's, User names, Slang words, Symbols etc. in sort separating the types of twits 3 Establishing connection between Twitter database using twitter API 4 Various Machine-learning algorithm are used like in this paper Naive Bayes, Support Vector Machine for classifying purpose of twits into various classes. 5 Result displayed on the basis of Polarization of the overall twits considered.

### 2.6 A sentiment Analysis method of short Texts in microblog [6].

The paper proposes a sentiment analysis of text in microblog by checking dependencies of the sentences of text and obtained dependency relation of words. Dependency parsing:

The analyzing dependency relation of text that reveals its syntactic structure. In this text checked the subject, verb and object relations among text.

Example: "I really like cats."

In this sentences verb is "like" that relation with "I", "really" and "cats".

Dependency Relation=<IDm, content.m, IDn, content.n, relation>

### 2.7 Opinion Mining in Twitter: How to Make Use of Sarcasm to Enhance Sentiment Analysis [7].

In this paper, the feature extraction is done in two sets i.e. Non-textual features and Textual features. In non-textual features, 6 features are extracted by counting the number of positive and negative Hashtags, Emoticons and slang words. In textual features, firstly the "NOT" tag is added to all the words between a negation word and the first following punctuation. Then, two lists of words i.e. positive words list and negative words list are used. The two lists contain respectively words that have a negative

emotional content and words that have a positive emotional content. Using these lists 4 features are extracted: the number of positive words and that of negative words, highly emotional positive words and that of highly emotional negative words.

### 2.8 Sarcasm Detection of non # tagged statements using MLP-BP [8].

In the proposed system, first tokenization is performed on the dataset then, feature extraction is done. After this all the features extracted are given as an input to the Multi-Layered-Perception Back-Propagation algorithm which classifies the sentences into sarcastic and non-sarcastic. Feature Extraction is done in two categories: 1. Structural features, 2. Affective features 1. Structural features: These features are based on structures i.e. structures of Punctuation marks, Discourse Markers and Emoticons, these features help in detecting sarcasm by analyzing how much time a punctuation mark is used or what type of emoji the user has used in the sentence, etc. 2. Affective features: In this, Whissell Dictionary of Affect in Language (DAL) is used to calculate the scores for pleasantness, activation and imagery of words.

### 2.9 Sentiment Analysis using Latent Dirichlet Allocation and Topic polarity world cloud visualization [9].

Alternate method for sentiment analysis is Latent Dirichlet Allocation (LDA) that extracts the topic of documents where the topic is represented as the appearance of the words with topic probability. Several methods have been implemented to analyze sentiment on documents, such as Naive Bayes classifier and SVM. Data comments are collected from students toward their university then preprocessing is performed on the collected data then Tokenization, stop words removal and forming bag of words. It shows negative sentiment in six variation and positive sentiment in three variations.

### 2.10 Parsing-based Sarcasm Sentiment Recognition in Twitter Data [10]

In this sentiment is done on famous micro blogging site twitter using two approaches for sarcasm detection. Firstly the parsing-based lexicon generation algorithm (PBLGA) and then to detect sarcasm based on the occurrence of the interjection word. The combined approach also been shown and compared with the existing state-of-the-art approach to identify sarcasm. The first section is parsing-based lexicon generation algorithm (PBLGA). A parsing based lexical generation algorithm (PBLGA) for sarcasm identification in Twitter data is required to recognize sarcasm in both forms of tweets: (a) contradiction of negative sentiment and positive situation (b) contradiction of positive sentiment and negative situation.

### 2.11 Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm [11].

In this paper sentiment is done on famous microblogging website, twitter which produces millions of views on certain topic which is use to predict various views on business related topics like sales and marketing. The limitations of available systems are not sufficient to deal with the complex structure of the big data. In the Map Phase the tweets are segregated using SentiWordNet dictionary in various POS according to the polarities and mapped them to the table using hashes further using this hashes it is further pass to the Reduce Table. In Reduce Phase the overall polarity of each tweets are taken and transformed in 5 different categories as extreme positive, positive, extreme negative, negative, neutral.

## III. ANALYSIS TABLE

The Analysis table is the summary of the studied research papers and the different techniques used on Sentiment analysis of various texts.

Sr No.	Title	Technique Used	Dataset	Accuracy
1	Sentiment Analysis for Sarcasm Detection on Streaming Short Text Data [1]	Data Preprocessing with emoji and slang dictionary and data preparation	Manually classified twitter dataset	A roughly of 1 % increase has been gained

2	Combining Naïve Bayes and Adjective Analysis for Sentiment Detection on Twitter [2]	Naïve Bayes and corpus of adjectives and adverbs	IMDB Movie Dataset	Increased by 8% i.e. 88.5
3	SentiReviews: Sentiment analysis based on Text and Emotions[3]	Machine learning approach for constantly analyzing the polarities.	Twitter API Weibo	Stating various methods
4	Improvement sarcasm Analysis using NLP and corpus based Approach [4]	NLP and CORPUS based approach for detect sarcasm on Twitter.	Twitter API	It performs analysis on real time data.
5	Real Time Sentiment Analysis of Tweets Using Naive Bayes [5]	Naïve Bayes and Natural Language Toolkit (NLTK) integrated with WordNet	SENTIWODNET Twitter API, Large Movie review Dataset (LMRD)	Previous shows about 58.4% increased using SentiWord along with Naive Bayes
6	A sentiment Analysis method of short Texts in microblog [6]	Dependency parsing with sentiment relationship migration and modified distance for sentiment analysis of short text in microblogs.	Microblog data  COAE2014	Accuracy is 73.8%  Increased by 4%
7	Opinion Mining in Twitter-How to make use of sarcasm to enhance sentiment analysis [7]	Various feature extraction for sentiment and sarcasm detection.	Twitter dataset	Accuracy was 83.9% Increased by 2% i.e. 85.9%
8	Sarcasm Detection of non # tagged statements using MLP-BP [8]	Structural, affective and semantic similarity features using MLP-BP	Reddit dataset	As sarcasm is considered accuracy will improve.
9	Sentiment Analysis using Latent Dirichlet Allocation and Topic polarity worldcloud visualization [9]	Determined the best combination of parameters for sentiment analysis using LDA.	Data comments are collected from students toward their university	F-measure by 61% as compare to Naïve Bayes and Logistic Regression
10	Parsing-based Sarcasm Sentiment Recognition on Twitter Data [10]	PBLGA along with Interjection Word Start (IWS)	Twitter API	P- .89 R- .96 F- .90
11	Sentiment analysis on twitter Data-set using naïve Bayes algorithm [11]	Frame worked implementation of Naive Bayes Algorithm	SentiWordNet, Twitter API	Accuracy increased by converting emotions to equivalent word

#### IV. CONCLUSION

In this paper various machine learning algorithms for sentiment analysis has been reviewed and compared and found that pre-processing helps a lot in achieving better accuracy. After reviewing all the machine learning algorithms, it is proved that, Naïve Bayes when combined with adjective analysis gained a remarkable accuracy.

#### REFERENCES

- [1] Anukarsh G. Prasad, Sanjana S, Skanda M. Bhat, B S. Harish, "Sentiment Analysis for Sarcasm Detection on Streaming Short Text Data", IEEE 2<sup>nd</sup> International Conference on Knowledge Engineering and Application, 2017, pp 1-5.
- [2] Mohit Mertiya, Ashima Singh, "Combining Naïve Bayes and Adjective Analysis for Sarcasm Detection on Twitter", IEEE International Conference on Inventive Computation Technologies (ICICT), 2016, pp 1-6.
- [3] Ms. Payal Yadav, "SentiReview: Sentiment Analysis based on Text and Emoticons" International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2017, pp 467-472.
- [4] Manoj Y. Manohar, "Improvement sarcasm Analysis using NLP and corpus based Approach" IEEE International Conference on Intelligent Computing and Control Systems ICICCS, 2017, pp 618-622.
- [5] Ankur Goel, Sitesh Kumar, "Real Time Sentiment Analysis of Tweets Using Naive Bayes", IEEE International Conference on Next Generation Computing Technologies (NGCT), 2016, pp 257-261.
- [6] Jie Li; Lirong Qiu, "A Sentiment Analysis Method of Short Texts in Microblog" IEEE International Conference on Computational Science and Engineering (CSE), 2017, pp 776-779
- [7] Mondher, 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 Analysis and Mining, 2015, pp 1594-1597
- [8] Pornima Gidhe, Dr. Leena Ragha, "Sarcasm Detection of non # tagged statements using MLP-BP", IEEE International Conference on Advances in Computing, Communication and Control (ICAC3), 2017 pp 1-4.
- [9] Mohammad F. A. Bashri, Retno Kusumaningrum "Sentiment Analysis Using Latent Dirichlet Allocation and Topic Polarity Wordcloud Visualization" Fifth International Conference on Information and Communication Technology (ICoICT), 2017, pp 1-5.
- [10] 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 Analysis and Mining, 2015, pp 1373-1380
- [11] Huma Parveen, Prof. Shikha Pandey, "Sentiment Analysis on Twitter Data-set using Naïve Bayes Algorithm" IEEE 2<sup>nd</sup> International Conference on Applied and Theoretical Computing and Communication, 2016 pp 416-419.