SENTIMENT ANALYSIS USING DATA MINING ON SOCIAL PLATFORMS SUCH AS TWITTER

Asif Siddiqui , Vibhu Priyadrashi , Nikita Warke , Anuja Kumbharde

Department of Computer Engineering

Dr. D.Y. Patil Institute of Technology,Pimpri,Pune

Abstract **:-** In recent years, there has been a rapid growth in online communication. As a result of this, many social networking services and related mobile applications have emerged. Huge amount of data is generated by these sites on a daily basis and this data can be used as a source for various analytical purposes. Twitter is one of the most popular networking sites with millions of users. There are users with different views and a variety of reviews are generated by them. Nowadays, Opinion Mining has become an emerging topic of research due to a lot of opinionated data available on blogs and social networking sites. Tracking different types of opinions and summarizing them can provide valuable insight into products, services or other topics. Analysis of opinions and their classification on the basis of polarity (positive, negative, neutral) is a challenging task factoring in sarcasm.

Keywords ***:-Opinion Mining; Social Networking; Polarity; Sentiment Analysis.***

1. Introduction

Sentiment analysis involves natural language processing, text analysis and computational linguistics for identification and extraction of subjective information in source materials. Generally speaking, the aim of sentiment analysis is determining the attitude of a speaker or a writer with respect to the topic in question or the overall contextual polarity of the content. The attitude may be his or her judgment or evaluation of affective state, or the intended emotional communication. Humans have the innate ability to determine sentiment. However, this process is time consuming, inconsistent, and costly in a business context. It's just not realistic to have people individually read thousands or even millions of user customer reviews and evaluate them for the polarity of the sentiment along with factoring in the sarcasm.

Approaches And Methods

1.1. Rule-based Approaches

Usually, rule-based approaches define a set of rules in some kind of scripting language that identify subjectivity, polarity, or the subject of an opinion. The rules may use a variety of inputs, such as the following:

- S.K. Bharti and B. Vaccha "Sarcsatic sentiment detection in tweets Streamed in real time" Techniques like *stemming*, *tokenization*, *part of speech tagging* and *parsing*. Other resources, such as Lexicons (i.e. lists of words and expressions).

The following is an example of a rule based approach:

1. Defining two lists which consist of the polarized words-
   For example: negative words such as poor, underwhelming, *worst*, bad, ugly, etc and positive words such as overwhelming, best, good, beautiful, etc.
2. For a given text:

- Counting the number of positive words appearing in the text.
- Counting the number of negative words appearing in the text.

3. If the number of occurrences of positive words is greater than the number of occurrences of negative words, then a positive sentiment is returned. Return a negative sentiment for an opposite scenario. Otherwise, for a comparable number of positives and negatives, return neutral. Such a system is very naïve as it doesn't account for the combination of these words for the formation of sentences, which may result in sentences with a completely different meaning especially when sarcasm is used. A more advanced processing can be made, but such systems get very complex quickly. These systems are also very hard to maintain as new rules have to be introduced anytime a new expression is encountered. Besides, there might be undesired outcomes upon introduction of new rules conflicting with the old ones. As a result, these systems require important investments in manually tuning and maintaining the rules.

1.2. Automatic Approaches

Apoorv Agarwal and Boyi Xie "Sentiment Analysis of Twitter Data" Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. The sentiment analysis task is usually modeled as a classification problem where a classifier is fed with a text and returns the corresponding category, e.g. positive, negative, or neutral (in case polarity analysis is being performed).

The Training and Prediction Processes

In the training process, the model learns to associate a particular input (i.e. a text) to the corresponding output (tag) based on the test samples used for training. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g. *positive*, *negative*, or *neutral*) are fed into the machine learning algorithm to generate a model. In the prediction process, the feature extractor is used to transform unseen text inputs into feature vectors. These feature vectors are then fed into the model, which generates predicted tags (again, *positive*, *negative*, or *neutral*).

Detecting Sarcasm

General sentiment analysis holds major drawback for sarcasm detection. When we consider twitter data, the use of sarcasm is ever increasing. There is a wide range of sarcastic tweets being published every second. As a result these tweets bring down the accuracy of the classification algorithms.

For example, a tweet saying "Did you know that you can actually ignore the new whatsapp update and carry on with your life?" will be classified as simple positive-or-negative tweet rather than being classified as a proper sarcastic tweet. However one can clearly classify the given tweet in a negative emotion conveyed very sarcastically.

Proposed Approach

We can analyse the data based on the sentence structure and also take into account the context with respect to which the message has been posted. Several messages on their own may have a very different meaning than what was intended by the user. Several abbreviations such as r(are), u(you), fyi(for your information), cmon(come on), wid(with) etc. will also be processed by the algorithm. The initial stage will consist of lexical analysis which detect the lexical features such as unigram, bigram and n-gram. The next stage will have Hyperbole analysis where interjections such as 'wow', 'aha' etc. will be processed. Also, along with interjections, punctuation marks (?, ! etc.), quotes ("", '') and intensifiers (Adjectives and Adverb) will also be analysed. The last stage will be of pragmatic analysis where we can deal with emoticons, smilies and replies to messages.

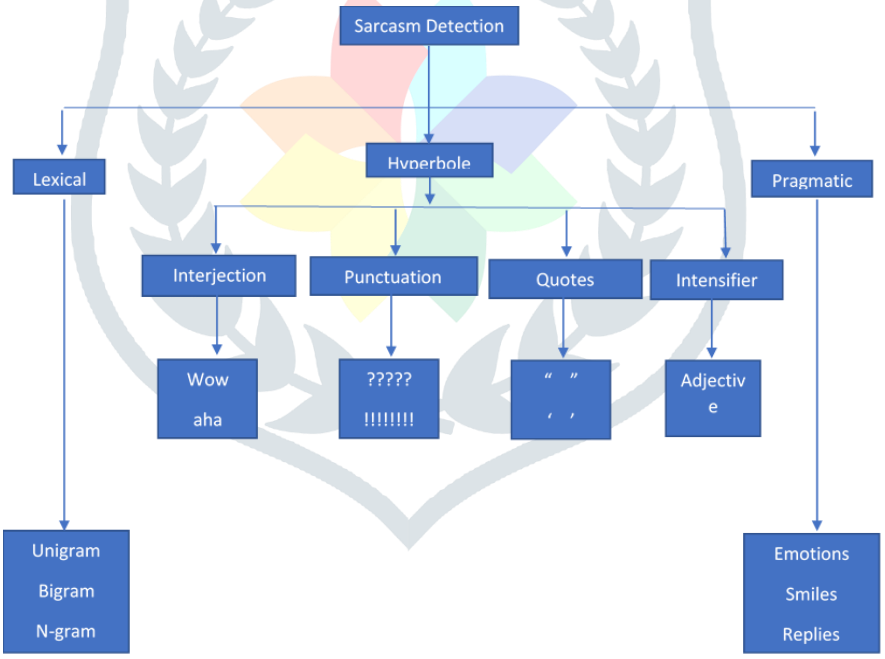

Fig: Proposed Architecture for Sarcasm Detection

System Architecture

The purpose of the implementation is to be able to automatically classify a tweet as a positive or negative tweet sentiment wise. The list of word features are extracted from tweets i.e., Twitter API and are collected together in the database which is followed by the pre-processing of the collected data. The classifier is trained by the training set and then classifies the tweets as per the type of sentiment that it detects. The classifier is able to detect the polarity of the tweet and generate the results.

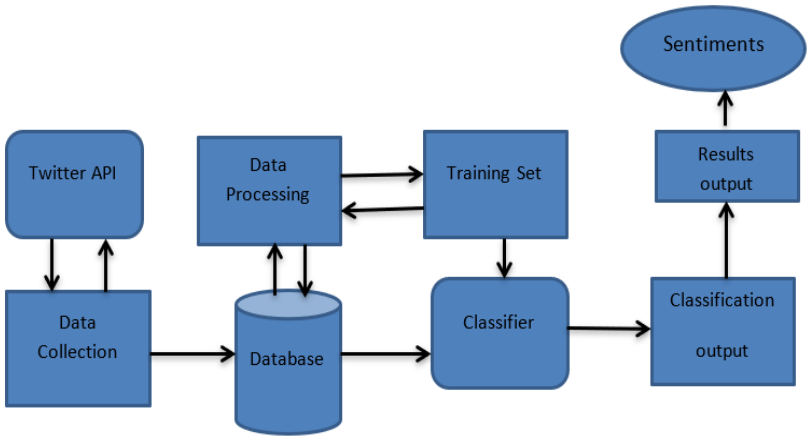

Fig: System Architecture

Conclusion And Future Work

Final system will be an application which will be able to find out the polarity of the Twitter users regarding several topics. Users will get all information at one place. Users can analyse and refer the results obtained and choose to take appropriate steps. This system will provide a platform for the users to retrieve information related to their interest. Currently there is no such common platform available in the market which facilitates this.

Any future work will consist of modifications to the system so that it can make suggestions to the users based on the results of the analysis.
Also, it will focus on improving the accuracy of the application further and keep it as close to ideal as possible.

Acknowledgements

This project was supported by Dr. S.N Mali, Principal, Dr. D.Y. Patil Institute of Technology,Pimpri,Pune. We thank our Head of Department Dr. Pramod D. Patil who provided insight and expertise that greatly assisted the research. We thank our project guide Mrs. Ashwini Dhoke for assistance and comments as well as sharing her pearls of wisdom with us during the course of this research, that greatly improved the project.

References

1. Apoorv Agarwal., Boyi Xie., Ilia Vovsha(2013) Sentiment Analysis of Twitter Data, Department of Computer Science, Columbia University

2. Bilal Alatas., Cem Baydogan., (2018) Sentiment Analysis Using Konstanz Information Miner in Social Networks[2018 6th International Symposium on Digital Forensic and Security (ISDFS)]

3. Chae Won Park., Dae Ryong Seo.,(2018) Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants [2018 5th International Conference on Industrial Engineering and Applications (ICIEA)]

4. Chiyu Cai., Linjing Li.,(2016) New Words Enlightened Sentiment Analysis in Social Media [2016 IEEE Conference on Intelligence and Security Informatics (ISI)]

5. Hanif Bhuiyan., Jinat Ara.,(2017) Retrieving YouTube Video by Sentiment Analysis on User Comment[IEEE International Conference on Signal and Image Processing Applications (ICSIPA)]

6. Mondher Bouazizi., Tomoaki Ohtsuki., (2015) Opinion Mining in Twitter How to Make Use of Sarcasm to Enhance Sentiment Analysis [2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining]

7. Paras Dharwal., Tanupriya Choudary., (2017) Automatic Sarcasm Detection using Feature Selection[International Conference on Applied and Theoretical Computing and Communication Technology]

8. Pornima Gidhe., Dr. Leena Ragha., (2017) Sarcasm Detection of non tagged statements using MLP-BP [2017 International Conference on Advances in Computing, Communication and Control (ICAC3)]