Political Perception in India about Public Opinion using Sentiment Analysis on Twitter

Sayali Jori, Shraddha Phansalkar

Abstract: *The political inclination of a country plays important part in a country. Twitter data is analyzed by using various sentiment analysis techniques to give insight of public opinion regarding political perception. The use of emoticons and emoji has increased on large scale to express feelings. The sentiment orientation of emoticons and text are related. The analyzing sentiment orientation of emoticons and text together boosts the performance of sentiment analysis system. The effect of sarcasm in political tweets is analyzed The proposed system consists of analyzing political tweets using four different techniques of sentiment analysis word wise sentiment analysis, emoticons wise sentiment analysis, emoji wise sentiment analysis and combined sentiment analysis. The accuracy of a system is validated by using Support Vector Machine. The effect of sarcasm on political tweets is analyzed using Support Vector Machine algorithm.*

Index Terms: Sentiment Analysis, Pre-processing, Natural Language Processing, SVM Algorithm, Training Data.

I. INTRODUCTION

Research Motivation

Twitter is popular micro blogging site. The wide variety of people like politicians, journalists share their opinion and feelings on the twitter. The lot of literature survey has been done on twitter related to political data. It is quite interesting to know political inclination of people. All this points along with data motivated us to do this research.

Twitter

Twitter is most popular social media site. The lot of tweets is generated every day. There are a million numbers of users on the twitter. The people discuss current affairs and share their feelings on twitter. The short text message size of twitter is flexible to use.

Sentiment Analysis

Sentiment analysis aims to find out opinion expressed regarding particular topic in the text.

There are different classes of sentiment analysis:

Positive class – These are good words about the target.
Negative class – These are bad words about the target.
Neutral class – These are neither good nor bad word about the target.
Different techniques of sentiment analysis:
Machine learning based sentiment analysis.
Lexicon based sentiment analysis.
Hybrid based sentiment analysis.
Machine learning based sentiment analysis provides better performance, but it require huge labeled training data set and it is domain dependent. Machine learning techniques required labeled positive and negative examples. The features are extracted from these training examples, the classifier is trained and then the test data is classified.
Lexicon based techniques gives good results by using large dictionaries but it provides high precision and low recall. Lexicon based techniques makes use of dictionaries. The dictionary consists of a list of words with their polarity scores.
Hybrid techniques give accuracy of machine learning algorithm and stability of the lexicon based system.
A choice of methodology heavily depends on application, domain and language.
Sarcasm Detection: Sarcasm detection is used to express negative feelings. In political tweets sarcasm detection helps to identify dissatisfaction of people about particular political party or political event.
Sarcasm in the twitter is in the form of:
Positive words with negative emoticons.
Negative words with positive emoticons.

II. RELATED WORK

Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Noureen Hameed gave detail analysis about a variant of SVM from year 2012 to 2017 [1].
Widodo Budiharto, Meiliana Meiliana proposed a system for sentiment analysis of data related to 2019 presidential election. The data collected using twitter 4J API and analysis was done based on tweet counting and sentiment analysis process using R based on pre-processing [2].
Munir Ahmad, Shabib Aftab, Iftikhar Ali proposed a system using SVM algorithm in weka for polarity detection of tweets [3].
Sandip Roy proposed a system for predicting public opinion about the winning party of 14th Gujarat Legislative Assembly election, 2017. The emotion lexicon was used to determine polarity using eight emotions. Deep learning tool was used for predicting results [4].

Ayeena Malik, Divya Kapoor, Amit Prakash Singh proposed a system to analyse political tweets using AFINN dictionary [5]. Josemar A. Caetano, Hélder S. Lima, Mateus F. Santos and Humberto T. Marques-Neto proposed a system for analysis of homophily in election tweets using lexicon dictionary SentiStrength [6].

Martin Ringsquandl and Dusan Petkovic proposed a system for prediction of sentiment analysis of presidential candidates of the Republican Party in the USA and their campaign topics using lexicon based techniques word wise and aspect based sentiment analysis [7].

Mohammed O. Shiha, Serkan Ayvaz analyses use of emoji characters increases overall sentiment analysis performance [8].

Wiesław Wolny analyzed use of natural language processing and symbol analysis for analyzing emoticons [9].

D. K. Tayal, Sumit Yadav, Komal Gupta, Bhawna Rajput, Kiran Kumari, proposed a system to identify sarcastic tweets and determine polarity of sarcastic tweets [10].

Sana Parveen, Sachin N. Deshmukh proposed a system for sarcasm detection using classifiers like Support Vector Machine, Naive Bayes and Maximum Entropy [11].

III. METHODOLOGY

The main aims of the system are as follows:

Detection of sentiment of political tweets using word wise, emoji and emoticon wise and combined sentiment analysis techniques.

Compare the accuracy of a system using Support Vector Machine.

Study the effect of sarcasm on political tweets.

System Architecture

Fig1. System Architecture

Technical Requirements

Software required is java jdk 1.7, net beans, MySQL database. The system should have internet connectivity to fetch real-time tweets.

Fetching data

The real time twitter data is fetched using Twitter 4J API. The data related to different hash tags related to political events, political candidates can be fetched. Here the data related to political leaders like #namo, #rahulGandhi, #mamta, #kejriwal are fetched. The data related to political events like #Notebandi, #Rafale is also fetched. The analysis of system is done on tweets collected prior to the elections. The data set size grows by fetching data.

The tweets contain some unwanted noise. It is necessary to remove unwanted noise. So preprocessing is applied.

Fig2. Data set related to the political leader

Pre-processing

It is the process of converting data in understandable format.

The preprocessing is done using:
Removal of stop words
Removal of URL
Removal of special symbols
Applying porter stemmer algorithm

Removal of stop words: The stop words have provided no meaning in sentiment analysis. Hence, stop words are replaced by blank spaces. In this project the dictionary of stop words is maintained.

Removal of URL and special symbols: The URL and special symbols have no meaning in sentiment detection. Hence they are replaced with the blank spaces.

Applying porter stemmer algorithm: It is process of converting word to its root word because of these many words are merged, dimensionality is reduced and accuracy of a system increases.

Sentiment analysis using different techniques
Word wise sentiment analysis using Natural Language Processing:
The dictionary of words is maintained along with the polarity score of words. The polarity of individual word is added for single tweet, and sentiment is analyzed for that tweet.
In this project AFFIN dictionary is used.
It consists of words with polarity valence score from plus 5 to minus 5. It was developed by Finn Krup. It internally consists of POS-Tagging.
The words are tokenized in tweets. The word is found in AFFIN dictionary, and its score is determined.
Emoticon wise sentiment analysis
The emoticons represent visual expressions. They are form using punctuation marks. In this project different emoticons lists is maintained.
The positive emoticon list, negative emoticon lists and neutral emoticon list is maintained.

The emoticon from tweets is split, and it is compared with these lists. According to match the score is assigned to emoticons.

Emoji wise sentiment analysis

Emoji represents facial expressions. In this project 410 emoji are stored in MySQL with their sense, tags, meaning, Unicode and scores. The emoji wise sentiment analysis is performed using Unicode. Based on Unicode matching the sense of emoji is determined.

Combined sentiment analysis

The emoticons and emoji scores are compared and the total score is added in polarity score of text.

Sarcasm detection

The sarcasm detection is done using SVM in weka. The LibSVM library is used to classify the instance. Here training data set file stress.arff is used. The stress.arff file consists of four different parameter combinations. The score values of emoji, emoticon, sentence sentiment analysis and fourth parameter is result. The testing tweet is given to different sentiment analysis techniques. The output of emoticon wise, emoji wise, sentence wise analysis is stored in feature array. The feature array is tested against training data set and result is determined as sarcastic or non-sarcastic. The sarcasm detection is validated in weka tool using four different algorithms

Accuracy using SVM

SVM algorithm is linear classifier where separating hyper plane minimizes classification error of test data.

The labeled data set from kaggle is considered as training data and preprocessed data set of a project is given as testing data. The classification is done using SVM in python using scikit-learn library of machine learning.

The training data set consists of 53,883 records, based on unigram, bigram feature vectors, extracting features and transformation process using TF-IDF and applying linearSVC classifier classification is done. The accuracy of SVM algorithm is 68%.

IV. APPLICATION

The system is used to give insight of public opinion regarding political candidates as well as political events using different sentiment analysis techniques.
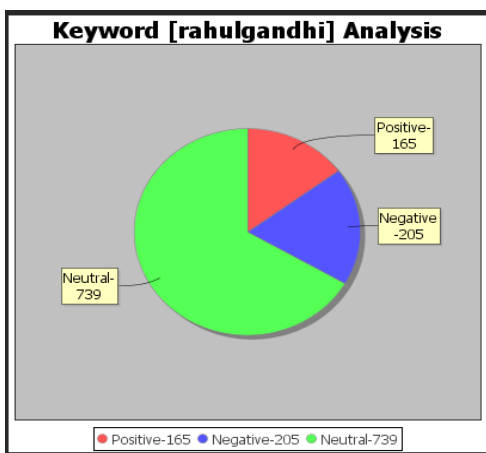
V. RESULTS

Fig3. Sentiment analysis of tweets related to hash tag Rahul Gandhi using hybrid approach
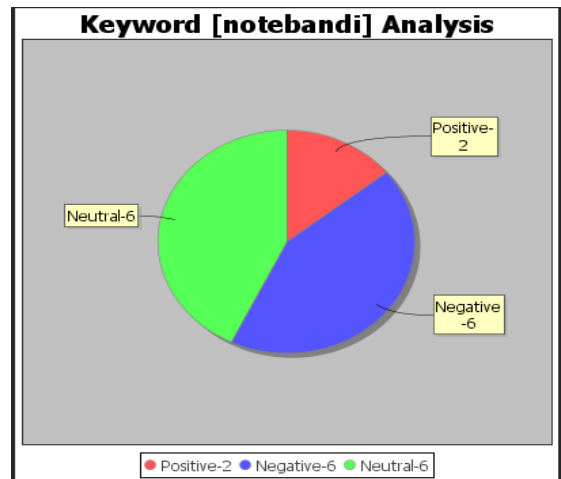

Fig4. Sentiment analysis of a political event Note Bandi using hybrid approach
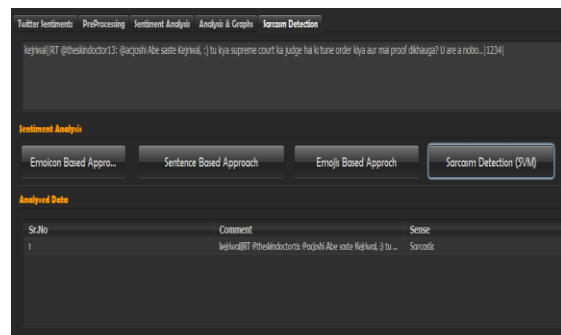

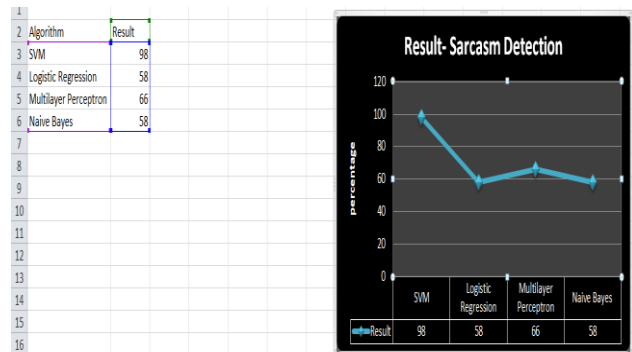**Fig5. Sarcasm detection in political tweets using SVM**

Fig6. Sarcasm detection accuracy using different algorithms in weka

Fig7. Accuracy of a system using SVM algorithm

VI. CONCLUSION

Lexicon based technique along with the emoticon and emoji based technique together boosts the performance of sentiment analysis system by giving insight of public opinion regarding political perception in India. The effect of sarcasm on political tweets is studied.

REFERENCES

1. Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Noureen Hameed, "Sentiment Analysis Using Svm: A Systematic Literature Review", (Ijacsa) International Journal Of Advanced Computer Science And Applications, Vol. 9, No. 2, 2018.
2. Widodo Budiharto And Meiliana Meiliana, "Prediction And Analysis Of Indonesia Presidential Election From Twitter Using Sentiment Analysis", Journal Of Big Data December 2018.
3. Munir Ahmad, Shabib Aftab, Iftikhar Ali, "Sentiment Analysis Of Tweets Using Svm", International Journal Of Compute Applications 177(5):975-8887 · November 2017.
4. Sandip Roy, "Analyzing Political Sentiment Using Twitter Data", May 2018.
5. Ayeena Malik, Divya Kapoor, Amit Prakash Singh, "Sentiment Analysis On Political Tweets", Vth International Symposium On "Fusion Of Science & Technology", New Delhi, India, January 18-22, 2016.
6. Josemar A. Caetano, Hélder S. Lima, Mateus F. Santos And Humberto T. Marques-Neto, "Using Sentiment Analysis To Define Twitter Political Users' Classes And Their Homophily During The 2016 American Presidential Election", Journal Of Internet Services And Applications 2018 9:18.
7. Martin Ringsquandl And Dušan Petković, "Analyzing Political Sentiment On Twitter", 2013, Association For The Advancement Of Artificial Intelligence.
8. Mohammed O. Shiha, Serkan Ayvaz, "The Effects Of Emoji In Sentiment Analysis", International Journal Of Computer Electrical Engineering, Volume 9, Number 1, June 2017.
9. Wiesław Wolny, "Emotion Analysis Of Twitter Data That Use Emoticons And Emoji Ideograms", 25th International Conference On Information Systems Development (Isd2016 Poland).
10. D. K. Tayal ; Sumit Yadav ; Komal Gupta ; Bhawna Rajput ; Kiran Kumari, "Polarity Detection Of Sarcastic Political Tweets", 2014 International Conference On Computing For Sustainable Global Development (Indiacom).
11. Sana Parveen, Sachin N. Deshmukh, "Opinion Mining In Twitter – Sarcasm Detection", International Research Journal Of Engineering And Technology (Irjet) Volume: 04 Issue: 10 | Oct -2017.