Analysis and Prediction over Indian Election Dataset

Goutam Majumder

Assistant Professor, Department of Computer Science & Engineering,

Lovely Professional University, Jalandhar, Punjab – 14411, India.

Abstract – Sentiment analysis on online reviews has become a hot research field for every industry. Study of social network is the study of the interactions and correspondence of people on different topics and it has gained more attention nowadays. Millions of people on social media such as Facebook and Twitter give their opinions on different topics on a daily basis. It has many uses in different fields of research, from the social sciences to industry. Elections are made to view the public opinion, where a group of people chooses a candidate by means of votes, many methods are used to predict the outcome. Many agencies and media companies conduct pre-screening surveys and expert opinions to predict election results. We use twitter data to predict the outcome of the election by collecting twitter data and analysing it to predict the outcome of the election by analysing the sentiment of the candidate's twitter data. We used machine-based lexicon approach to finding emotions in twits and predicting sentiment score.

Keywords – Sentiment Analysis, prediction analysis, Twitter API

1. Introduction

This works presents an analysis of elections in India of static and real-time data. For this purpose, we have used big data and machine learning approaches to visualize and analyse the data of the general election [1]. This work also contains the sentimental analysis of twitter stream data. In the era of technology, every political fraternity using these technologies as a weapon to contest the election and this is the resulting game-changer of the election. In the recent general elections in India, technology has made it easier for parties to measure the mood of the voters and so that they make their strategies accordingly [2], [3].

Political parties use big-data analytics to gain insights into voting preferences based on their socio-economic status, caste, local issues and other parameters. Based on the voter opinion and segments, customized election campaigns with the most important advertisements and videos are created and geared towards the specific target groups [4]. Once the campaign is started, data is collected to evaluate its effectiveness and further tweak it according to the response it produces. Not only the campaign, but also the election strategy and the fielding candidates were science with big data also leading the parties here. Analysis of data from various applications – offering insights into the political experience of MPs and MLAs and evaluating their performance – provides insight into a candidate's popularity (or otherwise) and helps parties to make informed decisions about the best candidate for an electorate [5].

Twitter is the best platform for social media that has sensible comments and tweet data. At this platform person who is tweeting then he is serious about his tweet. Twitter's developer framework offers many API products, tools, and services that allow you to harness the power of the free, global, and real-time network of Twitter communications [6].

In this work, we propose to scrape twitter for varied posts and order them into a real-time dataset. We then propose to examine the structure of this data set. Through this data set, we propose to analyse what the current topics of discussions are about, what the perceived opinions of the users are about the various political leaders and parties. Analyse data regarding discussions about political trends, events, and other policy decisions. We propose to scrape the data using APIs provided by twitter and apply data mining and sentiment analysis techniques on the scraped data. Through this analysis, we expect to gain a clear insight into how users of such sites behave and interact with one another while also gaining insights into their opinions and biases towards various issues [7], [8].

Twitter is a social news aggregation and discussion website with large communities of people across different categories. This work will focus on mining/extracting data efficiently from real-time tweets and analysing that data to form meaningful conclusions about the usage of the website. Sentiment analysis or opinion mining is the statistical study of the beliefs, thoughts, behaviours and emotions of people expressed in written language. In recent years, it is one of the most active areas of research in the natural language processing and text mining. So we used the twitter framework to analyse the data in real-time and stream. Apart from this, we have also analysed the static structured data of general election result through which can able do the analysis of results. This project also contains an android app that has a feature of rate your leader and party.

2. Scope of the work

Social Media platforms have become quintessential for user-generated content and consumer opinions. The Twitter API and code are open-sourced so twitter's comments can be readily fetched. To expand on that recent reviews have been given by using framework tools like python, R, twitter API, matplotlib, and NLP. Sentiment and data analysis has become a million-dollar industry and it becomes necessary for political parties to invest in it. With the help of this technology, they can able to gauge the sentiments of voters and put their message forward accordingly in a very effective way. Not only sentiments of voters also parties can set the fielding of their candidates on the basis of caste, religion, culture, colour, and underdog candidates. This work will help to run the propaganda and campaign of the election with statistic of various state wise general Lok Sabha election party wise results.
It also concludes age-wise candidate win and loss so it can help in the distribution of tickets.

3. System Design

Flow of the system is depicted in Fig. 1. At first tweets are crawled and consider this data as unlabelled data. There after stop words are removed and stemming also performed as part of the preprocessing. In second step crawl word mood is applied over unlabelled data pass it for two different process. Various machine learning algorithms are adopted for prediction purpose and for sentiment analysis. At last stage tweets are analysed based on tweet moods, retweets count and locations.

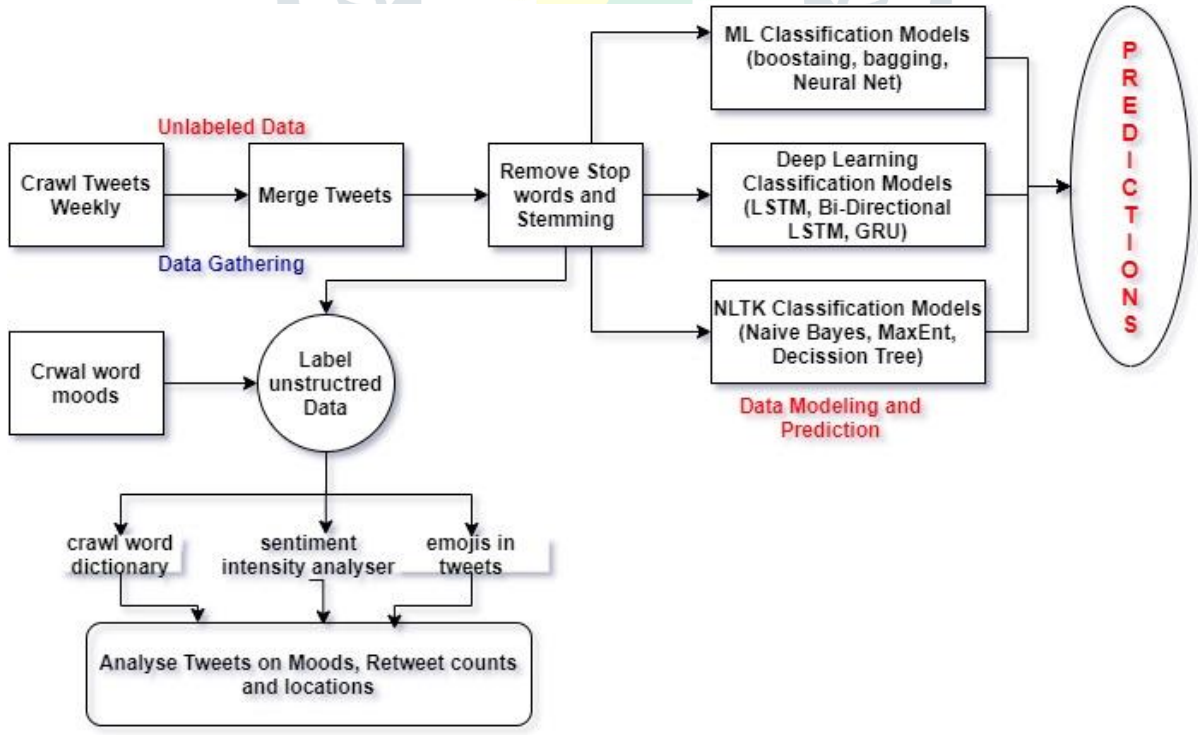

Fig. 1: Flow of the proposed system

3.1. Proposed algorithm for Sentiment Analysis of the Tweets

- Create a Developer Twitters Account
- Navigate to preferences using apps
- Get the client id and client secret keys by filling in the details.
- Create a class the following contents- Twitter username, password, client id, client secret
- Create an empty file called commented.txt
- Start authentication
- Read data from the init file
- Build a JSON object and pass data to twitter
- Authenticate the bot and stored the returned data in an object
- Start data gathering
- Get new comments from twitter
- If comment text matches a regex of our interesting topic
- Save comment id
- Add id to a commented text file via file handling
- Save the comment body
- Then go to the next comment
- Scheduling process started
- To work around Twitter API limits, set scraping schedule to a few minutes.
- Call fetch data
- Wait for a few seconds.
- Repeat
- Start Analysis
- Get total comments parsed
- Get total comments that are relevant
- Get percent of relevant comments
- Run analysis on the data collected

The above algorithm is useful in scraping the Twitter site and hence the data set created from it can be further used to read and analyze the emotions and the sentiments of the users as a community. This analysis can be used to bifurcate the topics on which the current socio-political situations is taken as a positive or negative impact by the masses.

3.2. Implementation of the sentiment analysis

- Generating the Twitter API key
- Get the authentication, you need to sign an app through your twitter account to get tweets through the Twitter API. For the same thing follow these steps:

- Go the apps.twitter.com: ‘Create New App’
- After filling application details callback URL can leave empty.
- After creating the app we will redirect to new tab to open “keys and access tokens” tab.
- Note the ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token Secret’

The username is the username used by the user to login and the password is the authentication key for the user. The consumer_key and the consumer_secret_key were generated by app.twiiter, which allows the bot to scrap its content for analysis.

- Run the Python Code to fetch the data in real-time and to calculate and plot the data for further analysis
- The data set taken can be then used to find the sentiments of users generally as well as for specific topics. It can also be used to find frequency distribution of words appearing in the corpus
- Plot the values to have a graphical overview of the analysis that is just the visualization part of the process.

- Use tools like NLP to find the positive response, negative or the neutral responses of the users and that analysis, in turn, depict the nature of the topic.

3.3. Implementation of the android system

In this App, there are two programming languages used. One is XML and another is JAVA. XML is used for layout description and the user interface which the user will experience, and the JAVA is used for the back-end purpose. That means all the functionalities that have been included in this application those functionalities run for this JAVA files.

- Certain built-in functions have been implemented.
- Widgets like toolbar, Relative Layout, Linear Layout, page viewer are used
- Background design for custom toolbar and list view is used which makes the appearance of the interface of the application better.
- Certain images in the drawable folder have been pasted and used in various objects.
- A raw folder is also set.
- For storing data shared preferences is used

The idea of this App came from my friend's discussion and the interface which we have chosen has been researched in various apps and been choose form random applications. Then we have chosen the application and built-in step by step with these following techniques:

- First, we started with the login page
- Then we build my sign up page.
- Then we focused on the homepage.
- Then we went into sub-modules.
- Then we created my all reaming pages
- Then we have made the video playing application this app is an android app that displays the Voting related information of two-party. This app will help you to support your party. The app is very simple to use and you will find very easy Interface for User

4. Conclusion

This study has made an attempt to answer whether twitter trends can predict election results. Our research setting in Indian context offers a complex political landscape. For example, in India vote share might not be a good proxy for number of seats own. In this paper, we have developed a twitter based sentiment analysis of the current election trading. We have developed an app, which display the trading of a politician of a particular political party based on twitter streaming data.

References

1. Jyoti Ramteke, Samarth Shah, Darshan Godhia, and Aadil Shaikh. "Election result prediction using Twitter sentiment analysis." In 2016 international conference on inventive computation technologies (ICICT), vol. 1, pp. 1-5. IEEE, 2016.
2. Adam Bermingham, and Alan Smeaton. "On using Twitter to monitor political sentiment and predict election results." In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011), pp. 2-10. 2011.
3. Rincy Jose, and Varghese S. Chooralil. "Prediction of election result by enhanced sentiment analysis on Twitter data using Word Sense Disambiguation." In 2015 International Conference on Control Communication & Computing India (ICCC), pp. 638-641. IEEE, 2015.
4. Rincy Jose, and Varghese S. Chooralil. "Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach." In 2016 international conference on data mining and advanced computing (SAPIENCE), pp. 64-67. IEEE, 2016.
5. Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams. "140 characters to victory?: Using Twitter to predict the UK 2015 General Election." Electoral Studies 41 (2016): 230-233.
6. Michael P. Cameron, Patrick Barrett, and Bob Stewardson. "Can social media predict election results? Evidence from New Zealand." Journal of Political Marketing 15, no. 4 (2016): 416-432.

7. Tariq Mahmood, Tasmiyah Iqbal, Farnaz Amin, Wajeeta Lohanna, and Atika Mustafa. "Mining Twitter big data to predict 2013 Pakistan election winner." In INMIC, pp. 49-54. IEEE, 2013.
8. Murphy Choy, Michelle Cheong, Ma Nang Laik, and Koo Ping Shung. "US presidential election 2012 prediction using census corrected Twitter model." arXiv preprint arXiv:1211.0938 (2012).