Twitter Sentiment Analysis

On

The Pulse of India's Elections

**Introduction**

Twitter has declared the 2014 Indian General Elections is a "Twitter Election". Indian Elections 2014 have completely redefined the way India has witnessed political battles fought between individuals and parties. Election saw 56 million election-related Tweets during 5 months, when the polls ended. Eventually it provides whose fortunes are rising and falling in the eyes of the web, and to learn who the web thinks will be the overall nominee.

Here, candidate of main 3 parties are considered for the analysis. Candidates are

* Narendra Modi- Bharatiya Janata Party(BJP)
* Rahul Gandhi- Indian National Congress(INC)
* Arvind Kejriwal- Aam Aadmi Party(AAP)

Twitter became the medium of choice for people to engage in and consume political content. Take any metric: original content generated, engagement by political leaders, user engagement with content, news breaks, influence on political discourse or capacity to set media agenda — it happened on Twitter.

Since, Indian Twitter user has had a greater impact on this mammoth election. This is why sentiment analysis has been considered here to analysis twitter related to election to understand pulic opinion on candidates

Here twitter data has been considered for campaign period to find out followings:

* Observe which Indian politician has had the highest positive/negative sentiment on Twitter.
* Also, find out "Tweets by Emotional Categories" and which word occured maximum times at a descending order. Atlast, comparing the frequencies of words among candidates.

Candidates always face challenge to get public opinion and feedback during campaign period. Since, twitter became the medium of choice for people to engage in and consume politcal content, candidates can use this social media to design more effective campaign and gain popularity which evetually helps them to win the election.

**Dataset**

Three dataset has been used here Twitter data has been captured from ........ to perform the sentiment analysis. Three datsSince, the focus on measure the performance of each cadidate duringHere, tweets for

India first time introduce social media to get public opinion during 4 months started from January 1 to April 30'15. campaign period.

* Here, consider Twitter data related captured on Election campaign onsider, twitter data during campaign period, starting

Here, India candi...... dataset from .... website has been considered here. Since, sentiment analysis has been considered here so focus Give the description of the dataset that you are using along with the individual attributes you will or will not use in your analysis.

Also mention the source of the dataset (where did you get it from). In case the data is curated and created by you please explain the details.

Descriptive statistics of the attributes and datasets can also be provided here.

**Problem Definition**

**Research Goal**

**Conclusion**

**Approach**

Our approach is to use different machine learning classifiers

and feature extractors. The machine learning classifiers are

Naive Bayes, Maximum Entropy (MaxEnt), and Support

Vector Machines (SVM). The feature extractors are unigrams,

bigrams, unigrams and bigrams, and unigrams with

part of speech tags. We build a framework that treats classi-

fiers and feature extractors as two distinct components. This framework allows us to easily try out different combinations

of classifiers and feature extractors.First, create a block diagram for the steps of your approach to clearly provide an overview. For example, if you first scrapped twitter, second applied NLP techniques to extract keywords, third labelled the tweets as positive and negative using a set of keywords, and fourth build a classifier, then you should create a box for each of the steps with arrows connecting one step to the next one. A sample block diagram is shown below.

Second, explain each of the steps in detail. What are you planning to do in each step or have already done. For example, in the above case you would create subheadings for each of the steps.

**Step 1: <Collect data>**

Collect tweets from twitter data related to 3 select candidates. Only content of the tweet and tweet publish timeframe have been considered here as the focus on public opinion towards candidates. Python programming language has been used to convert text file to csv file and reduce the size of the raw dataset before importing to R environment. Python is a dynamic, and multipurpose programming language. So, it is easy to handle large data in Python. With the help of Python 2 attributes with 10,000 data have selected from each of 3 datasets.

**Step 2: <Import dataset into R environment>**

After loading 3 datasets, generate 3000 random sample data of each candidate to get better result out of the dataset.

**Step 3: <Perform Sentiment Analysis on each candidate>**

Classify the text based on emotions categories and polarity return with higher count. Learning method has been used here to classify the text. Use ggplot function to visualize the output of each candidation. Eventually sentiment analysis speaks the overall contextual polarity of a document.

**Step 4: <Estimate Tweets sentiment>**

Get the score of negative and positive sentiment by performing “Sentiment scoring algorithm”. Here to get the result, “Breen’s Approach” has been considered. This method is called Lexicon which is nothing but word dictionary

**Step 5: <Text Mining>**

At this stage, introduce "Text Mining" machine learning tools to determine the characteristics of the message. First, transforming text by building a corpus, that is a collection of text documents. Pre-processing text includes removing stopwords, changing letters to lower case, removing punctuations and numbers. After that, stemming words t retrieves the root from, so that words look normal. TermDocumentMatrix function provides the frequency of terms in document and also importance of the words by using performing findAssocs function.

**Step 6: <Analyze the frequency of words of each candidate>**

Comparision cloud and comonality cloud function have been using here to find out the words of different frequencies and the words in common from the text referring to public opinion on candidates.

**Step 7: <Check the accuracy of the model>**

Finally, text has been classified as either negative and positive by using Naive Bayes algorithm. Task is to classify new cases as they arrive, i.e., decide to which class label they belong, based on the currently exiting objects.

**Results**

Explain your results here. Consider that you need to communicate your results to executives in an organization. For example:

* Insert tables and/or charts showing the results
* Write description of the tables and charts, such that they show the usefulness for an organization
* Identify the evaluation measures, such as accuracy, precision, recall, etc.

 ￼**￼**

* **Sentiment Analysis of Tweets by polarity shows what people think about candidates.**

 



Among 3 candidates Rahul and Modi has greater positive sentiment than Kejriwal on twitter. So, both Rahul and Modi are likely to get more votes than Kejriwal. On the other, Kejriwal has more neutral sentiment compare to other 2 candidates. It may depict that people may not sure about Kejriwal election menifesto.

**Finding Word Associations**

**We see that "ppg" has a high 0.65 correlation with "rpg", or rebounds per game. This makes sense as a tweet which contacts statistics about the points per game would also include other statistics like rebounds as well as "apg"- assists per game and "fg" field goals.**

* **word cloud to visualize the top occurring terms in the term document matrix for tweets on 3 candidates**

"Women" and "India" appear more often than any other word in most tweets related to candidate Rahul. Most likely he has been trying to talk about engaging women more for the development of the country.

Aam Aadmi Party(AAP) formally launched its political activity on 26 November 2012. And soon gained popularity in the courty. Word "Aamaadmiparty" appear most due to party's self-promotion and highlighting its party manifesto.people talk about party menifesto and their future direction.

Whereas, Modi's BJP party hold ralli one after another during campaign period. As a result, **Modi’s predominance appears more clearly by comparing the occurrence of his name including "ralli" term in election-related tweets versus other names and relevant search terms.**

**AamAllegations of corruption against politicians, bureaucrats, big business houses and the media dominate AAP leader Arvind Kejriwal's speeches**

**The chief minister of the western Gujarat state frequently mentions his government's "pro-development", "pro-youth" and "clean governance" policies.**

**Mr Gandhi is representing a party that has governed India in the last decade under the leadership of his mother Sonia Gandhi and Prime Minister Manmohan Singh.**

**Mr Singh's government has been marred by several high-profile corruption cases and analysts say this has severely affected the popularity of the Congress party.**

**A word cloud of Mr Gandhi's speeches and statements shows that he has been trying to highlight his own idea of "intra-party democracy" and governance.**

**He often talks about his party's "pro-poor" policies and praises programmes like the Right to Food Bill that promises cheap food for two-thirds of the population.**

**Allegations of corruption against politicians, bureaucrats, big business houses and the media dominate AAP leader Arvind Kejriwal's speeches.**

**He says India needs "independence" from corruption and pledges to punish all corrupt politicians and officials.**

**The former chief minister of Delhi also talks about a "nexus" between the national parties, industrialists and the media.**

**In his speeches, Mr Kejriwal taps on issues related to the "common man" such as water scarcity, high electricity and gas prices.**

**Mr Modi appears to be more direct about his vision for India and less discreet in verbal attacks on his political rivals.**

**.**

**He says India needs "independence" from corruption and pledges to punish all corrupt politicians and officials.**

**Modi’s predominance appears more clearly by comparing the occurrence of his name in election-related tweets versus other names and relevant search terms.urther elements to support the predominance of the Modi discourse on Twitter can be seen in the large following of the BJP candidate’s handle: nearly 4 million accounts follow Modi, versus 1.8 million for Kejriwal (as of May 12).**

* Visualize the words of different frequencies and the words in common from all the tweets text referring to public opinion on candidates



Commonality cloud shows that people talk about more on "India" and mass "People", when it comes to 3 candidates. It means mass people likely to expect that future leader would work for betterment of the common people of India. Whereas, comparison cloud show "gandhi" word appear most while talk about the Indian candidates. People likely to see Gandhi and nandra modi as there were the leading contender to take charge of the country.

* Apply Naïve Bayes Classification model: Naive Bayes is a simple model which works well on text.

Here, Naive Bayes model compare the prediction with reality. Result shows that, it is a very good model as the accuracy rate is 100%. Which means people predict negative and positive sentiment accurately for all the 3 candidates.

**Conclusions**

Analysis shows that people talk about positive things about Rahul and Modi. They are the 2 candidates people most likely wants to form a new government . Although, Kejriwal enter into politics with mass popularity, during campaign period, Kejriwal face tough challenge to establish their party menifesto, as a result he gets more tweets which classified by neutral.

Finally, the analysis shows that how politicians and citizens cooperate to create an e-democracy based on information dissemination and political awareness especially during elections. Further to add that more successful parties used Twitter to push timely updates on online and offline campaign activities, to their followers.