

Analysis of Indian Election using Twitter

Gayatri P. Wani

Department of Computer Engineering,
GES's R. H. Sapat College of Engineering,
Savitribai Phule Pune University, India

Nilesh V. Alone

Department of Computer Engineering,
GES's R. H. Sapat College of Engineering,
Savitribai Phule Pune University,

ABSTRACT

Modern civilization's one of the key player is social media. Social media is changing existing information behaviour by giving users access to real-time online Information channels without the constraints of time and space. This generates a huge unstructured data for data mining. This gives scientist an enormous scope for data analysis challenge. In this project system is going to study the general user tweets from the election point of view. Here the System will study the user view of Indian election. Based on the users tweets system analyses if there exist a pattern between the tweets and to analyse and draw meaningful inferences from the collection of these tweets collected over certain period; the proposed system identify the feasibility of development of a classification model to identify the political orientation of the twitter users based on the tweet content and other user based features. There are voting advice applications (VAAs) are online tools which are popularly used in deciding which party/candidate to vote for during an election in countries like Greece, Cyprus but still there in India there is no such application which focus on this . The proposed system plans to provide an online portal for recommendations and comparison of user's political opinions.

Keywords

Social media, Election campaign, Indian politics, Tweets, Twitter, Facebook and Hadoop

1. INTRODUCTION

The 21st century is the century of Knowledge. For every answer people first use Google. The theme line is connecting people using Internet. When we talk about Internet along with knowledge we have Social sites are specially used for connecting world. People especially youth spend its lot of time on sites like Facebook, Twitter etc... Every current topic due to social media is supported with general public's reviews and comment which makes a clear picture of that particular topic to world. The credit goes to popularity of social media.

According to Com. Score India is now at third position taking place of Japan becoming the world's third largest Internet user after China and United States [1]. As per Internet and Mobile Association of India (IAMAI) in June 2014 India is having more than 243 million internet users; which is 23 % increase in count compared to year 2013. The use of social media is significantly complex. The main challenges with social media are security, privacy, accuracy but still it has given new opportunities for sharing, collaboration, and engagement of users. It provides the platform for a politics. Social media platforms, like Twitter, Facebook and YouTube motivate people to get involved in all the political activities by sharing their view about party and candidates. Elections are having a great impact of social media. It specially used by both general public for getting updates on political events and

parties/Candidate to update their strategies and policies based on this social media's valuable feedback.

This paper is organized as follows. Section II reviews existing system used for election based analysis. In Section III, present proposed system. In Section IV, it provides the implementation details. Section V evaluates the dataset and results.

Problem Definition

With all the real time data collected over period of time, the system will analyze and draw meaningful inferences from the collection of tweets. Proposed system will analyze tweets data from many perspectives to make meaningful inferences. Trend analysis, sentiment analysis volume analysis are major parts of proposed system. In trend analysis, system will try to find trending discussions, parties, personalities throughout the period of time. From literature, K-means is more suitable algorithm for clustering of tweet data and to find trends. Volume analysis of tweets will give idea of popularity of particular topic or person over a period of time. Volume analysis with respect to geo-location and date will help to make certain conclusions. Sentiment analysis of tweets will help draw conclusion for political orientation of overall users respective to political parties, topics.

1) Input

- A tweeter tweets related to election
- A hash-tag List

2) Output

- Volume Analysis
- Trend Analysis
- Sentiment Analysis

2. LITERATURE REVIEW

Social media has become this significant part of modern civilization. Social media is changing existing information behaviour by giving users access to real-time online information channels without the constraints of time and space. This gives scientist an enormous scope for data analysis challenge. Most previous studies have adopted broad-brush approaches that typically result in limited analysis possibilities.

Author Min Song ,MeenChul Kim and Yoo Kyung Jeong [3][6] has studied a twitter dataset for 2012 Korean election by collecting real time tweets. Topics extracted from tweets and related real time events relation was identified and they were traced chronologically using term co-occurrence retrieval technique.

For India 2014 general election; User's orientation towards parties and candidates was studied by Author Abhishekbhola[7] using Tweeter. A dataset consisting of 17.60 million of tweets was analyzed for identifying user, candidate, party popularity based on peak of time, topic or location. A sentiment analysis was performed using classification algorithms

Voting advice applications (VAAs) [8] are online tools designed for election in countries like Greece. It is becoming increasingly popular and they are helping users in deciding which party/candidate to vote. It is designed based on the concept called community-based recommendations system. It provides the comparison of users' political opinions, and becomes a channel for user communication. This system proposed various approaches for community-based vote recommendation. The approaches were evaluated on five real VAA datasets in terms of prediction accuracy. Using Data from Facebook and Twitter; Lars Kaczmirek and his team from GESIS [9] gathered various aspects of the communication structures. They compared data gathered from social media with local survey and added new insights to social media by providing how social media can be used during elections. Based on this studies ;German Longitudinal Election Study (GLES), a long term research project is designed ; that examines the German federal elections for the years 2009, 2013, and 2017. The main aim of this project is to track the German electoral process over an extended period of time; by collecting Twitter and Facebook data about the German Bundestag elections.

3. PROPOSED WORK

The main goal and challenge of the system is analysing twitter data for Indian election to see the impact of tweeter on Indian or particular state election. The proposed algorithm is consists of three main steps.

Our proposed system is analysing system which is based on the mechanism that analyses User Tweets using Hashtags and Keywords. The proposed system collects tweets using this HashTags which are nothing but the popular personalities/Parties. General public orientation toward these parties can be studied using the tweets the people have posted on the Tweeter. Tweeter is generally lauded by academicians, journalists and Politicians; for its potential political value. Many politicians make use of this micro blogging site to express themselves in the limit of 140 characters. These tweets can be categorized on various policies such as geo-location analysis to analyses the peoples view for that particular area which might help parties to design their winning strategy. The proposed system mainly focus on collection of tweets to make volume analysis to and out the popular days of election ; A trend analysis to and a popular or trending party/candidate and a sentiment analysis to actually bifurcate the positive and negative tweets for the party/candidate so that making trend analysis on this tweets can help this party/candidate to act accordingly to improve their reputation at the same time it might help user to actually make a clear opinion about any party/candidate. This will be conducted in 3 phases. To brief about it the phase one is connecting with tweeter and downloading the tweets. The second phase deals with loading these tweets on HDFS for further analysis and the third phase is the actual analysis and they are volume analysis, Trend Analysis and Sentiment analysis.

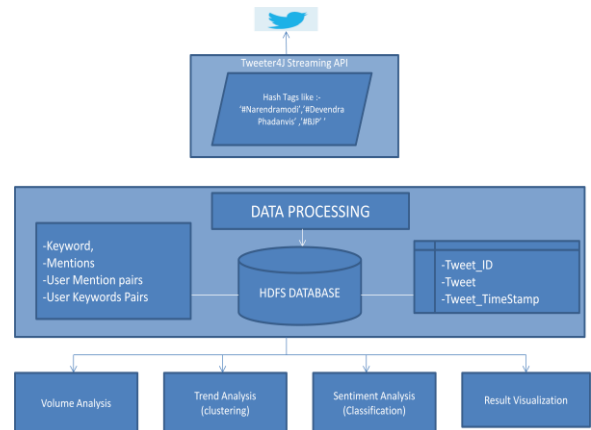


Figure 1 : Tweet Analysis

MATHEMATICAL MODEL

$$P = \{I, O, F, U\}$$

Where
I: Input
O: Output
F: Functions
U: User

Where
 $I = \{U, UT, HL\}$
Where
U = User having a tweeter developer profile
UT = User Tweets which will be downloaded using authentication key
HL = Hashtags list provided as input for downloading election related tweets

$O = \{UP, PT, TA, VA, SA\}$
Where below are the output generated from system processing
UP= Retrieved User Profile details
PT= Processed Tweets by removing unwanted keywords stop words etc..
TA= Trend analysis on processed tweets will provide trending topic/Politicians
VA= Volume Analysis will provide the analysis based on date wise or location wise volume analysis
SA= Sentiment Analysis of tweets by categorizing them in positive, negative or neutral sentiments

$U = \{SV, TU, A\}$
Where
SU = System Visitor
TU = Tweeter User whose tweets are used in system as input
A= Administrator

$$F = \{F1, F2, F3, F4, F5\}$$

Where
1) Function F1 : This function download the tweets through secure authentication using OAuth
2) Function F2 : This function process the downloaded tweets for stop word removal
3) Function F3 : This function performs volume analysis using Map Reduce Framework

- 4) Function F4 : This function performs trend analysis using clustering of user, parties and related details.
5) Function F5 : This function performs sentiment analysis by classifying tweets into positive ,negative and nutral types

ALGORITHMS

The algorithm for analysis is outlined as follows.

Algorithm 1: Porter Stemmer Algorithm

Input

- Let T be the set of downloaded tweets.

Output

- Processed tweets with all unwanted word, space and special character removal.

Algorithm 2: K-Means Clustering Algorithm

Initially, topic wise tweets are set as a canter of clusters. For every iteration, distance between center and sample is checked and sample is added to respective cluster. Distance between center and sample is measured using TF-IDF .Clusters are updated at every iteration. Based in TF-IDF weightage

Input

- Let X be Set of data points where $X = x_1, x_2, x_3, \dots, x_n$
- Let V be Set of centre's where $V = v_1, v_2, \dots, v_c$

Output

- Clusters formed

Steps

- Choose numbers of clusters to be determined
 - For initial centres of the clusters; randomly choose the centroid
 - Repeat
 - Assign each object to their closest cluster center using Levenshtein distance.
 - Compute new cluster center by calculating mean points.
 - Until
 - No more changes in the cluster's center
- OR
- object's clusters is not changed further

Algorithm 3: Naive bays Classifier

Political orientation of users towards party, topics can be analyzed from tweets. Map Reduce version of nave bayes algorithm will be implemented to classify tweets into positive, negative and neutral classes.

Steps

- Create a data for the classifier
 - Create a list of positive tweets
 - Create a list of negative tweets
- convert this two list in to single list with two parts word array for each tweet and its type
- Design a Classifier
 - Extract the word feature list from the list with its frequency count
 - Using this words list create feature extractor which contains the words which will matched with a dictionary created by us indicating what words are contained in the input passed
- Training the Classifier using training dataset
 - Generate Lable_Prod List which contains positive and Negative Label.

- Generate Feature_Prod List which contains the featured words.
- Calculate the probability for the positive and Negative Label.
- Compare this probability to identify the tweet category as positive, negative or neutral.

4. RESULTS

We have implemented this system based on java and Hadoop platform for Map Reduce framework

Dataset

- The dataset for the system is the election related downloaded tweets.
- For Sentiment analysis positive and negative thesaurus will be provided.

Results

As discussed above the system will generate volume analysis, trend analysis and sentiment analysis. The tweets downloaded from twitter database are processed with help of porter stemmer algorithm and users define functions. These filtered tweets will be used as input for various analysis modules generating the trend, volume or sentiment analysis

Centralized system: Experiment performed on single machine with core i3 processor 2.0 GHz x 2 and 4GB of main memory. The operating system is Ubuntu 14.04 and all algorithms are implemented in JDK 1.7.

The data was download from twitter and the further analysis was conducted on it as shown in results

Table 1. Hash Tag Wise Tweet

| Hash Tag | Count | Hash Tag | Count |
|----------------------|-------|-----------------------|-------|
| #ShivSena | 2477 | #MaharashtraElection: | 149 |
| #Shivsena | 1035 | #NCP | 142 |
| #BJP | 840 | #Yuvasena | 118 |
| #Mission150 | 568 | #Modi | 114 |
| #UThaMaharashtra | 494 | #Proud2Follow | 112 |
| #MaharashtraElection | 459 | #Congress | 111 |
| #shivsena | 426 | #ElectionResult | 105 |
| #Maharashtra | 330 | #FinalElectionResult | 104 |
| #ShivSena: | 283 | #UT4CM | 104 |
| #UTha4CM | 214 | #MaharashtraElections | 103 |

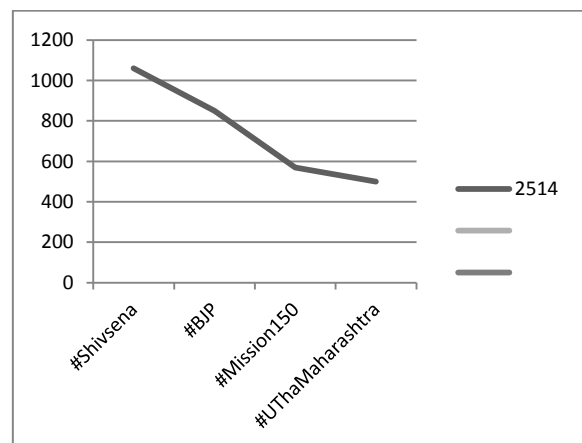


Fig.2 Hashtag Wise TweetCount

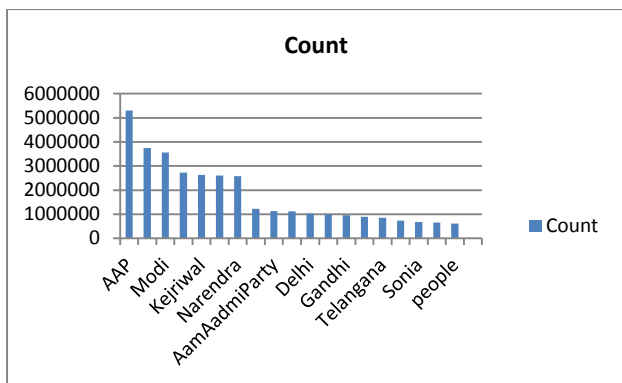


Fig.3 Trending Topic of Elections

As discussed earlier the system also performs sentiment analysis and trend analysis.

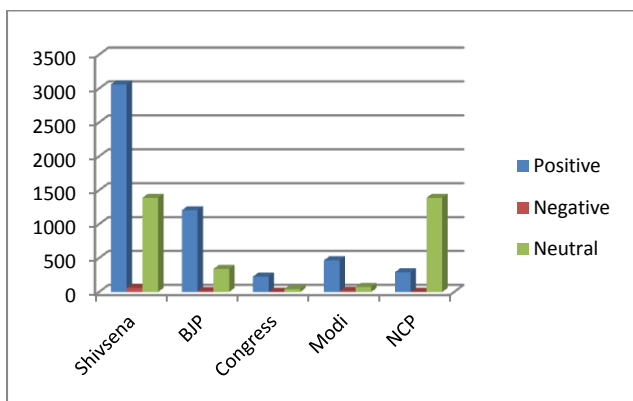


Fig.4 Sentiment analysis of tweet dataset during Maharashtra state assembly election

As the system also studies orientation of the users towards parties which is calculated using K-Means can be described as below:-

Table 2. Clustering of the twitter data to understand user's orientation

| key: cluster0 | value: 951 | len: 8773; |
|-------------------------|--------------|------------|
| 0:0.0317439599588505; | | |
| 1:0.007785511650683655; | | |
| 2:0.00911601437145913; | | |
| 3:0.008685158879466285; | | |
| 4:0.00600840075081049; | | |
| 5:0.010097762163994141; | | |
| key: cluster1; | value: 1311; | len: 8773; |
| 0:0.06794883023775404; | | |
| 1:0.06200068035778367; | | |
| 2:0.026126748577468096; | | |
| 3:0.035674424429240244; | | |
| 4:0.01583430022806395; | | |
| 5:0.013401314555699321; | | |
| 6:0.012500457491432028; | | |

5. CONCLUSION

With the increased use of social media the current paper focused mainly on use of social media as a tool for election campaign. India which is known to be one of the wired countries in the world with having more than 65 % of its youth below age-group of 35; Social media plays vital role in the life of this young youth. The proposed systems will try to analyse the Maharashtra state assembly election; to study the impact of social media on Maharashtra politics system found people can express their views in 140 characters more efficiently and openly.

6. FUTURE SCOPE

Further work can be carried on to find verifying the changed sentiments of the user before and after election. The study can be enhanced to actually classify the gender of the user and to match it with actual demographics it can include the scope for verifying whether the user is human or a bot. The study can also include a machine learning approach to train a system to automatically classify the tweets and do a sentiment analysis of the tweets.

7. ACKNOWLEDGEMENT

I would like to express my sentiments of gratitude to all who rendered their valuable guidance for this work. I would like to thank Dr. P. C. Kulkarni, Principal, GES's. R. H. Sapat College of Engineering Nashik, for providing me strong platform to develop my skills and capabilities.

I am also thankful to Prof. N. V. Alone, Head of Department, Computer Engineering and my guide Prof. A. S. Vaidyaour PG co-ordinator for helping and guiding me with the topic and also providing me with adequate facilities, ways and means by which I was able to complete this paper.

8. REFERENCES

- [1] India is now world's third largest Internet user after U.S., China. <http://www.thehindu.com/sci-tech/technology/internet/india-is-now-worlds-third-largest-internet-user-after-us-china/article5053115.ece>
- [2] Use of New Media in Election Campaigning (Lok Sabha Elections 2014) https://www.academia.edu/7486078/Use_of_New_Media_in_Election_Campaigning_Lok_Sabha_Elections_2014
- [3] Gayatri Wani , Nilesh Alone, "A Survey on Impact of Social Media on Election System" <http://www.ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506100.pdf>
- [4] <http://www.bbc.com/news/world-asia-india-762391?OCID=fbbcbindia>
- [5] Social media for political campaign in India. <http://www.slideshare.net/RaviTondak/social-media-for-political-campaign>
- [6] Min Song MeenChulKim ;Yoo Kyung Jeong, Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter 1541-1672/14/ Published by the IEEE Computer Society.
- [7] AbhishekBhola "Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General Elections2014" arXiv:1406.5059 [cs.SI]
- [8] IoannisKatakis, Nicolas Tsapatsoulis, Fernando Mendez, VasilikiTriga, and ConstantinosDjouvas "Social Voting Advice Applications - Denitions, Challenges, Datasets

- and Evaluation" IEEE TRANSACTION CYBERNETICS ,VOL. 44 No. 7
- [9] Lars Kaczmirek, Philipp Mayr, Ravi Vatrapu, Arnim Bleier, Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter
- [10] Aibek Makazhanov, Davood Rafiei : Predicting Political Preference of Twitter Users, ASONAM '13 Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining Pages 298-305
- [11] Use and Rise of Social media as Election Campaign medium in India, Narasimhamurthy N, (IJIMS), 2014, Vol 1, No.8, 202-209
- [12] Indian general election, 2014. http://en.wikipedia.org/wiki/Indian_general_election,_2014 .
- [13] Population of India 2015. <http://www.indiaonlinepages.com/population/india-current-population.html>
- [14] Twitter Data Analytics. <http://tweetracker.fulton.asu.edu/tda/TwitterDataAnalytics.pdf>