

# Indian Election Using Twitter

Gayatri P. Wani and Nilesh V. Alone

**Abstract** Today's world is ruled by Social Media. It is giving user, an access to real information without space and time constraints. Along with this information this gives a great opportunity to users to share their views freely on open platform. This current happening with people's opinion generates huge database. This generated database can be studied by scientist to analyze the generated data. The proposed system studies user's opinion on Indian election. System studies user opinion to understand if there exist any similarities in user's views on election. The system tries to identify the feasibility of classification model. The system studies the political orientation of Twitter users by analyzing the tweet content and other user's Twitter features. There are systems like voting advice applications (VAAs), online tools which are popularly used during an election in countries like Greece, Cyprus for getting real opinion on parties/Candidates during election but still in India there is no such application which can guide user. The proposed system plans to develop an application for recommending and comparing user's political opinions.

**Keywords** Social media · Election campaign · Indian politics · Tweets · Twitter · Facebook and Hadoop

---

G.P. Wani (✉) · N.V. Alone

Department of Computer Engineering, Gokhale Education Society's  
R. H. Sapat College of Engineering, Nashik, Maharashtra, India  
e-mail: gayatriwani@gmail.com

N.V. Alone

e-mail: nilesh.alone@gmail.com

G.P. Wani · N.V. Alone

Management Studies and Research, P T A Kulkarni Vidyanagar, Nashik  
Maharashtra, India

G.P. Wani · N.V. Alone

Savitribai Phule Pune University, Pune, Maharashtra, India

© Springer Science+Business Media Singapore 2017

S.C. Satapathy et al. (eds.), *Proceedings of the International Conference on Data Engineering and Communication Technology*, Advances in Intelligent Systems and Computing 468, DOI 10.1007/978-981-10-1675-2\_27

251

# 1 Introduction

We call today's world a connected world. The best way of connectivity is Internet. For every new term or question; we first use Google search to find out relevant information. India is at third position in using Internet [1]. In India people, especially, youth spend lot of their time on sites like Facebook, Twitter, etc. Every current topic is getting discussed on social media generating huge public's reviews which can make a clear picture of people's views on that particular topic to world. The main challenges with social media are privacy and especially accuracy or reality of the opinion of the user [2]; but still it has given new opportunities for collaboration, sharing, and engagement of users. It provides the platform to various subjects and one of them is politics. Social media 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 [3]. It is 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 [4]. This paper is organized as follows. Section 2 reviews existing system used for election-based analysis. In Sect. 3, we present our proposed system. In Sect. 4, it provides the implementation details. Section 5 evaluates the dataset and results.

## 1.1 Problem Statement

Using the real-time data (tweets) collected over election period, the system analyzes the collection of tweets to draw meaningful inferences. The given system analyzes the tweets with the perspective of volume analysis, trend analysis, and sentiment analysis. Using volume analysis, system studies the popularity of candidate/Party, terms, active users during election period. In sentiment analysis, the system studies the sentiments of the user's tweets in terms of positive, negative, and neutral tweets for identifying user's orientation toward political party/candidate. In trend analysis, the system studies the popularity by area and tag wise. The input for the system is set of hashtags which download relevant tweet set which becomes the input for further analysis.

## 2 Related Work

Social media is a very important part of today's modern civilization. The social media influence is generating huge information which challenges analyst to make use of this generated information for various purposes starting from health to market needs, finance, global cliental, etc. These users' opinions are also largely

generated when it comes to any political activity such as election; which is very crucial for any country. Most of the countries like Greece, United States are studied by analyst from social media like Twitter and Facebook point of view. They have adopted broad-brush technologies and limited analysis possibilities.

Author Song et al. [5, 6] have studied a election relevant to Twitter dataset for 2012 Korean election by storing real-time tweets by using tweet4j api of Twitter. They developed a system to analyze 2012 Korean presidential election, by carrying out Dirichlet-multinomial regression (DMR) topic modeling, network analysis using mention-based Twitter network, and term co-occurrence analysis relevant to presidential issues embedded in Twitter data.

Indian general election 2014; was studied by Author Bhola [7] using Twitter. The system studies user's orientation toward parties and candidates. The dataset used consists of 17.60 million of tweets; which were analyzed for popular user, candidate, and party; based on location, topic, and peak of time. A sentiment analysis was performed with the help of human annotators using classification algorithms.

Countries like Greece use online voting advice application called voting advice applications (VAAs) [8]. It is quite popular and users are getting benefit from it by choosing which party/candidate to vote. VAAS proposed variety of approaches for community-based recommendation systems which are evaluated on five different datasets. First it was evaluated on VAA dataset for identifying prediction accuracy of the system.

Kaczmirek and his team from GESIS [9] have compared Facebook and Twitter dataset for German election with a local survey for giving new insight on importance of social media during election period. This research is part of research project called German Longitudinal Election Study (GLES). This research will examine the German federal elections for the years 2009, 2013, and 2017.

### 3 Proposed System

The main goals of the system are analyzing Indian election Twitter data for studying the impact of twitter on Indian election and Maharashtra state. The system uses Twitter streaming and rest API for tweet collection. Our system also uses data collected for Indian election by Mr. Shreyansh Bhatt from Knoesisn University for Twitris project [10]. For collecting tweets for Maharashtra, hashtags like BJP, Namo, UthaMaharashtra, etc. are used as filters. Twitter gives platform for discussion which is quite popularly used by academicians, Journalist, and Politicians. Twitter has potential political value. Many politicians express their agendas and strategies with common man using Twitter. These tweets with geolocation and date, time, follow, and follower along with retweet facility provide quite popularity of party. This analysis can be used by politicians for defining their winning strategies. The system mainly covers volume analysis to identify popular days, topics, parties,

and candidates along with active users during election period. Sentiment analysis is performed in system in order to understand positive and negative popularity of parties/Candidates. Trend Analysis covers geolocation and word-based clustering to analyze popular parties and candidates. The system is developed in three phases starting from tweet download followed by the analysis of collected data and lastly visualizes the results discussed in Fig. 1. The tweets are preprocessed before actually being used in the system.

The algorithms used in the system are discussed below

**Algorithm 1: Volume Analysis**

Input: - Preprocessed tweets Output: - Top hashtags, Top active users, Top trends  
Begin

1. Transfer all tweets from local file system to HDFS.
2. In map phase, all tweets are split into words with whitespace as a separator and mapped to one.
3. In reduce, all similar words are counted and written to HDFS.
4. Transfer all output tweets to local file system.
5. Sort all tweets as per the count and filter all with hashtags, users, and trends.

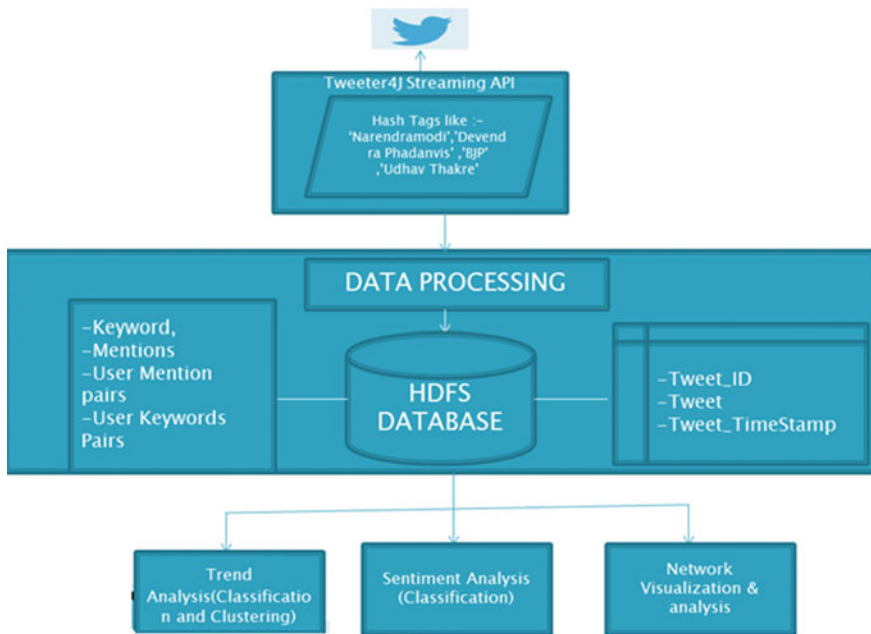


Fig. 1 System architecture

6. Put all top hashtags, users, and trends.
7. Visualize results for better perspective.

End

### **Algorithm 2: K-Means Clustering Algorithm**

Input: Preprocessed tweets Output: Clusters formed

#### **K-Means Clustering Algorithm Using Geolocation**

Input: Set of geolocation points where  $X = \{X_1, X_2 \dots X_n\}$ , Output: Clusters Begin

1. Provide the number of tweets to be used for clustering [11–14].
2. Provide numbers of clusters to be determined
3. Transfer vectors and clusters from local file system on HDFS
4. For initial centroid of the clusters; randomly choose the geolocation value
5. Repeat
  - a. In map phase, calculate closeness in terms of longitude and latitude of geolocation of each vector with every cluster centroid.
  - b. In reduce phase, assign vector to cluster which is the most close to centroid.
6. Until
  - a. No more changes in the clusters center OR
  - b. Object's clusters are not changed further
7. Classify each tweet from cluster on provided party type.

End

#### **K-Means Clustering Algorithm Using TF-IDF**

Input: Set of TF-IDF of tweets where  $X = \{X_1, X_2 \dots X_n\}$  Output: Clusters

Begin

1. Find out unique words and TF-IDF of each unique [11–14].
2. Create set of tweet vectors using dictionary of unique words.
3. Provide the numbers of clusters to be determined.
4. For initial centroid of the clusters, randomly choose/provide the TF-IDF
5. Transfer vectors and clusters from local file system on HDFS
6. Repeat
  - a. In map phase, calculate closeness in terms of TF-IDF of each vector with every cluster centroid.
  - b. In reduce phase, assign vector to cluster which is the most close to centroid.
7. Until
  - a. No more changes in the clusters center OR
  - b. Object's clusters are not changed further

End

**Algorithm 3: Naive Bayes Classifier**

Political orientation of users toward party, topics can be analyzed from tweets. Map Reduce version of Naive Bayes algorithm is implemented to classify tweets into positive, negative, and neutral classes.

Input: Preprocessed tweets, Positive and Negative dictionary Output: Classified Tweets.

Begin

- 1: Create a data for the classifier
  - 1.1: Create a list of positive words
  - 1.2: Create a list of negative words
  - 1.3: Provide a tweet file which needs to be analyzed
- 2: Design a classifier
  - 2.1: Extract the word feature list from the list with its frequency count
  - 2.2: Using this words list, create feature extractor which contains the words which will match with a dictionary created by us indicating what words are contained in the input passed
- 3: Training the classifier using training dataset
  - 3.1: Generate Label Positive Probability which contains total number of positive words in input file
  - 3.2: Generate Label Negative Probability which contains total number of Negative words in input file
- 4: Calculate the probability score for the positive and negative word for individual Tweet
  - 4.1: Calculate positive score by total number of positive words in the tweet divided by Positive Probability
  - 4.2: Calculate negative score by total number of negative words in the tweet divided by Negative Probability
- 5: Compare this probability to identify the tweet category as positive, negative, or neutral

## 4 Results

The system uses two datasets for election-related analysis. The first dataset is about India's 16th general election conducted in 2014 and the other database is for Maharashtra state assembly election for year 2014.

### 4.1 Volume Analysis

The system using the importance of hashtag analyzes the election for finding important messages or themes of communication during election period. Similarly trending users and trending topics are identified by the system.

Figures 2 and 3 show top hashtags of the Maharashtra and Indian Election.

### 4.2 Sentiment Analysis

Sentiment analysis was performed to understand user’s orientation toward any party or candidate. During Maharashtra election, it was observed that people were having more orientation toward Shivsena party and Mr. Modi. Whereas for Indian election, the results were quite in favor of BJP due to the popularity of Mr. Narendra Modi.

The pie chart in Fig. 4 below describes the popularity of BJP, Shivsena, Congress, Mr. Arvind Kejriwal, and Mr. Narendra Modi.

### 4.3 Trend Analysis

The trend analysis provides the popular trends like popular personality or topics and geolocation-wise popular parties.

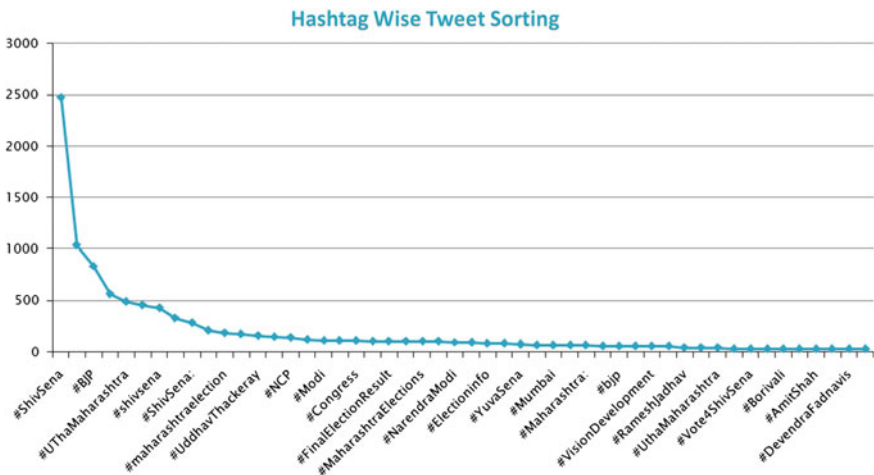


Fig. 2 Top Hashtag for Maharashtra State Assembly Election-2014







Fig. 5 Word cloud of cluster0 for Maharashtra State Assembly Election-2014



Fig. 6 Party-wise region-wise geolocation-based tweet Clustering (Color figure online)

### 4.3.2 Trend Analysis Using Geolocation

The trend analysis of the system is performed using K-Means clustering. The K-Means clustering algorithm works on the principal of distance-based clustering of the given point. For India’s 16th Loksabha election the clustering is performed using geolocation coordinates. These tweets were classified on party basis for identifying which plotted geolocation belongs to which party.

## 5 Performance Evaluation

To check the performance of our system, we focused on India's 16th Loksabha Election-2014 as it consists of 2,71,49,422 tweets and its relevant details. Our system is developed on Hadoop. We analyze the performance of the system using single node and multi-node Configuration. The single node consists of 8 GB RAM and I7 Processor with 500 GB hard disc. The multi-node consists of one master along with three slave nodes having configuration of 4 GB RAM and i3 processor. We have evaluated performance of our system based on the time taken for processing of volume analysis and sentiment analysis. Speedup acquired for volume analysis is around 45 %. Similarly speedup is 25 % for Sentiment analysis.

## 6 Conclusion and Future Scope

The experimental results show that the algorithm used in designing system is implemented successfully and also capable of eliminating the drawback of existing system. As per result of 16th Loksabha election it is verified that the predictions computed using our system are correct. The speedup obtained by using distributed system for volume analysis is 45 % and for sentiment analysis is 25 %. As per our analysis of Twitter data, maximum people were pro-BJP/Modi. In Loksabha election, it could be observed that people from maximum states of India elected BJP to form the government of India. In Maharashtra state, assembly election, we analyzed that Shivsena, Mr. Udhav Thakrey, and B.J.P. were quite popular even though this election was just after the Loksabha election. People were having mixed feeling for both the parties.

## References

1. <http://www.thehindu.com/sci-tech/technology/internet/india-is-now-worlds-third-largest-internet-user-after-uschina/article5053115.ece>
2. [https://www.academia.edu/7486078/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. <http://www.bbc.com/news/world-asia-india762391?OCID=fbbbcindia>
4. <http://www.slideshare.net/RaviTondak/social-media-for-political-campaign>
5. <http://www.ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506100.pdf>
6. Min Song, Meen Chul Kim; Yoo Kyung Jeong, Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter 1541-1672/14/ Published by the IEEE Computer Society
7. Abhishek Bhola "Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General Elections 2014" [arXiv:1406.5059](https://arxiv.org/abs/1406.5059) [cs.SI]
8. Ioannis Katakis, Nicolas Tsapatsoulis, Fernando Mendez, Vasiliki Triga, and Constantinos Djouvas "Social Voting Advice Applications - Definitions, Challenges, Datasets and Evaluation" IEEE TRANSACTION CYBERNETICS, VOL. 44 No. 7

9. Lars Kaczmirek, Philipp Mayr, Ravi Vatrappu, Arnim Bleier. Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter
10. <http://twitris.knoesis.org/>
11. 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
12. Use and Rise of Social media as Election Campaign medium in India, Narasimhamurthy N, (IJIMS), 2014, Vol 1, No. 8, 202–209
13. [http://en.wikipedia.org/wiki/Indian\\_general\\_election,\\_2014](http://en.wikipedia.org/wiki/Indian_general_election,_2014)
14. <http://www.indiaonlinepages.com/population/india-current-population.html>