

Prediction of Election by Twitter



Kusum, Supriya Panda

Abstract: Nowadays social media like Twitter and Facebook etc. is one of the key players. Twitters are micro blogging sites by which users sent their opinions and views in brief. The information generated by one user can be seen by everyone. Therefore to analyze twitter sentiment can be a crucial task. For this task, we have used various approaches like novel based approach and machine learning and many other rules like context awareness are used for the detection of public opinion and prediction of results. We are studying the user tweets during elections. Meaningful tweets are collected on a definite period. The feasibility of the developed classification model is identified by our proposed work to identify the political orientation on the tweets and other user-based features. The technique for the collection of tweets in time has played an important role. When the outcome of applied technique competes with survey agencies result was published before elections result.

Index Terms—tweets, opinion mining, emotion detection, sentiment analysis, social media, classification.

I. INTRODUCTION

Nowadays almost every person expresses their sentiments on online social sites like blogs, websites, microblogs etc. Twitter is a popular microblogging site. People can give their ideas or views about these sites. Tweets have very short length therefore they are mainly used in analysis of sentiments. After analyzing, the short text messages large information is obtained about the thoughts of people [21]. To extract the information for natural languages, sentiment analysis is an effective technique. For the general election in India, 2014 user's direction belonging to parties and candidates was studied [23]. For this twitter was used. For extracting meaningful and important opinions NLP technique is used. Tools that work on mines and customs are not used in micro-posts due to lack of context [8]. This information is meaningful in many subjects where context is large. Micro-blogging services like Twitter is a very popular online service. Its users are several million because users post their views, opinion and every important thing by the short messages these short length messages have the same feature which is common to all. For example:

User Name: It is included to direct their message. A DeFacto standard includes @ it is used before the user name that is @ IPL 2015.

HashTag: Users can use '#hash tags' like expression. This expression is used to express keywords which tell the content of Tweet.

RT: It is known as re-Tweeting. When the Tweet is very interesting users republish that Tweet "RT" is used.

Revised Manuscript Received on February 05, 2020.

* Correspondence Author

Kusum*, Department of CSE, Manav Rachna International Institute of Research and Studies, Faridabad, India. E-mail: kusumprerak@gmail.com

Dr. Supriya Panda, Department of CSE, Manav Rachna International Institute of Research and Studies, Faridabad, India. E-mail: supriya.fet@mriu.edu.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Tweets containing opinion and views are very important for making decisions and are used as a tool in government work, corporation, schools and organizations.

Sentiment Analysis deals with the study of analyzing people's sentiments, ideas, feelings, and mood of people, etc. In brief sentiment analysis is opinion mining. It has three different levels:

Document Level

Sentence Level

Aspect Level

In Document Level complete document is used to express Sentiments. In sentence level, only involved sentence is used for the extraction of sentiments. In aspect level, fine-grained analysis is done by this method. It directly captures the ideas..

Sentiment Classifier: Classification is done by the sentiment classifier. It classifies sentence into subjective and Objective

class. Sentiment analysis is done only on

That sentences which gives personal facts

subjective sentences because the objective sentence does

not hold subjective information. Thus it reduces the

accuracy of the result.

II. RELATED WORK

The exponential growth of twitter has drawn the attention of Researches in various areas. For eg. Product marketing election campaign, etc. Twitter is used extensively for political deliberations. According to Tumasajan et al., [15] the only number of messages by party reflected election

results. Content analysis of 100000 messages containing a reference was conducted.

Our main goal and challenges are to analyze Twitter data for the Election system in India. The information available on the website are mainly of two types: facts and opinions. Arguments by social media sites were one of the causes of the presidential elections in the USA in 2008 won by Obama. This shows the success of micro-blogs such as Twitter [1]. Twitter-based opinion mining platform was

developed by [4]. The variables used for data collection are important keywords [16,17]. The sentiment value is addition of weights which is extracted from text. For the sentiment analysis two methods: Bag of words model and natural language processing are used. Bag of words means weight (textual record consisting of a token) referred to as sentiment in the context. Natural Language processing gives path to know the context, string of words and sentiments. For keyword selection the topic modeling technique used by Blei et al., [18] provides an explicit representation. Efficient techniques based on variation methods are applied. Popular Political Party Leaders give impression on voters during evaluation of elections [3]. He said "rallies and meetings increased as the political activities increased and election come closer.



Prediction of Election by Twitter

For this verification elections related topics are extracted by LDA based approach. Tumasjan et al., [15] mine opinions from incrementally generated queries a novel framework is used to consider the arguments.

To identify the opinions is a challenging task. The research started with a classification problem. Various machine learning approaches like Naïve Bayes, SVM, and Maximum Entropy model is used.SVM supports the best result according to [19]. To analyze Tweets automatically a new system architecture was proposed. This system was combined with manually annotated data. Opinion containing tweets was extracted and filtration was done on non-opinion tweets. Naive Bayes' Classifier with text classifier is used at the last step when the determination of sentiments is done as positive or negative.Naive Bayes use a probabilistic algorithm.It takes theory of probability to calculate probability of tag in a text.Features are taken from text that are the pieces of information.Some of the more techniques are used:

Stop words are removed.For example a,else,over,either etc.

Words are lemmetized by grouping together..For example elections elected were grouped and counted as the same word.

Instead of counting single word sequence of words was counted.

Words that appear in text penalizing frequently was used.

Our proposed work is based on a technique which analyses tweets by comparison. There are three major parts of the system:

- Analysis based on trend
- Analysis based on sentiment
- Analysis based on volume

In trend analysis trending discussion of various parties, personalities are found in the period .K- means Algorithm is used to detect the trend data.K is meaning of fixed number of clusters in a dataset.Through this points are collected of similar things.Random centre point is selected used as beginning point.Means refers to averaging. The reputation of the Party / Candidate helps the user to make a clear opinion.

Sentiment analysis helps to conclude. The combination of sentiment analysis using supervised learning and volume-based measures are used to predict results. For monitoring public sentiments no observations are made during the election campaign.TF-IDF method is applied [7].

To learn from tweets every week system is designed..Result is consolidated by eliminating duplicate tweets and number of retweets are impacted.

Volume analysis gives idea of popularity of topics and persons.

Generally all parties and candidates utilized social sites in order to encourage voters. It is a way to help people of obtaining information from the site. The social sites are taken as communication media sharing information with all citizens [24].

III. PROPOSED WORK

Our proposed work is on topic extraction related with real time events. 2019 legislative assembly elections of states in India.Twitter analyzation was done on a data set containing(1000) of tweets.These tools are designed for elections based on Community based recommendation system.It works as a channel for user communication system.

In this analyzing system, user Tweets are analyzed by the use of hashtags and keywords. Hashtag tweets are popular parties or popular personalities. The public direction is studied in tweets that people post on twitter. Content information from twitter is collected by crawling tweets during election time.

For scoring sentiments, different metrics are used by specific keywords. Grosse et al.,[5] performed a time series analysis for candidates and different topics. Sentiments and keywords for each candidate are compared(contesting in elections).

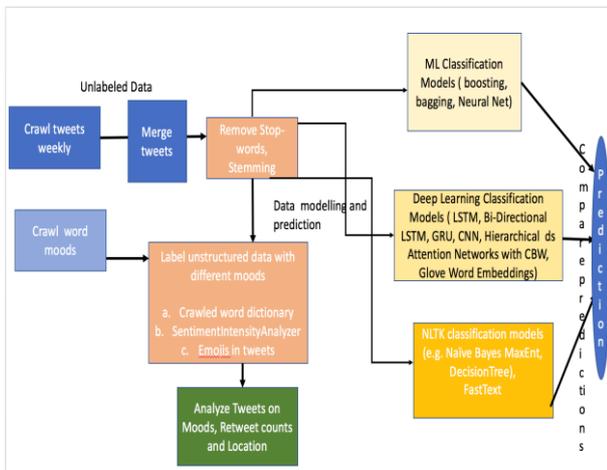
Many politicians also express them through the tweets. Our system has focused on the collection of tweets .Election reated tweets are downloaded mainly analysis based on volume, analysis based on trend, and analysis based on sentiment.

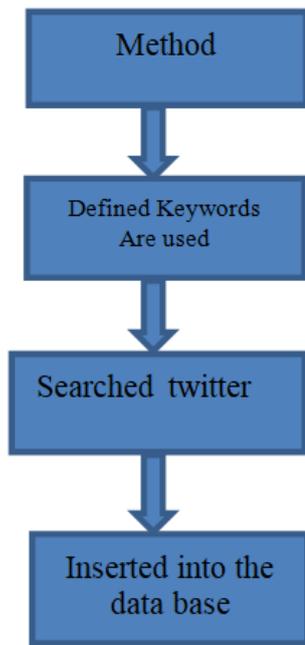
In analysis based on volume tweets are downloaded.. After load[ing actual analysis is done .

We have implemented this system based on python. Data set1: Election-related Tweets downloaded.

Important variables of data collection are keywords that help in meaningful tweets[16][17]. Many kinds of research of keyword collections are used: Candidate's name, Party's name, Election-related features, and campaign hashtags; Time bound to elections which is about 2 months. In this period data related to elections is discussed mostly which may be very important for predicting the election's result. Meaningful data can be collected by the use of more keywords.

In our proposed work method used for implementation was Labelling meaningful keywords.support vector machine,naïve bayes classifier,maximum entropy model for classification.Social sites data or microblogging data used hybrid unigram+bigram for feature selection.In this trended and dynamic keywords are used. For getting dynamic keywords sources like important newspapers, RSS and Topic Modelling technique(LDA) [18] are used.





Data Collection Steps

Keyword used from Tweets: it is based on the TFIDF approach.

The importance of words is measured in multiple documents. TF (Normalised Term Frequency)=How many times a word appeared in that document/Total number of terms in that document.

IDF(Inverse document frequency)=Total number of documents/Number of times a word in it

Keywords used from RSS: During the period of elections we collect words from top news agencies in India. K number of topics is fixed for a document. Collection of words=n.

Collection of trending hashtags from twitter: During the whole election period collection of top trending hashtags is done.

The data is collected before the declaration of the election date.

Data has been collected continuously by using dynamic words.

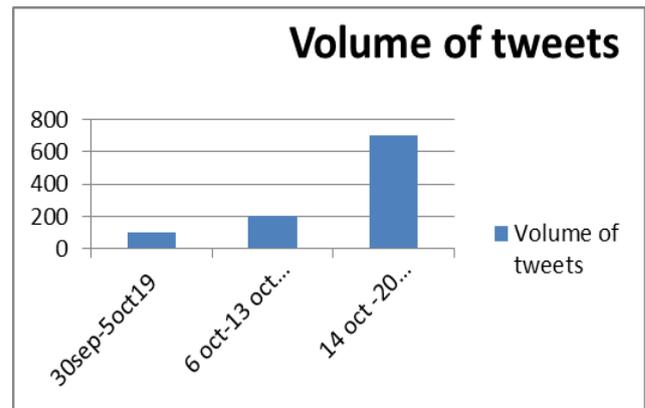
Preprocessing of data is done like this: Each word in the tweet is very important to give decisions. So preprocessing is a crucial task. Tweets are full of slangs, misspellings and using natural language processing. To overcome this problem preprocessing is done step by step like tokenization, lemmatization, feature waiting, stop word removal, stemming. The frequency-based method prepared by [22] is applied.

Data Analysis:

Various meaningful interfaces had been extracted when tweets are examined with different parameters like volume and sentiments. These interfaces help in election result's prediction.

a) Volume analysis: To detect user's sentiments towards Legislative assembly election states in India 2019 has been surveyed before the result outcome date.. All the parties such as Party1(Bhartiya Janta Party), Party2(CONGRESS), Party3(Jannayak Janta Party), etc. The election was declared on 27th Sept. 2019 by the election commission. The date of polling was 21.10.2019 and the result was declared on 24.10.2019. Tweets of each party were collected.

Tweets were focused on the speech, the popularity of leaders, and talking in a media news channel, influenced users and voters. LDA based approach is used to produce results.



In the formation of training and testing, 1000 tweets are taken at different times. The name of candidate and name of party are taken as features of keywords.. The performance follows popular machine learning approaches like SVM, Naïve Bayes, Random Forest, and decision tree are applied for meaningful attributes.

Natural language processing has bag of words model containing unigram (=1 word), bigram(=2 word) and trigram(=3 word). Words are fed as vectors to classify the text. When tweets are cleaned from URL's, stopwords, emojis, hashtags etc words are created.

For classification training sets tweepy and textblob are used. Tweepy provides convenient cursor interface to iterate different types of objects.

Three steps were followed in our program:

1. Authorize twitter API client
2. Fetch the tweets.
3. Parsing tweets for classification.

A twitter client class was created. init function was used to handle authentication. For fetching tweets we used

Fetches_tweets=self.api.search(q=query, count=count)

After parsing tweets sentiment classifier was used to classify sentiments as positive, negative and neutral. Sentiment Polarity method of textblob was used. Classifier follows a principle like this:

Tweets which have polarity value less than zero are negative. When polarity value=0 it is neutral and greater than 0 are positive. It depends on positive or negative words present in it. To analyze textual data textblob was used. For implementation of different tasks in Natural Language Processing it is used. A tweet that has many words (to express sentiment) are highly subjective. To analyze tweets by textblob subjectivity is an important factor.

This applied technique outcomes are compared with other survey agencies and result is published before election result. Result shows twitter is a good source for investing public opinion

Results:

A Dataset was prepared of total 1000 tweets from date 27th Sep 2019 to 20th Oct 2019 to read public moods and facts.. Among 1000 tweets 24 % were positive 22.7% negative and 53. 5% neutral as shown in table below:

Prediction of Election by Twitter

Total no of tweets	Positive Percentage	Negative Percentage	Neutral Percentage
1000	24%	22.7%	53.5%

IV. CONCLUSION

Microblog(Twitter) can help in the prediction of elections result before the actual declaration date. Analysis of different issues that is very important in event predictions like elections are done. This paper, provides a smart technique for prediction

Of election result. In this paper data is collected in different regions. Intelligent techniques are used by this process. Relevant features are introduced which are related with election result. It is based on sentiments of twitters Prediction results were published before election declaration date. This paper verifies that prepared keywords and hashtag-based approach used by social media data offers good results for coming generations. In the future by the use of tweepy and text blob, possibilities of improvement are increasing day by day because social media is one of the key players used by most of the citizens to express their feelings. People mostly use social media sites like Twitter, Facebook, etc. to give their views. The results show how the proposed approach is effective. The proposed system tries to analyze Maharashtra and Haryana assembly elections to study people's opinions. Results show people can express their feelings in the Tweeter by Tweets.

REFERENCES

- Fraser, M. and Dutta, S., Obama's win means future elections must be fought online, nov 2008.
- Wigand, F., 2011. Tweets and retweets: Twitter takes wing in government. *Information Polity*, 16(3), pp.215-224.
- Gayo Avello, D., 2011. Don't turn social media into another Literary Digest poll. *Communications of the ACM*, 54 (10).
- Grosse, K., González, M.P., Chesñevar, C.I. and Maguitman, A.G., 2015. Integrating argumentation and sentiment analysis for mining opinions from Twitter. *AI Communications*, 28(3), pp.387-401.
- Nooralahzadeh, F., Arunachalam, V. and Chiru, C.G., 2013, May. 2012 Presidential Elections on Twitter--An Analysis of How the US and French Election were Reflected in Tweets. In 2013 19th International Conference on Control Systems and Computer Science (pp. 240-246). IEEE.
- O'Connor, B., Balasubramanian, R., Routledge, B.R. and Smith, N.A., 2010, May. From tweets to polls: Linking text sentiment to public opinion time series. In Fourth International AAAI Conference on Weblogs and Social Media.
- Birmingham, A. and Smeaton, A., 2011, November. 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)
- Maynard, D. and Funk, A., 2011, May. Automatic detection of political opinions in tweets. In Extended Semantic Web Conference (pp. 88-99). Springer, Berlin, Heidelberg.
- Larsson, A.O. and Moe, H., 2012. Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New media & society*, 14(5), pp.729-747.
- Soler, J.M., Cuartero, F. and Roblizo, M., 2012, August. Twitter as a tool for predicting elections results. In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 1194-1200). IEEE
- Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S., 2012, July. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 system

- demonstrations (pp. 115-120). Association for Computational Linguistics.
- Cameron, M.P., Barrett, P. and Stewardson, B., 2016. Can social media predict election results? Evidence from New Zealand. *Journal of Political Marketing*, 15(4), pp.416-432.
- Fink, C., Bos, N., Perrone, A., Liu, E. and Kopecky, J., 2013, September. Twitter, public opinion, and the 2011 Nigerian presidential election. In 2013 International Conference on Social Computing (pp. 311-320). IEEE.
- Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L. and Hughes, M., 2013, June. Sentiment analysis of political tweets: Towards an accurate classifier. Association for Computational Linguistics.
- Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., 2010, May. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Fourth international AAAI conference on weblogs and social media.
- Narendra, B., Sai, K.U., Rajesh, G., Hemanth, K., Teja, M.C. and Kumar, K.D., 2016. Sentiment analysis on movie reviews: a comparative study of machine learning algorithms and open source technologies. *International Journal of Intelligent Systems and Applications*, 8(8), p.66.
- Wadhwa, P. and Bhatia, M.P.S., 2014. Discovering hidden networks in on-line social networks. *International Journal of Intelligent Systems and Applications*, 6(5), pp.44-54.
- Blei, D.M., Ng, A. and Jordan, M., 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* (3).
- Jain, V.K. and Kumar, S., 2017. Towards Prediction of Election Outcomes Using Social Media. *International Journal of Intelligent Systems and Applications*, 9(12), p.20.
- Sebastiani, F., 2002. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), pp.1-47.
- Blenn, N., Charalampidou, K. and Doerr, C., 2012, May. Context-Sensitive sentiment classification of short colloquial text. In International Conference on Research in Networking (pp. 97-108). Springer, Berlin, Heidelberg.
- Bao, Y., Quan, C., Wang, L. and Ren, F., 2014, August. The role of pre-processing in twitter sentiment analysis. In International Conference on Intelligent Computing (pp. 615-624). Springer, Cham.
- Bhola, A., 2014. Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General# Elections2014. arXiv preprint arXiv:1406.5059.
- Narasimhamurthy, N., 2014. Use and rise of social media as election campaign medium in India. *International Journal of Interdisciplinary and Multidisciplinary Studies*, 1(8), pp.202-209.

AUTHORS PROFILE



Kusum Mehta, received her B.E. in Computer Science and Engineering in 2002 from Vaish College of Engineering ,Maharshi Dayanand University Rohtak. And received her Master's Degree in Computer Science and Engineering from Banasthali Vidyapeeth in the years 2006. She is pursuing her PhD in Department of Computer Science and Engineering, at Manav Rachna International Institute Of Research and studies, Faridabad, India. Currently, she is doing her research in the area of Big Data under the guidance of Dr. Supriya Panda.



Dr. Supriya Panda, received her Doctorate in Computer Science in the year 1990. She has around Thirty one years of Academic experience in the field of Computer Science. Currently She is working as the Professor in the Department of Computer Science and Engineering at Manav Rachna International Institute Of Research and Studies,, Faridabad, India. She has been the best Teaching Fellow at BGSU, Ohio , USA during MS and Ph.D(1985-90).