Election Prediction Using Deep Learning and Opinion Mining

Smeet Rupapara[1], Saksham Tiwari[2], Manas Kore[3], Kunal Waval[4], Prof.Reena Mahe[5]

[1](Student, Information Technology, Atharva College of Engineering, Mumbai University, India)
[2](Student, Information Technology, Atharva College of Engineering, Mumbai University, India)
[3](Student, Information Technology, Atharva College of Engineering, Mumbai University, India)
[4](Student, Information Technology, Atharva College of Engineering, Mumbai University, India)
[5](Professor, Information Technology, Atharva College of Engineering, Mumbai University, India)

Abstract : *Election is a very popular word. As per definition, the term "election" means a formal and organized choice by vote of a person for a political office or other position. But the process and properties may vary for different sectors. There are different types of elections like General Election (Lok Sabha), Rajya Sabha Election (Upper House), State Assembly Elections or Local Body Elections. As election is very famous, same way election prediction is again not a new keyword. It has same long life as election. But it is quite a challenging task to predict accurate result. The use of opinion mining and deep learning in social science to evaluate the prediction is the core part of this paper.*

Keywords- *Election Prediction, Opinion Mining, Sentimental Analysis, Text Classification, Training Dataset*

I.   INTRODUCTION

Opinion mining is the computational study of a person's sentiments, outlook, approaches and emotions towards an entity That entity could be the individuals participating or contesting in the election, an event or a political topic. It is a type of natural language processing for tracking mood of the public about the particular subject. Opinion mining, also popularly known as sentiment analysis, involves building a system to collect and categorize opinions about a subject. The system that we have developed conducts debates and discussions on political issues. It allows general public to post and express their views about the political parties or individuals that are participating in the upcoming election.

The abundant use of social networks in bygone times has led to end users a potent platform to express their opinions. Platforms like LinkedIn, Facebook, Twitter, Instagram and Google+ are being profoundly used to share opinions, reviews, suggestions and ratings. Application of opinion mining on these social media data has been seen by many as an effective tool to monitor user preferences and inclination. We attempt to provide one such platform to the general public, one that will be specific for political discussions and debates. Also, political parties can manipulate the information gathered from this platform to get intuitions about user's sentiments and thus plan their political campaigns. Political parties can also get a brief insight on their probability of winning the election. Therefore, huge funds are being poured by political parties for social media campaigns during elections. The system that we have developed focuses on assigning a polarity or a strength to each and every comment posted by its users on the ongoing discussion or debate that express their opinions, emotions, sentiments, and so on about the political parties using popular text classification algorithms like Naive Bayes and Porter Stemming Algorithm which are Supervised Learning Algorithm, meaning, they require a training data set to perform sentiment analysis.

II.   LITERATURE REVIEW

In [1], the authors compare the performance of two popular sentiment analysis algorithms based on Supervised Learning, namely Naive Bayes and SVM. Supervised Learning is a branch of Machine Learning which needs a training data set to perform classification. In [2], the authors predict that the credibility of the tweets being analyzed can be checked by taking reviews from other website and sources and then the prediction or sentiment value could be generated. The sarcastic element would be further reduced in the future by better processing. Upon studying Levin's verb categories we determined that only certain classes of verbs are useful in opinion analysis.A. Yue Jiao, Yu-Ru Syau and E. Stanley Lec (May 2005) :This project purposes a new way to predict the presidential election by using a fuzzy adaptive network (fan). Because, election process is inherently vague and cannot be written easily in precise numbers. Linguistic variables are used for public opinion and economic conditions. 7 these variables cannot be expressed in precise numbers due to the mutual dependencies between the various factors. Various models are discussed in this research project e.g. National time series model. But this model was totally based upon the economic conditions and public opinion was not considered. Some factors like "national security" were used in FAN to calculate the prediction. This factor played vital role

in 2004 elections, but was not so important in previous elections. The values for some variables were used as low, medium and high. In this proposed model, more number of linguistic variables will be used which implies the more accurate results.

B. Gleeiber Fernandes Royes, Dr. Rogerio Cid Bastos (2001): Decision making is an important aspect in the field of computer science. But how can we use this decision making into political decision making? Several areas are now using the concept of neural network or fuzzy logic to predict the uncertainty. Appling the Fuzzy logic in social science is the suitable technique because it provides the same result as like an expert in that area. Due to the direct relation with human habits, social science is always considered as different from exact science. In this research the level of acceptance is calculated by using membership function via the concept of fuzzy logic. The quality, to predict the result in election system shows the reliability of fuzzy logic in the area of political science. This research was also tested in the election for mayor in the state of Santa Catarania, Brazil on October 2000 and the result declared by the administration was nearby with the predicted result. This shows the accuracy of result prediction by using fuzzy logic.

According to [4,5], the Big Data in social media represents all the all data generated by social media. These data are recognized by : the large volume, the noise that can generate spam and the dynamic aspect the frequent changes per day [6].

Social Media Platform	Total Number Of Active Users
Twitter	330 Million
Facebook	2.2 Billion
Google+	111 Million
LinkedIn	467 Million

Table 1: Number of active users on popular social media

III. PROPOSED SYSTEM

The system that we have created needs to be as efficient as possible. User first need to register on the website. He/she can then take part in discussions and debates conducted on the website and post his/her opinion about any politician or party taking part in the upcoming election.

Fig 1. BASIC WORKING OF THE SYSTEM

A. Data collection

All these reviews posted by the users on a specific topic are grouped together for data preprocessing.

B. Data Preprocessing

Each and every review is then processed upon and broken down into components using spaces, commas, special characters, etc. Also, common words (like "the", "and", "is", "a", "or", etc.) that do not hold any value are neglected.

C. Applying Algorithms

The algorithm we have used for removing the commoner morphological and in flexional endings from words in the ending is Porter Stemming (or Porter Stemmer). All other remaining words are then mapped to previously created dictionary that contain value (ranging from -1 to +1, where -1 to 0 represents negative polarity, 0 represents neutrality and 0 to +1 represents positive polarity) for each of those words.

For better understanding, let us take an example. Suppose a user posted a comment "I am not feeling good about Donald Trump being our President" during the American Presidential Election. If we break down this comment into terms (or words) and map them to the pre-defined sentiment types (previously created dictionary) it would look something like this:

Token	Sentiment Mapping
I	-
Am	-
Not	Negative
Feeling	-
Good	Positive
About	-
Donald	-
Trump	-
Being	-
Our	-
President	-

Table 2: Basic example of sentiment analysis

IV. CONCLUSION

The use of social media platform like ours for prediction of election results poses challenges at different stages. The very first challenge we faced was the scarcity of training dataset for text classification. We propose our model for election result prediction which uses the labeled data created using our framework. While our model alone may not be sufficient to predict the results, however, it becomes a crucial component when combined with other statistical models. However, this model can be extended in the future to create an automated framework which mines data for months since election result prediction is a continuous process and requires analysis over long periods of time. Thus we recommend creating an Active learning model wherein the model itself recommends what data should be labeled. This would minimize the efforts for labeling while making sure that there is no compromise on contextual relevance.

From the papers we referred we learned that by training and verifying the sentiment classification by the same person, we could archive a high degree of accuracy. This method is suitable to train and classify sentiments from various blogs/twitter and reviews on political points or issues. This method is also a good candidate to assist human/operator to classify a large number of tweets and also suitable for political or business Sentiment classification.

REFERENCES

[1]   Y. Yang and F. Zhou, "Microblog Sentiment Analysis Algorithm Research and Implementation Based on Classification", 2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES) ,2015.

[2]   Rushlene Kaur Bakshi, Navneet Kaun, Ravneet Kaur, Gurpreet Kaur, Opinion Mining And Sentiment Analysis, IEEE International Conference On Computing For Sustainable Global Development, October 2016.

[3]   Harmanjit Singh, Election Results Prediction System Based On Fuzzy Logic, IEEE International Journal Of Computer Applications, September 2012.

[4]   R.R. Mukkamala, A. Hussain, and R. Vatrapu. Fuzzy-set based sentiment analysis of big social data. In Enterprise Distributed Object Computing Conference (EDOC), 2014 IEEE 18th International, pages 71–80, Sept 2014.

[5]   Duc T Nguyen, Dosam Hwang, and Jason J Jung. Time-frequency social data analytics for understanding social big data. In Intelligent Distributed Computing VIII, pages 223–228. Springer, 2015

[6]   Geoffrey Barbier and Huan Liu. Data mining in social media. In Social Network Data Analytics, pages 327–352. Springer, 2011.