

# Election Prediction on Basis of Sentimental Analysis in 3<sup>rd</sup> World Countries

Hafiz Syed Muhammad Bilal\*, Muhammad Asif Razzaq\*\*, Sungyoung Lee\*

{bilalrizvi, sylee@oslab.khu.ac.kr}\*, 10msitmrazzaq@seecs.edu.pk\*\*

\*Dept. of Computer Engineering, Kyung Hee University, Korea.

\*\*Dept. of Computer Science, SEECS NUST, Pakistan.

The detection of human behavior from social media revolutionized health, business, criminal and political prediction. Significance of it, in incentive transformation of public opinion had already proven for developed countries in improving democratic process of elections. In 3<sup>rd</sup> World countries, voters poll votes for personal interests being unaware of party manifesto or national interest. These issues can be addressed by social media, resulting as ongoing process of improvement for presently adopted electoral procedures. On the optimistic side, people of such countries applied social media to garner support and campaign for political parties in General Elections. Political leaders, parties, and people empowered themselves with social media, in disseminating party's agenda and advocacy of party's ideology on social media without much campaigning cost. To study effectiveness of social media inferred from individual's political behavior, large scale analysis, sentiment detection & tweet classification was done in order to classify, predict and forecast election results. The experimental results depicts that social media content can be used as an effective indicator for capturing political behaviors of different parties positive, negative and neutral behavior of the party followers as well as party campaign impact can be predicted from the analysis.

## 1. INTRODUCTION

In virtual environment, communication of people from different terrestrial regions through which they communicate is the basic theme behind social media. Social media comprises of group of internet-based applications constructed for the creation and discussion of user based contents. Individuals, organizations and different communities are benefited from social media applications. People need to keep themselves in touch with others to be better informed. This explosive growth is very much seen generally for Facebook, LinkedIn, FriendFeed, MySpace etc., and especially Twitter, which has loads of users generating contents called, status messages periodically. These contents are available publically in most of cases utilized by researchers for the purpose of opinion-mining, sentiment analysis, predictions and finding masses attitudes and trends. Our current focus is such social media site, with rich contents of user created data comprising of text, media and images URLs. User based textual contents are readily available and can be obtained by using APIs. These APIs cater for researchers in obtaining real-time as well as some third party APIs also give access to crawl historical social data. Once data is collected, it would undergo several validated steps before classification, required for opinion-mining and sentiment prediction. The results obtained will pave us in predicting future for certain brands, movies, polls, public moods, and also for many other individuals, organizations and community's objectives. Popularity of movies and forecasting

box-office revenues was discussed in depth by Asur and Huberman. In political context, predicting German Elections 2009[3], United States Elections 2010 [6], Singapore General Elections 2011 [2] was studied and sentiment analysis was proven as valid indicator in deducing results in contrary to original polls [3]. Public moods were studied using social media feeds during 'Arab Spring' in which Twitter got special attention and usage. It started from self-immolation of Tunisian which triggered Arab World, resulting in ouster of Tunisian President. Despite all pros and cons social media applications have the overall impact of societal improvement. A virtual crowdsourcing society is formed over the internet sharing impartial viewpoints access to each other which increases cooperation and collaboration. We can say it is not a prerequisite for nation development but it can contribute significantly in purification of nation building processes.

## 2. BACKGROUND

### 2.1 Twitter

Twitter a social networking service was launched in 2006 and gained popularity among people with over 500 million registered users. Twitter is in the top ten most used social media applications on the internet. Twitter simply broadcast tweets with 140 characters or less to the followers (one who chooses other's tweets to be posted in their timeline) around the network. These tweets represent any information in text form or shared link relating personal activity, entertainment, sports, or politics.

Nowadays importance of Twitter has risen as it has proved to be fast, low cost medium for disseminating real time information for any sporting event, disaster, academic conference, political activities, elections etc. It was found to be most effective during some of the worth mentioning events like Michael Jackson's Death 2009, Iranian election protest 2009-10, FIFA World Cup 2010, NBA Finals 2010, Egyptian revolution 2011, Sandy Hurricane 2012 and US Election 2010. During these events twitter users updated their statuses and tweeted thousands of tweets per second and setting up tweet propagation per unit time records. These tweets resulted in revolutions for certain nations, sometimes celebrating one's win, sometimes benefitting people suffering from disasters like earthquake, flooding, and sometimes aiding politician for their campaigns. These events got popularity on the basis of terms, words or phrases referred by users commonly in their tweets, so most frequent used words would rather help in trend topic setting. These trending topics would quickly suggest for a user "what is happening right now". These trending topics mostly range between real-time events. People send their reactions, opinions, endorse someone else ideas and discuss about these events in the form of tweets. These status updates are reflected on user's page and also in the timeline of follower's pages. Direct messages can be sent to targeted users using '@username' for one-to-one correspondence. Tweets starting by 'RT' is indicator for re-tweet in which user endorse and propagate someone else's interesting post to rest of his followers. These tweet statuses and re-tweets will form trends if more frequently discussed over twitter.

## 2.2 Social Media & Politics

Campaigns for elections have found totally new platform provided by different social media applications. The popularity rise can be studied for developed and under developed countries. Political parties, party workers, and politician post their campaign messages over fan pages and twitter accounts, maintained by themselves or paid party employees. Number of followers reflects the popularity of political figure and agenda as a simple rule. Conventional campaigns for elections involves cost, time, exertion for dispersing ones motivational opinion in order to get attention of voters, in contrary to this, social media campaign especially twitter involves no such cost. The politicians and political parties can outreach voters free of cost and in no time in deliberating party's agenda and ideology. The political parties can address users from all walks of life mostly involving younger generation by disseminating appealing agenda goals and objectives. While campaigning on twitter, political parties projects their positivity, in the meantime also propagates negativity of opponents. This sets a mobilization mechanism for voters and followers.

The political developments and events are reflected in tweets and even resulted in top trends

maintained by twitter. German Federal Elections were held in 2009, Tumasjan revealed tweets obtained for related political parties, leader asserted resemblances with originally compiled results and concluded twitter can act as a mirror to offline political landscape [3]. He predicted winner or loser for a candidate with accuracy of around 88.0%. The importance of twitter and other social media applications was also unveiled during Singapore General Elections 2011 which found twitter to be integral part of election campaign and mobilization of citizens to cast vote [1]. The conventional Irish General Elections 2011 results were related by social analysts and researchers with combined approach involving volume based and sentiment based results obtained from twitter [2].

## 2.3 Twitter as Prediction Tool

In 2009 American Association of Public Opinion Researchers (AAPOR) spent \$2 billion for online research, out of which \$1.7 billion were spent for traditional forums, weblogs, and political discussion boards. The main focus of AAPOR was contents obtained from online surveys are beneficial and reliable rather than user generated contents. On the other hand Tamasjan worked around the point that Micro-blogging contents may be effective for opinion mining and predictions. His investigation proved that Micro-blogging content's information do base upon opinions of certain users or vetting of opinions of reliable users in their social networks [3]. Thus their opinion on Micro-blogging forum may have certain weight which cannot be ignored and can be negotiated for opinion mining and predictions. Tamsjan further proved that accurate results and prediction for elections outcomes based on the data obtained from social media rather than political forums, and weblogs already supported by AAPOR. Social media also possess same features pertaining to data as of obtained from financial market containing aggregation mechanism from dispersed bits of information. Price system information can be referred to Micro-blogging twitter data by considering the size of followers and re-tweet rates, and most frequently used terms. These features influenced by human behavior can be used to predict currently occurring events [6]. In addition to above mentioned salient features, detected sentiment also holds valuable data which be aggregated in to meaningful, predictable information. Previous studies as of US Elections, Singapore elections with extensive statistical outcomes proved twitter to be a social sensor for the prediction of electoral results which can be related to poll results.

### 3. RESEARCH QUESTIONS

Using social media with clear-cut little content information, we would like to investigate whether it can be used to predict election results comparable with actual General elections results at different level. It mentioning party names, party leader names are reflective of vote bank in actual or not. We would like to investigate political sentiment analysis with bottom-up approach considering individual political affiliation to the party level campaign sentiment. Following are the compiled research questions to be answered.

- 1 How accurately we can predict election results and detect causes if any, for sentiment shift during Pre-Election Campaign.
- 2 Effect of social media on voters for underdeveloped countries and Factors effecting campaign from outside the Domain.

### 4. METHODOLOGY

#### Qualitative Behavior

To study qualitative behavior of social media by using API we have collected 612,802 responses with average responses per day mentions in Figure 1. These were based on switches / keywords. This resulted in downloading huge amount of relevant and irrelevant responses. Here irrelevant means same keywords but from other geographical region, and in different languages.

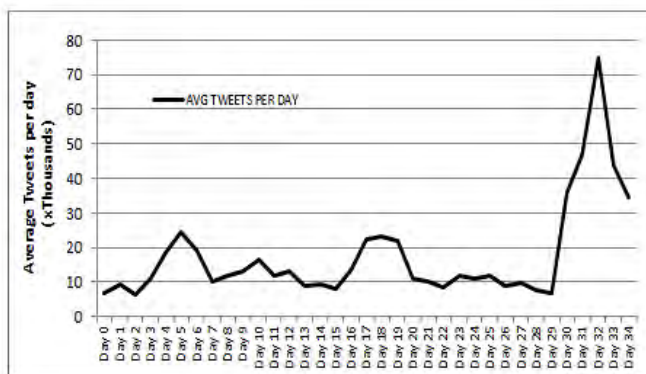


Figure 1. Average tweets per day

After going through around 7,000 responses, about 50% of total were labeled positive, negative and neutral. The labeled data in a text file with each line representing individual response, was then pre-processed.

#### 4.1 Algorithms & Approaches

Using Rainbow we used classifications methods Naïve Bayes (NB), k-nearest neighbor (KNN), term frequency-

inverse document frequency (TFIDF) and probabilistic indexing (Prind).

### 5. CONCLUSION

Sentiment detection / opinion mining from text obtained from social media has been appealing topic for Natural Language Processing. We have tried to find out the predictive power of social media. By using different classifiers we achieved around 70% accuracy for positive and negative sentiment. We also achieved around 50% accuracy for manual labeled data for three classes positive, negative and neutral. This work of forecasting elections for developing countries is of unique attempt. We have found social media user contributing for all political parties. We also deduced that there are certain political parties and leaders who have low electability but high popularity.

#### Acknowledgement

This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2014-(H0301-14-1003)

#### REFERENCES

- [1] Marko Skoric, Nathaniel Poor, Palakorn Achananuparp, Ee-Peng Lim, Jing Jiang. 2012. Tweets and Votes: A Study of the 2011 Singapore General Election. In: 2012 45th Hawaii International Conference on System Sciences.
- [2] Adam Bermingham and Alan F. Smeaton. On Using Twitter to Monitor Political Sentiment and Predict Election Results. Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP) Nov 2011, Chiang Mai, Thailand.
- [3] Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welp. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. International AAAI Conference on Weblogs and Social Media 2010.
- [4] Jessica Chung and Eni Mustafaraj. Can Collective Sentiment Expressed on Twitter Predict Political Elections? Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence
- [5] Brendan O'Connor, Ramnath Balasubramanian, Bryan R. Routledge, Noah A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media

- [6] Daniel Gayo-Avello , Panagiotis T. Metaxas, Eni Mustafaraj. Limits of Electoral Predictions Using Social Media Data. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, July 17-21, 2011