Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Tweets and Votes: A Four-Country Comparison of Volumetric and Sentiment Analysis Approaches

Saifuddin Ahmed
Department of Communication
University of California, Davis
sfdahmed@ucdavis.edu

Kokil Jaidka
Big Data Experience Lab
Adobe Research
jaidka@adobe.com

Marko M. Skoric
Department of Communication
City University of Hong Kong
mskoric@cityu.edu.hk

Abstract

This study analyzes different methodological approaches followed in social media literature and their accuracy in predicting the general elections of four countries. Volumetric and unsupervised and supervised sentiment approaches are adopted for generating 12 metrics to compute predicted vote shares. The findings suggest that Twitter-based predictions can produce accurate results for elections, given the digital environment of a country. A cross-country analyses helps to evaluate the quality of predictions and the influence of different contexts, such as technological development and democratic setups. We recommend future scholars to combine volume, sentiment and network aspects of social media to model voting intentions in developing societies.

Introduction

The wide outreach and popularity of online social networking platforms such as Twitter make them important venues for surveying the political attitudes of the general milieu, who use them to discuss political issues, parties and political leaders. By using frequency of mentions and sentiments expressed on Twitter, several studies have illustrated that Twitter posts can accurately predict election outcomes (Tumasjan et al., 2010; Boutet, Kim, and Yoneki, 2012; Livne et al., 2011). However, some scholars are skeptical about the rigor and reproducibility of the results (Gayo-Avello, 2013).

A review of related literature highlights some important research gaps. Firstly, scholars have mainly focused on economically developed, highly wired and politically stable democracies, such as the United States (Livne et al., 2011), United Kingdom (Boutet et. al, 2012) and Germany (Tumasjan et al., 2010), which comprise a two-party or a multi-party system with low fragmentation. The question that arises is, how would social media-based predictions perform in developing democracies, with localized Internet access and a fragmented political environment? Secondly, there is no clear agreement about which approach, whether volumetric (Tumasjan et al., 2010), sentiment (Bermingham and Smeaton, 2011) or social network (Livne et al., 2011), would yield the most accurate predictions of election outcomes from social media, or why. In a meta-survey, Skoric, Liu and Lampe (2015) found that a volumetric approach was common, but a multi-approach analysis was more accurate.

This study aims to predict elections from Twitter posts, following a cross-country approach to compare the quality of predictions and the role of different technological infrastructures and democratic setups in the General Elections of three Asian democracies (India, 2014; Pakistan and Malaysia, 2013) and a constitutional monarchy (United Kingdom, 2015). We have employed volumetric and sentiment analysis (and a combination of both approaches) based on lexicon and probabilistic modeling to examine approximately 6 million tweets. These countries make a worthy comparative analysis because although social media played an important role in all these elections, the countries are very different in terms of Internet connectivity and political environments. UK (89.8%) and Malaysia (MY) (66.9%) are highly connected as compared to India (IN) (15.1%) and Pakistan (PK) (10.9%). UK and Malaysia have two main competing national parties, while India and Pakistan have several national and regional political parties in their electoral competition.

Method

Data collection and pre-processing

Data collection

We collected tweets from Twitter's streaming API by using Tweet Archivist to track the mentions of political parties and their top two leaders. Approximately 6 million

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

tweets were collected between candidate nomination date and voting day, for 10 parties in UK (2.3 m), 14 parties in Malaysia (1.1m), 15 parties in India (1.2m) and 11 parties in Pakistan (1.1m).

Data cleaning and filtering

There were several unrelated and spam tweets in our dataset; for e.g., the search term “Greens” (referring to the Green party’s members in the UK) created ambiguity. We filtered our dataset with the top election hashtag (#GE13, #GE2013, #GE14, #GE2014, #GE15 or #GE2015), to retain 4.5 million tweets.

Language detection

Python's natural language toolkit was used to identify and separate English tweets from non-English tweets. 100% of the UK dataset, and over 90% of the India and Pakistan datasets were in English; accordingly we have considered only English tweets in their analyses. However, 77% of the Malaysia dataset was in Malay and only 23% was in English; accordingly, we used the complete Malaysia dataset in the volumetric analysis. However, its sentiment analysis was limited to English tweets, because of the lack of an available Malay sentiment lexicon.

Analytical Approach

Election polls

We have used Mean Absolute Error (MAE), the average error in terms of the difference between a set of predicted values (tweets) and actual values (votes):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | e_i |$$

where n is the number of forecasts and e_i is the difference in actual result and predicted result for the ith forecast. However, the relative error of models using MAE is not always observable. By focusing only on the mean, the infrequent big errors are overlooked which are especially important when studying political uses of Twitter, as studies have shown that sometimes minor political parties are typically more popular on social media platforms as compared to majority parties (Ahmed and Skoric, 2014; Jaidka and Ahmed, 2015). To regulate for large errors, we calculated the root mean square error (RMSE) which unlike MAE penalizes variance as it assigns errors with large absolute values more weight than errors with smaller absolute values.

Volumetric analysis

Our volumetric analysis was aimed at measuring the volume of attention or support (i.e., frequency of mentions, supporters, likes etc.) measured by simple counts. We define our volume-based measure as:

$$V_x = \frac{c_x}{\sum_{j=1}^{m} c_j}$$

where *V(x)* represents the volumetric share of tweets for a party *x*, in a system of *j* parties, and c_x is the count of the tweets in the dataset which is relevant to party x. We have calculated the following volumetric predictors for each party: *general mentions* (when a tweet mentioned more than one party), *specific mentions* (when a tweet mentioned just one party), *number of retweets* and *number of @-mentions* (proportion of retweets and tweets with mentions, relevant to a party) and number of *authors of general* and *specific mentions*. Here authors are the Twitter users.

Sentiment analysis

Our sentiment analysis was aimed at measuring the net, positive and negative impressions of each party. We have followed one unsupervised and one supervised approach to identify sentiments. In the former approach, we applied a sentiment lexicon called SentiStrength (Thelwall et al., 2010), to assign a positive score (on a scale of 0 to 5) and a negative score (on a scale of -5 to 0) to all the tweets relevant to a party. SentiStrength draws from several other lexicons; its scores are based on the presence of positive, negative or slang words and emoticons. It is known to have high consistency and reliability and is a popular tool for sentiment analysis in social media (González- Bailón and Paltoglou, 2015). We used the following model to calculate the net sentiment score *senti(t)* for each tweet *t* as:

$$\begin{aligned} senti_t = \{ & pos_t, pos_t > | neg_t | \\ & neg_t, | neg_t | > pos_t \\ & 0, pos_t = | neg_t | \} \end{aligned}$$

where *pos(t)* and *neg(t)* are the postive/negative Sentistrength scores for the tweet *t*.

In our second approach, we implemented supervised text classification for sentiment identification at the tweet-level. This is based on the argument that the opinions of the people posting on social media can be deduced, including the seemingly “neutral” usage of language (Monti et al., 2013). We trained a Naïve Bayes classifier on a hand-annotated dataset comprising tweets about the Ireland’s election in 2011 (Birmingham and Smeaton, 2011) and sentences from political news articles from the New York Times (Sanders, 2011), to identify the top 6000 features predictive of sentiment. The model was evaluated on 669 tweets and produced an overall accuracy of 89%. The precision and recall for identifying positive sentiments was 79% and 93% and the same for negative sentiments was 96% and 87%. The binary sentiment classes were then mapped to binary numeric scores as: $f : [pos, neg] \rightarrow [1, -1]$.

Using sentiments identified from SentiStrength and our Naïve Bayes classifier, we separately measured the corresponding sentiment shares of parties, through the following basic sentiment metrics: *positive (Pos), negative* and *net sentiment shares (Net)*. Combining with volumetric measures, we calculated the *unique authors of positive (PUU)* and *negative tweets*. We also calculated *sentiment* *reach (Rch)*, referring to the outreach of net sentiment by taking into account the number of Twitter followers, thus combining sentiment with network data. We found that the raw negative sentiment share, negative sentiment strength and authors of negative tweets had very high MAEs; they have hence been dropped from the Results section.

Adjustment and controls

While estimating the relationship between the predictors and vote shares, most studies report raw correlation or regression coefficients and ignore other confounders. To represent the correct estimation of metrics across countries, we adopt two strategies – firstly, while calculating the correlations between our metrics and vote share, we adjust our metrics based on the Internet penetration within the country. Secondly, we calculate partial-correlation between our measures and vote shares while controlling for four factors – a party's last election's a) vote share b) seat share c) their current share of Twitter followers d) number of candidates fielded in the present election.

Results

Table 1 reveals that for three out of four countries, the general and specific mention models perform better than both user analyses models. These models work better for UK and Malaysia with smaller variance in the errors as suggested by closer MAEs and RMSEs. Large variances are observed in India and Pakistan predictions. These variances were largely attributed to the over-representation of minority parties – PTI in Pakistan (up to 29.54% error) and AAP in India (up to 10.71% error). Individual party predictions are not presented here for brevity.

Table 1: Summary of volumetric analyses

Metric		UK	MY	IN	PK
General Mention	MAE	1.84	1.63	3.58	6.38
	RMSE	2.42	2.29	5.64	10.34
Specific Mention	MAE	1.62	1.64	3.09	5.88
	RMSE	1.94	2.27	4.45	9.14
General Mention User	MAE	2.05	1.83	2.77	7.85
	RMSE	2.77	2.86	4.05	10.93
Specific Mention User	MAE	1.94	1.69	2.23	5.94
	RMSE	2.66	2.38	3.38	8.53

The results from Table 2 suggest that within a volumetric approach, specific mentions of political parties on Twitter is the best indicator of vote share. All four metrics were found to be highly correlated with actual vote count with statistically significance; the strength of these correlations reduce when the variables are adjusted for the respective Internet penetration rates. Once the controls were included, the lower partial correlations in Table 2 show that the strength of previously observed associations weakened – but these remain the best indicators of vote predictability.

Table 3 provides the results for both sentiment approaches for the respective countries. We find that at the high level supervised analyses outperformed unsupervised analyses in all but three models, with lower errors than the actual vote shares. The difference in performance is especially marked in the cases of the South Asian countries, India and Pakistan.

Table 2: Correlation (r) and Partial Correlation (r_p) of volumetric measures and predictions

	General Mentions		**Specific Mentions**		**General Mention Author**		**Specific Mention Author**	
	UK	**MY**	**UK**	**MY**	**UK**	**MY**	**UK**	**MY**
Raw *r*	$.91^{**}$	$.97^{*}$	$.95^{**}$	$.97^{*}$	$.81^{*}$	$.94^{*}$	$.86^{*}$	$.94^{*}$
Internet *r*	$.89^{**}$	$.94^{*}$	$.94^{*}$	$.96^{*}$	$.80^{*}$	$.93^{*}$	$.85^{*}$	$.93^{*}$
Raw r_p	$.88^{*}$	$.80^{*}$	$.91^{*}$	$.82^{**}$	$.65^{+}$	$.87^{**}$	$.73^{*}$	$.87^{**}$
Internet r_p	$.86^{*}$	$.78^{*}$	$.90^{*}$	$.81^{**}$	$.64^{+}$	$.87^{**}$	$.72^{*}$	$.85^{**}$
	IN	**PK**	**IN**	**PK**	**IN**	**PK**	**IN**	**PK**
Raw *r*	$.86^{**}$	$.65^{*}$	$.85^{**}$	$.74^{*}$	$.71^{*}$	$.64^{*}$	$.76^{*}$	$.68^{*}$
Internet *r*	$.85^{**}$	$.64^{*}$	$.94^{*}$	$.73^{*}$	$.69^{*}$	$.62^{*}$	$.74^{*}$	$.67^{*}$
Raw r_p	$.77^{**}$	.69	$.81^{*}$	$.71^{*}$	.53	.62	$.73^{*}$	$.65^{*}$
Internet r_p	$.75^{**}$	.66	$.79^{*}$	$.70^{*}$	.52	.61	$.72^{*}$	$.64^{*}$

**. Correlation is significant at the 0.01 level; *. Correlation is significant at the 0.05 level; +. Correlation is significant at the 0.10 level.

Drilling into a country-wise comparison, we observe that in the case of UK and Malaysia, all four measures have performed well with low MAEs and RMSEs. For these countries, raw sentiment metrics have performed at par with a combination of sentiment with volumetric or network metrics. However they did not necessarily do better than respective volumetric measures. For India and Pakistan, again, minority parties were over-represented – especially PTI in Pakistan (up to 30.03% error for unsupervised and 27.77% for supervised) and AAP in India (up to 11.66% error for unsupervised and 13.34% for supervised). Individual party predictions are not presented here for brevity. On the other hand, the metrics combining sentiment with volume or network information have performed well for these two nations, as compared against just sentiment or volumetric models. The correlation and partial correlation results (Table 4) follow the same patterns as the volumetric analyses.

Table 3: Summary of sentiment analysis

Unsupervised Sentiment (SentiStrength)					
		UK	**MY**	**IN**	**PK**
Positive (Pos)	MAE	1.61	2.12	4.37	6.68
	RMSE	2.15	2.82	7.48	9.53
Net	MAE	2.49	2.59	3.57	3.82
	RMSE	2.93	3.66	5.16	5.55
Pos Author (PUU)	MAE	1.75	2.01	3.94	6.71
	RMSE	2.21	2.82	6.07	11.06
Reach (Rch)	MAE	2.13	2.58	3.88	4.29
	RMSE	2.42	3.52	6.66	6.11

Table 3 (contd): Summary of sentiment analysis

Supervised Sentiment (Naïve Bayes)					
		UK	**MY**	**IN**	**PK**
Positive (Pos)	MAE	1.77	2.01	3.49	5.39
	RMSE	2.20	2.93	5.48	7.94
Net	MAE	1.97	2.12	3.00	3.44
	RMSE	2.46	3.31	3.89	4.05
Pos Author (PUU)	MAE	1.58	2.10	3.80	6.58
	RMSE	2.21	3.05	6.08	9.00
Reach (Rch)	MAE	1.80	2.14	3.66	6.69
	RMSE	2.41	2.95	5.24	10.26

Table 4: Correlation (r) and Partial Correlation (r_p) of sentiment measures and predictions

	Unsupervised (SentiStrength)				**Supervised (Naïve Bayes)**			
UK	**Pos**	**Net**	**PUU**	**Rch**	**Pos**	**Net**	**PUU**	**Rch**
Raw *r*	.91**	.97*	.91*	.93*	.88*	.89*	.91*	.94*
Internet *r*	.89**	.94*	.91*	.92*	.88*	.88*	.91*	.93*
Raw r_p	.89**	.93**	.88**	.91**	.87**	.87**	.89**	.87**
Internet r_p	.86**	.90**	.87*	.91**	.85**	.87**	.88**	.85**
MALAYSIA								
Raw *r*	.90**	.95*	.88**	.89*	.86**	.89**	.90**	.92*
Internet *r*	.89**	.94**	.89*	.88*	.85**	.89**	.89*	.91*
Raw r_p	.89**	.91**	.81*	.86*	.84**	.87**	.88**	.91*
Internet r_p	.87**	.90*	.80**	.83*	.83**	.85**	.87*	.90*
INDIA								
Raw *r*	.70**	.71*	.72***	.75**	.71*	.73**	.76*	.81*
Internet *r*	.68**	.67**	.72*	.73**	.70*	.72*	.75*	.82*
Raw r_p	.68**	.69**	.71**	.69**	.68*	.72**	.74*	.79
Internet r_p	.63**	.66**	.71**	.68**	.67*	.71**	.73*	.78**
PAKISTAN								
Raw *r*	.62**	.63**	.69**	.67*	.66**	.68**	.71**	.69**
Internet *r*	.61**	.62**	.66**	.66**	.64**	.66**	.69**	.68**
Raw r_p	.61**	.61*	.68*	.64*	.65*	.65*	.69*	.65*
Internet r_p	.59**	.58*	.67*	.63*	.63*	.62*	.67*	.64*

***. Correlation is significant at the 0.001 level; **. Correlation is significant at the 0.01 level; *. Correlation is significant at the 0.05 level.

Conclusion

We have compared the efficacy of a simple volumetric approach, a supervised and unsupervised sentiment approach, and combination analyses in predicting the outcome of elections in very different countries and contexts. Our cross-country predictions suggests that in order to understand the political preferences of citizens through social media, it is rather important to first examine the digital nature of the environment under study. Social media and the Internet have played a key role in the social and political environments of UK over the past decade, and in Malaysia since the last election in 2008. From this, we infer that these societies are well-connected, with widespread access and balanced usage of Internet technologies; thus, simple frequency counts of social media activities are a good representation of the majority preferences of the demographic.

The choice of a methodological approach becomes more critical when one is analyzing contexts which do not have high Internet connectivity or prior experience in the use of Internet in elections, as in India and Pakistan; both nations witnessed the usage of Internet technologies for the first time, in the 2013/2014 elections. Recent studies in the region have found that minority parties are more active on social media (Ahmed and Skoric, 2014) which might be why they are over-represented in volumetric analyses. In these situations, a combination of approaches would help scholars to look beyond frequency counts, to identify which parties are in the lead. Our results also recommend supervised approaches to measure political sentiment – they can enrich a purely volume- or network-based approach and perform better than unsupervised approaches.

References

Ahmed, S., and Skoric, M. M. 2014. My name is Khan: the use of Twitter in the campaign for 2013 Pakistan General Election. In Proc. *HICSS*, 2242-225. IEEE.

Bermingham, A., and Smeaton, A. F. 2011. On Using Twitter to Monitor Political Sentiment and Predict Election Results. *SAIP* 2011:2.

Boutet, A., Kim, H., and Yoneki, E. 2012. What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election. In Proc. *ICWSM*, 411-414.

Gayo-Avello, D. 2013. A meta-analysis of state-of-the-art electoral prediction from Twitter data. *SSCR,* 31(6): 649-679.

González-Bailón, S., and Paltoglou, G. 2015. Signals of Public Opinion in Online Communication A Comparison of Methods and Data Sources. *AAPSS*, 659(1): 95-107.

Jaidka, K., and Ahmed, S. 2015. The 2014 Indian general election on Twitter: an analysis of changing political traditions. In Proc. *ICTD*, 43-47.

Livne, A., Simmons, M. P., Adar, E.; and Adamic, L. A. 2011. The Party Is Over Here: Structure and Content in the 2010 Election. In Proc. *ICWSM*, 17-21.

Monti, C., Rozza, A., Zappella, G., Zignani, M., Arvidsson, A.; and Colleoni, E. (2013). Modelling political disaffection from Twitter data. In Proc. ISDOM, 3.

Sanders, N. 2011. Twitter Sentiment Corpus. Retrieved from http://www.sananalytics.com/lab/twitter-sentiment/

Skoric, M., Liu, J.; and Lampe, C. 2015. Gauging Public Opinion in the Age of Social Media. Paper presented at *ICA* Annual Meeting, 9-13 June.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D.; and Kappas, A. (2010). Sentiment strength detection in short informal text. JASIST, 61(12): 2544-2558.

Tumasjan, A., Sprenger, T. O., Sandner, P. G.; and Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In Proc. ICWSM, 178-185.