Forecasting Canadian Elections Using Twitter

Kenton White[1,2(✉)]

[1] Advanced Symbolics, 41 York Street, 4th Floor, Ottawa, ON, Canada
kenton.white@advancedsymbolics.com
[2] School of Electrical Engineering and Computer Science,
University of Ottawa, 800 King Edward Dr., Ottawa, ON, Canada
kenton.white@uottawa.ca

Abstract. Experiments forecasting Canadian elections with Twitter are presented. A methodology for creating a representative Twitter sample is described and validated against census data. This sample and election polls are input into a VARX forecast model. The model covariance error monitors the forecast accuracy, measuring forecast confidence before the election occurs. The model is tested on several Canadian elections.

1 Introduction

Social media provides an unprecedented window into the political discourse – the tantalizing allure of tapping into a collective consciousness. Could it be possible, using platforms like Facebook and Twitter, to predict the next Prime Minister?

Tumasjan et al. found a correlation between share of Twitter traffic for the 6 main parties in the 2009 German national parliament election and the final percentages [13]. Based on this correlation, they speculated that Twitter could be a leading indicator of political opinion. O'Connor et al. [9], saw a similar leading indicator using topic sentiment. They compared the daily sentiment score for consumer confidence, presidential approval, and the 2008 presidential election race with Gallup polls. For consumer confidence and presidential approval, they found that the daily sentiment score was a leading indicator of consumer polls, but not for election polls. Bermingham and Smeaton [2] performed a similar analysis for the 2011 General Irish Election, comparing several features, like mention volume and sentiment, against the election results. They found that share of volume of mentions, followed closely by the share of positive volume, were the best predictors. Sang and Bos [12] studied the 2011 Dutch Senate elections, again comparing several features. Contrary to Bermingham and Smeaton, they found that sentiment scores were the best predictor. In a novel twist, the researchers introduced poll-dependent weights to correct for demographic differences between Twitter users and the Dutch electorate.

Contrasting these successful demonstrations, Gayo-Avello [4] and Metaxas [8] together analyzed the 2010 US Congressional election. After accounting for incumbency effects, they reported neither tweet volume nor sentiment had predictive value. In their analysis, Gayo-Avello and Metaxas made several critiques:

R. Khoury and C. Drummond (Eds.): Canadian AI 2016, LNAI 9673, pp. 186–191, 2016.
DOI: 10.1007/978-3-319-34111-8_24

– *A Priori Model.* Previous studies chose the best method and feature *a posteriori.* Method and features must be defined before the election starts.
– *Testable Models.* Previous studies could not measure forecast accuracy until after the election. A model must monitor the forecast accuracy.
– *Representative Samples.* Polls use a representative sample of the population for their analysis. Twitter forecasts must use a similar sample for analysis.

This paper directly addresses these critiques.

2 Sampling

Twitter studies collect tweets from the "Garden Hose", a down-sampled stream of real time tweets. These streams are also rate limited, being temporarily disabled when the number of streamed tweets exceeds a fixed amount in a set time. Down sampling and rate limiting produces a sample where the systematic errors are unknown, preventing the researcher from correcting for any biases.

Election studies are restricted to tweets within the election geography. These experiments use the self reported location in the Twitter profile. The location field text is queried against a reverse address lookup service[1] to determine if the user falls within the desired geography.

Following the polling professional methodology, analysis should be performed on a random sample of Twitter users. Samples can be created using the social network graph structure. Each user has public connections with other users. These connections are crawled, extracting a sample. Simple random walks on the graph generate biased samples, undersampling users with a small number of connections while oversampling those with a large number of connections.

Markov Chain Monte Carlo (MCMC) methods remove this bias [6], but with an important caveat: the algorithms converge to the stationary distribution by sampling the same user multiple times. Perfect samplers remove the statistical degeneracy from MCMC samplers. Coupling From The Past (CFTP) [10], a well known perfect sampler, can be modified to work on social networks using Conditional Independence Coupling (CIC) [14,15]. CIC retains the unbiased graph sampling from MCMC methods while avoiding degenerate samples. CIC is the sampling method used in these experiments.

3 Demographics

The representativeness of the sample is verified by directly measuring the demographics of 5,000 sampled Twitter users from Toronto. The 5,000 users were hand classified using Mechanical Turk, collecting 3 classifications for each account. Accounts where 2 out of 3 people agreed on all 3 of the characteristics were kept. The final data set had 3,032 accounts. Random error biases were removed using post stratification weighting [7].

[1] Microsoft's Bing Maps, https://msdn.microsoft.com/en-us/library/ff701713.aspx.

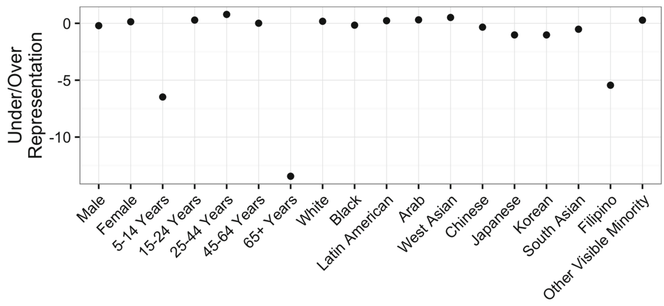

Fig. 1. Under/Over representation in weighted sample data compared against the 2011 Toronto census.

Figure 1 shows the deviation of the weighted Twitter data from the 2011 Toronto census data. The demographics of the online users follow the census demographics [3] with some important exceptions. Users under the age of 14 years were underrepresented, since Twitter requires that all users must be over the age of 14. Users over the age of 65 were also under represented, supporting anecdotal observations that Twitter is primarily used by younger people. Finally, Filipinos were underrepresented due to difficulty classifiers had differentiating Filipino ethnic traits.

4 Forecasting

O'Connor [9] had noted that Twitter data was a leading indicator of some public opinion polls. Achrekar et al. formalized this observation with a Vector Autoregression with Exogeneous Variables (VARX) model forecasting CDC influenza data [1]. VARX models require an output variable (the variable to be forecast) and an input variable (the variable from which the forecast will be derived).

Here a similar VARX technique is used for forecasting elections where election polls are the output variable and Twitter data is the input variable. The VARX model is implemented using the `R` package `DSE` [5]. The output variable is a time series of the aggregated daily polls for each major party. The input variable is an appropriate Twitter feature.

Which Twitter feature to use is determined by experiment on the 2014 Ontario Election. A sample of 32,339 Ontario Twitter users with 2,500,821 tweets between May 1, 2014 and 8 pm on June 12, 2014 (when the polls closed) was collected. For each of the 3 main parties (Liberals, Conservatives, and NDP) mentions of the party and the party leader were tracked.

Six separate features were tested:

– Tweet Volume (mentions of a candidate).
– Tweet SoV (unique people mentioning a candidate).
– Positive Volume (positive mentions of a candidate).
– Positive SoV (unique people positively mentioning a candidate).

Table 1. Comparison of different features for the 2014 Ontario election. Winning party is in bold.

	Without post stratification			With post stratification		
	LPC	CPC	NDP	LPC	CPC	NDP
Tweet volume	30.2 %	**50.7 %**	19.1 %	32.4 %	**46.7 %**	20.9 %
Tweet SoV	32.3 %	**50.4 %**	17.4 %	35.1 %	**50.3 %**	14.6 %
Positive volume	34.2 %	**54.9 %**	10.9 %	35.0 %	**51.9 %**	13.1 %
Positive SoV	38.5 %	**50.3 %**	11.2 %	40.5 %	**49.0 %**	10.5 %
Mean sentiment volume	**39.5 %**	37.4 %	23.1 %	38.8 %	**39.0 %**	22.2 %
Mean sentiment SoV	**41.1 %**	35.7 %	23.2 %	**40.1 %**	33.7 %	26.2 %
Election results[a]	**41.3 %**	33.3 %	25.3 %	**41.3 %**	33.3 %	25.3 %

[a] Election results are normalized to the top 3 parties

– Mean Sentiment Volume (mean sentiment score for each candidate).
– Mean Sentiment SoV (person's mean sentiment score for each candidate).

Mentions were normalized against the total people tweeting in a given day. Sentiment is calculated using word polarity [11]. For each of the 6 features the model was run with and without post stratification weighting. (demographics were determined as described in Sect. 3). Results are summarized in Table 1. Mean sentiment SoV provides the best final election prediction. Including demographic weighting does not improve the results. In the subsequent experiments, mean sentiment SoV without demographic weighting will be used.

The covariance matrix is an estimate on the forecast accuracy and can be used as a measure of the forecast confidence. This measure's effectiveness was tested on 2 recent provincial elections – Alberta 2012 and British Columbia 2013 – where the election polls missed the winning party by large margins. Since the model is trained on the election polls, the Twitter forecast should not be substantially different, but the covariance error should be large enough to cast doubt on the forecast accuracy.

A sample of 938,831 tweets from 24,015 Alberta Twitter users between March 22, 2012 and April 23, 2012 and 1,589,339 tweets from 32,603 British Columbia Twitter users between April 3, 2013 and May 14, 2013 was used. The polls are extrapolated to election day using a standard ARIMA model (without information from Twitter), which allows the election polls the benefit of any momentum changes from the days leading up to the election.

Table 2 summarizes the ARIMA and VARX errors. In both elections the VARX model, using the data from Twitter, has a much larger error than the corresponding ARIMA models that only use the poll data. For comparison, the error using an ARIMA model and a VARX model for the Ontario election is also shown. In the Ontario election, where the Twitter forecast agreed well with the election results, the covariance error is lower.

The model was tested live on the 2015 Federal Canadian election. A sample of 34,732,633 tweets from 130,816 Canadians between January 1, 2015 and

Table 2. Comparison of error for Alberta 2012 (top), British Columbia 2013 (middle) and Ontario 2014 (bottom) election. Winning party is in bold.

	Alberta			
	PC	Wildrose	Lib	NDP
ARIMA	35.8 ± 0.9 %	**38.1 ± 1.1** %	11.1 ± 0.6 %	11.4 ± 0.2 %
VARX	35.2 ± 4.6 %	**37.6 ± 6.5** %	12.1 ± 3.6 %	11.1 ± 1.5 %
Actual	**44.0 %**	34.3 %	9.9 %	8.5 %
	British columbia			
	Lib	NDP	Green	
ARIMA	36.4 ± 1.0 %	**44.3 ± 1.9** %	11.2 ± 1.5 %	
VARX	34.3 ± 3.1 %	**42.4 ± 9.7** %	11.8 ± 6.3 %	
Actual	**44.1 %**	39.7 %	8.2 %	
	Ontario			
	Lib	PC	NDP	
ARIMA	**36.3 ± 1.0** %	31.8 ± 1.4 %	23.5 ± 1.6 %	
VARX	**37.0 ± 1.1** %	32.3 ± 1.3 %	23.0 ± 1.7 %	
Actual	**38.7 %**	31.3 %	23.8 %	

Table 3. Summary of the 2015 Federal Canadian election. Winning party is in bold.

	Twitter			Polls			Actual		
	LPC	CPC	NDP	LPC	CPC	NDP	LPC	CPC	NDP
Canada	**38.1 ± 0.8**	31.1 ± 0.7	18.7 ± 0.9	**38.0 ± 2.1**	30.7 ± 1.1	20.5 ± 0.6	**39.5**	31.9	19.7
Atlantic	**54.0 ± 2.9**	18.7 ± 2.1	19.7 ± 2.1	**54.0 ± 4.6**	17.7 ± 2.5	21.1 ± 1.7	**59.1**	18.2	17.9
Quebec	**28.4 ± 1.9**	20.5 ± 1.7	24.3 ± 2.1	**30.1 ± 5.6**	18.1 ± 1.6	25.4 ± 4.1	**35.7**	16.7	25.4
Ontario	**45.6 ± 1.8**	32.7 ± 2.1	15.7 ± 2.1	**45.4 ± 0.7**	33.1 ± 1.2	16.6 ± 1.3	**44.8**	35.0	16.6
Prairies	33.5 ± 2.5	**42.5 ± 3.0**	18.4 ± 2.8	36.0 ± 4.4	**40.2 ± 5.6**	18.1 ± 0.9	34.3	**42.9**	19.5
Alberta	28.0 ± 2.0	**54.1 ± 1.9**	17.2 ± 2.0	25.3 ± 6.6	**52.8 ± 5.6**	18.1 ± 4.6	24.6	**59.5**	11.6
BC	**35.6 ± 1.9**	32.2 ± 2.0	23.2 ± 2.1	32.6 ± 4.5	**32.8 ± 4.6**	25.1 ± 2	**35.2**	30.0	25.9

October 19, 2015 was collected. Daily forecast models were run against national and provincial polls. Final forecast numbers were calculated when the polls closed.

Table 3 summarizes the Twitter forecast results. The mean value of the previous day's polls as well as the final election results are provided for reference. The Twitter forecast model correctly predicted the overall Canadian election as well as the provincial results. In contrast, the polls missed the correct winner in British Columbia. Both the polls and the Twitter forecast underestimated the Liberal's win in Quebec. Covariance errors in all of the Twitter forecasts were low, indicating confidence in the forecast result.

References

1. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.H., Liu, B.: Twitter improves seasonal influenza prediction. In: HEALTHINF, pp. 61–70 (2012)
2. Bermingham, A., Smeaton, A.F.: On using twitter to monitor political sentiment and predict election results. In: Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP 2011), pp. 2–10 (2011)
3. Canada, S.: Toronto, ontario (code 3520005) and ontario (code 35) (table) (2012). http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/index.cfm?Lang=E
4. Gayo Avello, D., Metaxas, P.T., Mustafaraj, E.: Limits of electoral predictions using twitter. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence (2011)
5. Gilbert, P.D.: Brief User's Guide: Dynamic Systems Estimation (2006). http://cran.r-project.org/web/packages/dse/vignettes/Guide.pdf
6. Gjoka, M., Kurant, M., Butts, C.T., Markopoulou, A.: Walking in facebook: a case study of unbiased sampling of osns. In: Proceedings of the IEEE INFOCOM, 2010, pp. 1–9. IEEE (2010)
7. Lumley, T.: Analysis of complex survey samples. J. Stat. Softw. **9**(1), 1–19 (2004)
8. Metaxas, P.T., Mustafaraj, E., Gayo-Avello, D.: How (not) to predict elections. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), pp. 165–171. IEEE (2011)
9. O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: linking text sentiment to public opinion time series. ICWSM **11**, 122–129 (2010)
10. Propp, J.G., Wilson, D.B.: Exact sampling with coupled markov chains and applications to statistical mechanics. Random Struct. Algorithms **9**(1–2), 223–252 (1996)
11. Rinker, T.W.: qdap: Quantitative Discourse Analysis Package. University at Buffalo/SUNY, Buffalo (2013). http://github.com/trinker/qdap 2.2.4
12. Sang, E.T.K., Bos, J.: Predicting the 2011 dutch senate election results with twitter. In: Proceedings of the Workshop on Semantic Analysis in Social Media, pp. 53–60. Association for Computational Linguistics (2012)
13. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: what 140 characters reveal about political sentiment. ICWSM **10**, 178–185 (2010)
14. White, J.K., Li, G.: Method and system for sampling online social networks, April 2015. https://patents.google.com/patent/US20150095415A1/en
15. White, K., Li, G., Japkowicz, N.: Sampling online social networks using coupling from the past. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), pp. 266–272. IEEE (2012)