Multi-cycle forecasting of Congressional elections with social media

Mark Huberty[*]
Travers Department of Political Science
University of California, Berkeley
markhuberty@berkeley.edu

ABSTRACT

Twitter has become a controversial medium for election forecasting. We provide further evidence that simplistic forecasting methods do not perform well on forward-looking forecasts. We introduce a new estimator that models the language of campaign-relevant Twitter messages. We show that this algorithm out-performs incumbency in out-of-sample tests for the 2010 election on which it was trained. That success, however, collapses when the same algorithm is used to forecast the 2012 election. We further demonstrate that volume-based and sentiment-based alternatives also fail to forecast future elections, despite promising performance in back-casting tests. We suggest that whatever information these simplistic forecasts capture above and beyond incumbency, that information is highly ephemeral and thus a weak performer for future election forecasts.

1. INTRODUCTION

Social media promises a real-time, readily available data source with which to introspect into the behavior of society at large. Many studies have suggested that this data can augment or supplant traditional measures of social attitudes like polling or surveys. Elections in particular have seen

This paper provides evidence of the difficulty of building effective forecasts of US elections using social media. We present the results of one of the first multi-cycle experiments in election prediction using social media data. We show that algorithms trained on one election perform poorly on a subsequent election, despite having performed well on out-of-sample tests on the original election. We provide ev-

idence that this failure stems from the volatility of the underlying data-generating process, even in an election system with very short periods between elections such as in the U.S. House of Representatives. We further show that this problem exists for other otherwise promising forecasting methods as well. In short, simplistic methods for forecasting elections from Twitter, even when their results are correlated with election outcomes, provide relatively little added benefit.

2. SOCIAL MEDIA AS AN ELECTION FORECAST

Twitter has become a popular medium for forecasting offline political behavior from visible online behavior. As an information-push medium, tweets promise an unvarnished, if also unstructured, look into individuals' political attitudes. However, successful predictors have proven ephemeral. Claims by [14] to have successfully forecast the 2009 German elections using Twitter data, were shown to be an artifact of researcher choices rather than research design [7]. Mixing sentiment analysis and relative attention on Twitter to different candidates appeared promising, but under-performed conventional polling in the 2011 Republic of Ireland general elections [1]. [13] performed somewhat better in the 2011 Dutch elections, but their best results relied on ad-hoc re-weighting using the very polling information that Twitter-based forecasts often aspire to replace. Finally, [11] show that Twitter sentiment may correlate with political polling, but nevertheless offers weak predictive power for actual election outcomes.

These problems indicate a much broader problem for election prediction via social media. Given the demographic differences between the Twitter user base and the voting population [9], the inherent dynamism of political language and activity, the partisan polarization of the Twitter community [3], and incentives for strategic behavior by campaigns and motivated partisans, simple heuristics appear unlikely to perform reliably as electoral predictors. At the very least, they argue, valid claims for any prediction should require the analyst to offer predictions ahead of time, clearly articulate how their predictive algorithm works, and establish a reasonable baseline–almost certainly not random chance– against which their predictions should be judged [8]. To date, very few forecasts have done so.

U.S. federal elections may pose a particularly hard task for social media-based forecasts. Most US elections are decided between only two parties. Of those districts in which two candidates actually run, only a fraction are actually competitive: incumbents win re-election more than 85% of the time,

even in anti-incumbent years like 2010. Even very close to the 50% win / loss cutpoint, evidence suggests that incumbents maintain advantages over their challengers [2], something possibly untrue of other political systems [5]. Partisan control of election district boundaries and other facets of election administration may reinforce this outcome. Hence US elections pose a very high bar: forecasts must beat a simple heuristic, incumbency, that reliably forecasts future winners with high accuracy, even in ostensibly competitive races.

3. AN N-GRAM FORECASTING MODEL

We first describe a new forecast based on the Twitter micro-blogging service. This method differs from earlier methods by modeling the language of candidates' Twitter feeds, rather than using simple counts or sentiment scores. Based on models built from the 2010 U.S. House of Representatives election, we generated forward-looking forecasts for the 2012 election outcomes. In line with recommendations from [8], we pre-released all data acquisition, cleaning, and forecasting code. Forecasts themselves were published daily ahead of the 2012 election.[1]

3.1 Data acquisition

All models used Twitter data acquired via the Twitter Search API.[2] Searches were performed each day, looking for all mentions of every known Republican or Democratic general election candidate in the prior 24 hours. We retrieved all tweets possible, up to the API limit of 1500 messages per query. Summary statistics on Twitter message volume by candidate are shown in table 1. Data gathering began on September 1 for the 2010 election, and on September 12 for the 2012 election. The final data sets included approximately 260,000 messages for 313 districts in the 2010 election, and 1.3 million messages for 369 districts in the 2012 election.

All data were filtered for noise prior to conversion to the bi-gram bag-of-words model described above. Filtering attempted to identify spam via a Latent Dirichlet Allocation topic model; tweets were subsequently excluded based on the presence of terms found in "noise" topics. For example, sports-related messages were common sources of irrelevant data. The sportscaster Stephen Smith shares a name with a candidate for California's District 34. Consequently, terms

---

[1]All code for data acquisition, cleaning, and prediction was released at `https://github.com/markhuberty/twitter_election2012`. All predictions were published in real-time at `http://californianewsservice.org/category/tweet-vote/`. The only exception to pre-release was a bugfix that altered the last several days of prediction prior to the 2012 election. An off-by-one error corrupted certain data and generated a spurious collapse in prediction accuracy. That collapse disappeared when we re-created the prediction inputs from raw data.

[2]Note that this differs from other papers, which tend to use variants of the Twitter streaming API. That API may not replicate the actual content of Twitter well [10]. We use the search API in this instance because it provides a means of gathering all mentions of a candidate. Exceptions here include very high-volume candidates like Nancy Pelosi or Paul Ryan, whose daily mention volume exceeds 1500. In those cases, a candidate's data is right-censored. However, most of those cases concern races that weren't very competitive anyway.

like `mlb` and `yankees` were signals for politically-irrelevant data that were inadvertently captured because of the name homonym. This cleaning reduced overall message volumes by 25,000 in 2010, and by 200,000 in 2012.

Year	Party	Inc. Party	Median	Mean	Std. Dev.
2010	D	D	123.0	566.6	2432.8
2010	D	R	18.0	88.8	256.4
2010	R	D	91.0	275.3	790.0
2010	R	R	51.0	345.1	1036.6
2012	D	D	585.0	2001.9	6679.9
2012	D	O	415.0	1338.7	2244.5
2012	D	R	144.0	818.4	2774.5
2012	R	D	164.0	753.7	2118.9
2012	R	O	481.5	1975.5	4724.0
2012	R	R	507.0	2256.5	7502.3

Table 1: Message volumes by party, district incumbency, and election. This table provides summary statistics for candidate Twitter message volume for the 2010 and 2012 elections. Incumbent party refers to the party of the district incumbent, regardless of whether the incumbent stood for re-election. 'O' refers to open districts created after the 2010 redistricting.

3.2 Data characteristics

The cleaned data illustrate two important results. First, Twitter volumes are strongly biased in favor of incumbents. As figure 1 shows, incumbents received significantly greater attention on Twitter than challengers. The detailed breakdown in table 1 shows that a Democratic incumbent received approximately 33% more messages than their Republican challenger in 2010; and nearly three times more in 2012. Similar results obtain for Republican incumbents. This imbalance suggests why volume-based forecasting algorithms (e.g., [14, 1]) may work: candidates' message volumes echo an ingrained bias towards incumbents that manifests itself across a variety of measures (fund raising, conventional media attention, name familiarity), and which correlates well with high incumbent rates of success. Second, as shown in figure 2 shows, highly competitive elections–those decided by small margins around the 50% cutpoint–receive significantly more attention from Twitter users than safe seats. Yet even there, incumbents continue to receive far more attention than challengers.

3.3 Model description and assumptions

Our model departs from earlier attempts at Twitter-based election prediction by modeling the language of the Twitter message feed itself. Other estimates have used message volume or naïve sentiment analysis. Either method assumes the meaning of Twitter language *a priori*: either through an "all news is good news" model that ignores such language entirely, or by ignoring linguistic context when applying sentiment lexicons. In contrast, we use one election as an opportunity to learn predictive weights for language bigrams, on the assumption that the salience of linguistic cues is relatively static over short election cycles.

The model works as follows. In 2010, we gathered data on contested U.S. House races between a Democrat (D) and a Republican (R). Messages were case-standardized and
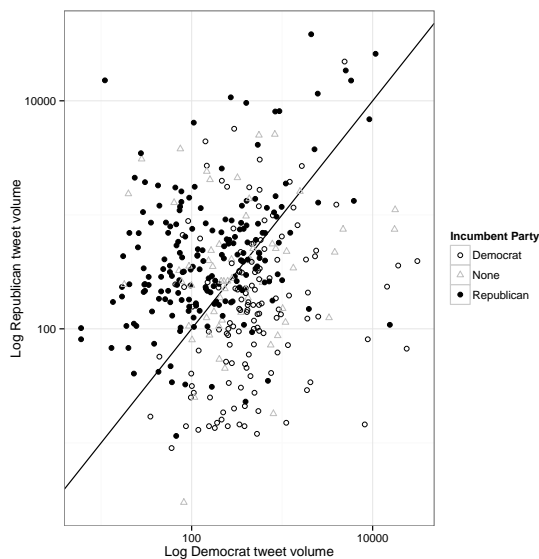
Figure 1: Incumbents receive far more attention from Twitter users than challengers. This figure shows the comparative message volume for candidate pairs in each district. Fill indicate the party of the district incumbent. The diagonal line illustrates where points would fall if both candidates in a district received equal message volume.
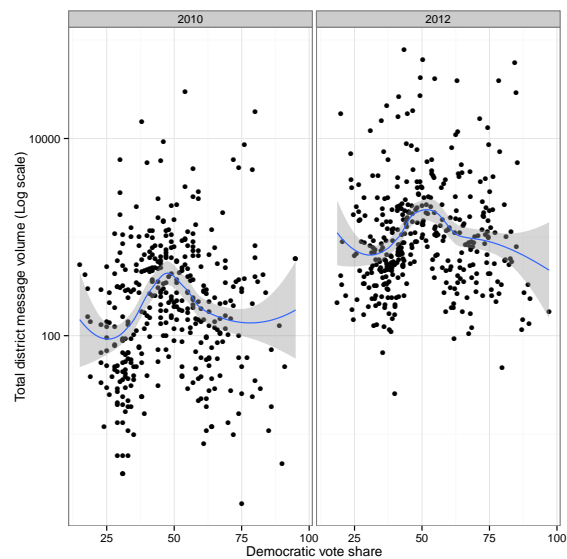

Figure 2: Competitive districts receive more attention on Twitter. This figure shows how total district message volume varies with district competitiveness. Elections decided around the 50% cut point receive orders of magnitude more attention from Twitter users than less competitive races.

stripped of English stopwords, URLs, and non-ASCII characters. Candidates' proper names were standardized to party-specific placeholders ("Rcanddummy", "Dcanddummy"). Each message was then converted to a bi-gram bag-of-words term-frequency representation. Term frequencies for all messages pertaining to a single race–to one R-D candidate pair–were then summed to generate a single bi-gram term-frequency vector for each election race. Terms present in fewer than 1%, or more than 99%, of races, were discarded. Vectors were normalized to sum to 1.

Using this bi-gram bag-of-words representation for districts, we trained an ensemble machine learning algorithm on district-level election outcomes for 2010. Outcomes were either continuous (Democratic share of the two-party vote) or binary (Democratic win / loss). Both algorithms relied on the SuperLearner ensemble supervised learning algorithm [16, 12], which trains a library of standard machine learning algorithms against labeled data.[3] Weights for each library member are learned via minimization of the cross-validated risk of the ensemble forecast. Accuracy is bounded on the low end by the best predictive performance of all the individual algorithms in the library.

Examination of the structure of the learned model provides insight into how the forecasting algorithm works. Examination of algorithm weights in the final SuperLearner

vote share algorithm showed that it was dominated by variants of the random forest algorithm with different tuning parameters. Figure 3 illustrates the most influential terms within one of those random forest variants. The final models are dominated by two sets of terms: one set that signal for the party of the incumbent candidate, and another that pick up on salient political issues. This is a sensible general model: given the high rate of incumbent re-election in U.S. politics, any reasonable model should start from the assumption of incumbent victory, and adjust from that baseline given politically-salient issues and other factors.

3.4 Model performance

All models were compared against the baseline rate of incumbent re-election. Forecasting models were trained on 2010 data. Back-casting 2010 results on an out-of-sample data set suggested that the algorithms could beat the incumbency baseline. As table 2 shows, the algorithm beat the rate of incumbent re-election in Democratic districts, and equalled it in Republican districts.

However, this model fared significantly worse when generating forward-looking forecasts of the the 2012 election. While it once again equalled the rate of incumbent success in Republican districts, forecasts for Democratic districts fell far short of the incumbency re-election baseline. For open seats, without incumbents, created by the post-2010 election redistricting, forecasts beat simple chance but only predicted two-thirds of races correctly.

Thus the multi-cycle test presented here invalidates the assumption implicit in the forecasting algorithm. The 2010 and 2012 elections were fought over similar issues, such as healthcare regulation and fiscal policy. But the underlying dynamics of Twitter use and content–including the generation of newly salient issues and changing representation

---

[3]The libraries in this case were specified to handle high-dimensional, sparse data. For win-loss prediction, the library included variants on: lasso, support vector machines with various kernel parameters, and random forests with various tuning parameters. For vote share prediction, this ensemble was expanded to include boosted regression, sparse partial least squares, step regression, ridge regression, and multivariate adaptive splines.

Year	Incumbent party	Voteshare accuracy	Win-loss accuracy	N	Incumbent win rate
2010	D	0.82	0.86	200	0.69
2010	R	0.98	0.99	113	0.98
2012	D	0.86	0.85	159	0.90
2012	O	0.67	0.67	21	
2012	R	0.90	0.90	189	0.91

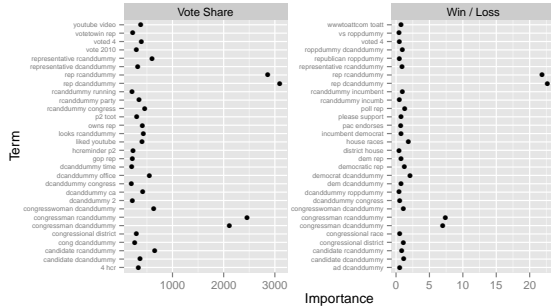Table 2: Predictive accuracy by election and district incumbent.

Figure 3: Prediction algorithms key in on incumbency indicators. This figure shows the estimated term importance for the random forest algorithm component of the vote share and win-loss ensembles. Term importance is estimated as the normalized change in predictive error upon random permutation of each term. Each panel shows the top 30 terms by importance for the highest-weighted random forest member of the SuperLearner library.
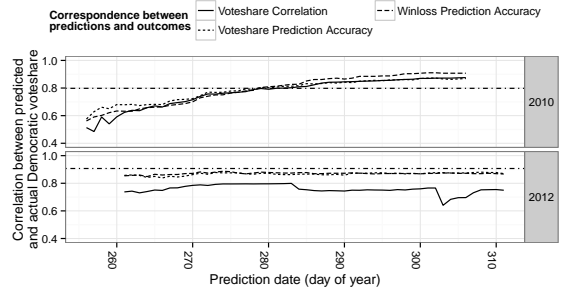

Figure 4: Predictive accuracy degrades between elections. This figure shows that algorithms capable of surpassing the incumbency baseline in 2010 were unable to do so in 2012. Predictions are back-cast for the 2010 election, using the trained algorithm; and forecast for the 2012 election. Vote shares were converted to win/loss predictions at the 50% cut point. Horizontal lines indicate the incumbent win rate for the districts in the total population of forecast districts.

of older issues–were not stable enough to permit an algorithm that showed promise in one election to perform well in the subsequent one. Instead, the incumbency portion of the forecast remained valid, while the adjustment from that baseline fell apart.

4. COMPARATIVE PERFORMANCE OF ALTERNATIVE FORECASTS

Other similarly promising Twitter-based forecasts may suffer from similar problems [6]. Given the data we have available, we are also able to test the multi-cycle performance of these methods. We provide two such tests: one based on relative candidate volumes as used in [14] and [4]; and the other based on naïve sentiment analysis using the OpinionFinder sentiment corpus [17]. Both methods attempt to add to the predictive power of incumbency in forecasting future elections. We show that neither method does so when applied to out-of-sample forecasts for future elections, as opposed to in-sample predictions for the elections used for algorithm training.

4.1 Volume-based forecasts

Volume-based forecasts assume a direct connection between relative message volumes for candidates and their performance at the polls. Following [4], we construct a measure of Twitter attention T_R as the ratio of the Republican message volume V_R to the total message volume $V_R + V_D$, as shown in equation 1.

$$R = \frac{T_R}{T_R + T_D} \qquad (1)$$

We then model election outcomes with OLS as specified in equation 2. Republican performance for an election at time t is modeled as a function of the Twitter proxy at t and the prior Republican vote share in that district V_{t-1}. This explicitly models incumbency separately from the contemporaneous, election-specific data derived from Twitter.

$$V_{R,t} = V_{R,t-1} + R \qquad (2)$$

Regression results are shown in table 3. Consistent with [4], we find that R remains significant even when explicitly accounting for incumbency, though its magnitude declines substantially.[4] Nevertheless, this estimator suffers the same regression to a weak incumbency signal as the n-gram model discussed in section 3. When back-casting the 2010 elections, the volume-based forecast beat the rate of incumbent re-election. But when used to forecast the 2012 election

---

[4]We note that we do not use the same sampling method as they do, and so the results here may not be directly comparable. However, the substantive conclusion of the regression remains the same: a candidate's share of Twitter mentions in a race remains significant even when conditioning on a measure of incumbency. Furthermore, the prior Congressional vote share is arguably a stronger measure of incumbency than prior Republican Presidential candidate performance.

using contemporaneous data, that forecast performed substantially worse than a simple incumbency heuristic. Figure 5 illustrates the performance degradation. Incumbency outperforms all model variants in 2012. Furthermore, forecast under-performance was worst for those elections that we care most about: those right around the 50% cutpoint. For races decided by a spread of 10 points or less (that is, where one candidate's share of the two-party vote was in the interval $(45, 55]$ percent), the complete model forecast only 63% of the races correctly in 2010, and only 53% in 2012. Finally, the estimator appears to gain little from the Twitter data itself. An OLS model trained without R performed nearly as well as the fully-specified model in both 2010 and 2012. This is true whether measured by the RMSE error for forecast vote share, or the binary win / loss accuracy rate.

Table 3: Regression table for a model of form $V_t \backsim V_{t-1} + R$. Current and prior vote share use the share of the two-party vote. Twitter ratio is defined as $R = \frac{T_R}{T_R + T_D}$ for Twitter message volumes $T_p, p \in Republican, Democrat$.

	Complete	Vote only	Twitter only
Intercept	21.99*	23.59*	38.92*
	(1.27)	(1.15)	(1.75)
Twitter ratio	5.39*		21.84*
	(1.85)		(2.89)
Prior vote	0.62*	0.65*	
	(0.03)	(0.02)	
N	298	298	298
R^2	0.71	0.70	0.16
adj. R^2	0.70	0.70	0.16
Resid. sd	9.20	9.31	15.51

Standard errors in parentheses
* indicates significance at $p < 0.05$

4.2 Naïve sentiment analysis

Finally, we implement a version of naïve sentiment analysis as a forecasting proxy. Earlier studies [11, 15] employed relatively simple sentiment analysis to generate either polling proxies or predictive measures for campaign outcomes. Here we use the OpinionFinder sentiment corpus [17] to assign sentiment scores to each candidate's tweets. Scores are computed as the sum of positive (+1) and negative (-1) OpinionFinder adjectives. A candidate's aggregate sentiment score is defined as $S = \frac{pos}{pos+neg}$. For a two-candidate campaign, we define the campaign sentiment ratio as $Sentiment = \frac{S_R}{S_R + S_D}$. Using this metric, we fit a regression of the form:

$$V_{R,t} = V_{R,t-1} + Sentiment \qquad (3)$$

Table 4 summarizes the regression in its fully-specified and component forms. We see that both prior vote share and sentiment return significant predictors of the two-party vote share. Once again, we see that the Twitter-based proxy remains significant in the regression specification even when explicitly modeling incumbency, though again its magnitude declines substantially. However, those results do not translate into accurate forward-looking predictions. Figure 6 summarizes both the win/loss accuracy and RMSE vote-share error for all model specifications. We see that while

back-casting the 2010 election could beat the baseline rate of incumbent re-election, forecasting 2012 performed somewhat worse. Moreover, the sentiment proxy provided no added predictive power: the model that used only vote share to forecast 2012 performed as well as the complete model. Conversely, a sentiment-only model performed substantially worse, and failed to beat the incumbency baseline either when back-casting or forecasting. Finally, performance was once again worst for the most contested races: for races decided by a spread of 10 points or less, the full model forecast only 65% correctly.

Table 4: Regression table for a model of form $V_t \backsim V_{t-1} + Sentiment$. Current and prior vote share use the share of the two-party vote. $Sentiment = \frac{S_R}{S_D + S_R}$, where $S_p = \frac{pos_p}{pos_p + neg_p}$ for $p \in \{R, D\}$.

	Complete	Twitter only	Vote only
Intercept	13.79*	44.12*	17.04*
	(1.65)	(2.83)	(1.23)
Prior vote	0.76*		0.77*
	(0.02)		(0.02)
Sentiment	7.09*	15.96*	
	(2.43)	(5.18)	
N	266	266	266
R^2	0.79	0.03	0.78
adj. R^2	0.79	0.03	0.78
Resid. sd	7.33	15.70	7.43

Standard errors in parentheses
* indicates significance at $p < 0.05$

5. DISCUSSION

These results add further weight to the argument that simplistic measures of political sentiment or intent in Twitter traffic will not suffice as valuable election forecasts. Each of the methods discussed here generated promising results when back-casting elections. None of them provided useful predictions for true out-of-sample forward-looking forecasts. Forecasts were particularly inaccurate for elections decided close to the 50% win/loss cutpoint. These results occurred despite a political system, the U.S. House of Representatives, with short intervals between elections, and in which adjacent elections are often fought over similar issues.

These results recommend against simplistic election forecasts with Twitter. None of the methods used here made vigorous attempts to account for the demographic or partisan differences between the Twitter universe and the voting public. Nor did they attempt to account for changes to that universe itself. Instead, they all sought to find a useful mapping between a snapshot of that universe, taken at one election, and actual election outcomes. The results here suggest that both elections and the Twitter universe more generally are sufficiently unstable as to quickly render such maps invalid.

6. CONCLUSIONS

We have provided three tests of heretofore promising approaches to forecasting U.S. House of Representatives elections with Twitter. Real-time forecasts based on n-gram patterns illustrated the degradation of model performance
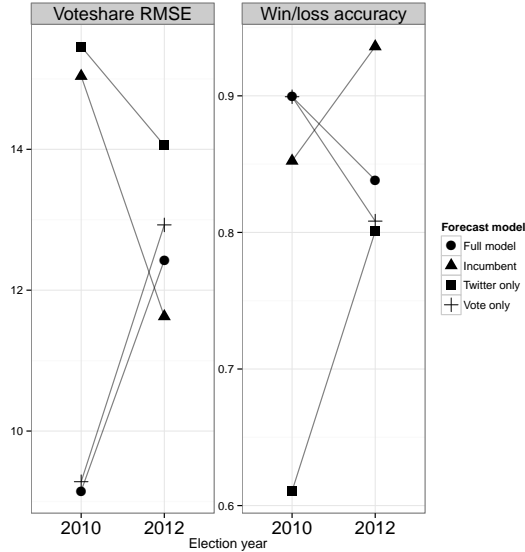
Figure 5: Summary of volume forecast performance. This figure summarizes the sentiment forecast performance in 2010 and 2012. In all cases, the incumbency-based forecast performed at least as well as the Twitter-based forecasts. For predicting the winner alone, incumbency out-performed all other models.
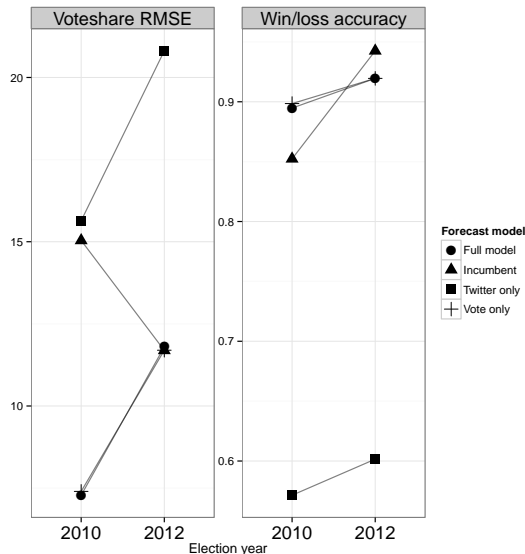

Figure 6: Summary of sentiment forecast performance. This figure summarizes the sentiment forecast performance in 2010 and 2012. In all cases, the incumbency-based forecast performed at least as well as the Twitter-based forecasts. For predicting the winner alone, incumbency out-performed all other models.

between election. Use of the data gathered for that experiment to test volume- or sentiment-based forecasts showed that the same thing was true of those methods. Examination of the data itself showed that Twitter data tends to reproduce known biases towards incumbents in the U.S. political system. Any predictive power above and beyond simply predicting that the incumbent will win thus appears to come from over-fitting to ephemeral phenomena unique to single elections. Real success at using social media to forecast general political behaviors thus appears to require much greater effort to detect and account for demographic, political, and other differences between Twitter users and the broader polity; and to do so continuously as both populations evolve and change. Whether, after having done so, Twitter will fulfill its promise as a simpler alternative to traditional polling, remains unclear.

7. REFERENCES

[1] Adam Bermingham and Alan F Smeaton. On using twitter to monitor political sentiment and predict election results. *Sentiment Analysis where AI meets Psychology (SAAIP)*, page 2, 2011.

[2] Devin Caughey and Jasjeet S Sekhon. Elections and the regression discontinuity design: Lessons from close US House races, 1942–2008. *Political Analysis*, 19(4):385–408, 2011.

[3] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer. Political polarization on twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[4] Josepha DiGrazia, Karissa McKelvey, Johan Bollen, and Fabio Rojas. More tweets, more votes: Social media as a quantitative indicator of political behavior. Working paper, Indiana University, 2013.

[5] Andrew Eggers, Olle Folke, Anthony Fowler, Jens Hainmueller, Andrew Hall, and James Snyder. On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *Available at SSRN 2265625*, 2013.

[6] Daniel Gayo-Avello, Panagiotis T Metaxas, and Eni Mustafaraj. Limits of electoral predictions using twitter. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM) 2011, July 17*, volume 21, page 2011, 2011.

[7] Andreas Jungherr, Pascal Jürgens, and Harald Schoen. Why the Pirate Party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T.O., Sander, P.G., & Welpe, I.M. "Predicting Elections with Twitter: What 140 characters reveal about political sentiment". *Social Science Computer Review*, 30(2):229–234, 2012.

[8] Panagiotis Takis Metaxas, Eni Mustafaraj, and Daniel Gayo-Avello. How (not) to predict elections. In *2011 IEEE third international conference on social computing (SocialCom)*, pages 165–171. IEEE, 2011.

[9] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosenquist. Understanding the demographics of twitter users. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSMâĂŹ11), Barcelona, Spain*, 2011.

[10] Fred Morstatter, Jurgen Pfeffer, Huan Liu, and Kathleen M Carley. Is the sample good enough? comparing data from twitterâĂŹs streaming api with twitterâĂŹs firehose. *Proceedings of ICWSM*, 2013.

[11] B. O'Connor, R. Balasubramanyan, B.R. Routledge, and N.A. Smith. From Tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, pages 122–129, 2010.

[12] Eric C. Polley and Mark J. van der Laan. Superlearner. Working paper, Divison of Biostatistics, University of California Berkeley, Berkeley, CA, 2010.

[13] Erik Tjong Kim Sang and Johan Bos. Predicting the 2011 dutch senate election results with twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pages 53–60. Association for Computational Linguistics, 2012.

[14] A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpe. Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review*, 2010.

[15] W. van Atteveldt, J. Kleinnijenhuis, N. Ruigrok, and S. Schlobach. Good News or Bad News? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology & Politics*, 5(1):73–94, 2008.

[16] M.J. Van Der Laan, E.C. Polley, and A.E. Hubbard. Super learner. *Statistical applications in genetics and molecular biology*, 6(1):25, 2007.

[17] Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. Opinionfinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 34–35. Association for Computational Linguistics, 2005.