ON USING SOCIAL MEDIA TO PREDICT ELECTIONS. A META-ANALYSIS

by

Brent Pretty

B.A., Vancouver Island University, 2014

A Major Research Paper

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Digital Media

in the

Yeates School of Graduate Studies

Toronto, Ontario, Canada, 2015

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A MRP

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

Brent Pretty

ABSTRACT

As of late there's been great interest in social media’s ability to predict elections. Platforms such as Twitter and Facebook, owing to their cultural ubiquity, offer a plethora of data and an opportunity to track public perception at a granular level in real time.

The ability to passively analyze public opinion is a massive step forward in the realm of political prediction, and has the potential to redefine the field of campaign strategy.

In this piece of research I analyzed the current state of social media based electoral prediction. I examined the methods and techniques used to collect and analyze data, and compared their results against both each other and other methods of prediction such as telephone polling.

In this I found a field that is still in its infancy. Much work remains to be done until a set of best practices surrounding social media based electoral prediction are accumulated.

ACKNOWLEDGEMENTS

I would like to thank the entire faculty of the MDM program, especially Sonya, Michael, and Matt, for their hard work in creating a truly innovative academic experience. I'd like to thank my faculty supervisor Dr. Greg Elmer for helping guide me through this process, and lastly I'd likely to thank my fellow students for all of the inspiration of the last year.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

As of late there's been great interest in the predictive power of social media data. Platforms such as Twitter and Facebook, owing to their cultural ubiquity, offer a plethora of data and an opportunity to track public perception at a granular level in real time. Over the last decade researchers have begun examining the predictive power of this data by integrating it into a variety of novel scenarios.

These range from tracking wikipedia activity in order to predict box office success (Meysyan, Yasseri, & Kertesz, 2013), creating a Minority Reportesque crime prediction model based on Twitter event data (Wang, Gerber, & Brown, 2012), to general activity prediction (Weerkamp & Rijke, 2012).

The broad interest in this topic reflects the reality of our information driven society. In an information economy the verisimilitude is as important as shipping products. Instruments such as stock market futures serve as evidence that speculation and prediction are ingrained within the very fabric of our economic system. Thus the accuracy of predictions, or at least a belief in their accuracy, is fundamental to the success of a hyper growth focused economy.

This appetite for prediction extends across all reaches of society, but perhaps the most concerned is the political sphere. Election campaigns are fueled on momentum. They stake out any claim to legitimacy or credibility as a means of gaining a critical mass of votes. This may mean a number of things. It may mean attempting to attract undecided voters, it could mean siphoning votes from an opposing candidate, or it could very well just simply mean getting your voters out to vote.

The ability to passively analyze public opinion is a massive step forward in the realm of political prediction, and has the potential to redefine the field of campaign strategy.

So how does this need for momentum interface with the multitudes of proposed predictive models? And is it beneficial for campaigns to adopt these approaches? Are social media based predictive models able to accurately predict elections? And if so, are they able to do it better than the current alternatives?

According to Schoen, Gayo-Avello, Takis Metaxas, Mustafaraj, & Strohmair (2013), “Social media provides a fluid, instantaneous, cheap, unstructured way of collecting data at largescale.” (p. 7)

I'll be examining the current strategic utility of the proposed methods as defined by their accuracy in relation to each other. We’ll delve into the claims made by researchers, highlighting both strengths and weaknesses, and we’ll make claims regarding possible future approaches. We'll do this in a manner similar to the meta-analysis performed by Gayo-Avello when he examined the field 3 years ago. (Gayo-Avello, 2012)

RELATED WORK

Social media based prediction has been a topic of considerable interest over the last decade. With the potential to offer low cost highly accurate predictive modeling researchers have spent a great deal of time and effort investigating how to utilize the data available. There are a number of different types of works within this evolving field.

Individual case studies are perhaps the most prominent form of inquiry. Works which fall into this category generally investigate specific elections or events with a focus on assessing the accuracy of individual models. This includes works such as the analysis relating to the 2012 US Primary performed by Lei Shi, Neeraj Agarwal, Ankur Agrawal, Rahul Garg, & Jacob Spoelstra (2012), the case study examining the 2012 US Presidential written by Murphy Choy, Michelle Cheong, Ma Nang Laik, & Koo Ping Shung (2012), as well as a host of others investigating elections around the globe.

Occasionally there are case studies that do not study a particular event, rather the accuracy of a particular tool. An example of this would be the study *A user-centric model of voting intention from Social Media* conducted by Vasileios Lampos, Daniel Preot¸iuc-Pietro and Trevor Cohn (2013), which proposes a predictive model for predicting polling averages during a non-campaign period.

Researchers also use social media data to examine political behaviour at a macro level. Notably, David Garcia, Fernando Mendez, Uwe Serdult, & Frank Schweitzer used social media data to investigate political polarization in online media

(2012). Stefan Stieglitz and Linh Dang-Xuan utilized this approach to propose a framework for social media based political communications (2012).

The predictive utility of social media is being studied not only in terms of electoral politics, but in other fields as well.

In the economic sector researchers are examining how to leverage the reactions and sentiment of the general news media and blogosphere to increase profits in the trading sector. Works such as *Trading Strategies to Exploit Blog and News Sentiment* (Zhang, Skiena, 2010), are pivotal to this movement towards social data based algorithmic trading.

Perhaps one of the most novel and interesting investigations into the predictive power of social media is the attempt to accurately model and predict criminal behaviour in order to better utilize police resources such as the work found in *Predicting Crime Using Twitter and Kernel Density Estimation* (Gerber, 2014).

Expectedly, behavioural scientists are using the multitude of personal data to model personality and modes of behaviour *Predicting Personality from Twitter* (Golbeck, Robles, Edmondson, & Turner, 2011), and *Social Media for Large Studies of Behavior* (Ruths, Pfeffer, 2014) demonstrate some of the most modern thinking in that regard

Social media data has been critical in the evolution of the way epidemiologists track and study the spread of disease. Being able to spot symptoms within the content of social media has proved invaluable in the fight against a variety of harmful pathogens. Works such as *Predicting Flu Trends using Twitter Data* (Achrekar, Gandhe, Lazarus, Yu, & Liu, 2011), have been the cornerstone in this effort.

The utlity of being able to predict the life cycle of a particular product is obvious, and that’s exactly what researchers utilized advanced sentiment analysis techniques to determine in *Prediction of product life cycles using Twitter data* (Prick, 2014).

In relation to this paper, the most influential work would be that of Daniel Gayo-Avello. His work generally features broad analysis pertaining to the state of social media based prediction. The meta-analysis authored by him in 2012 compared a collection of case studies covering elections from all over the world. In this meta-analysis he attempts to provide an accurate overview of what was then the current state of social media based electoral prediction (2012). Other works of his have covered the limits of social media based electoral prediction (2011), as well as an examination of the predictive power of social media in other areas (Schoen et al., 2013).

GENERAL CHARACTERISTICS OF SOCIAL MEDIA BASED PREDICTIVE ELECTORAL MODELS

Overview

Researchers have experimented with a wide variety of models to varying degrees of success. In his meta-analysis Gayo-Avello (2012) highlighted what was then the current state of the field. “Fortunately, as it will be shown, only two “flavors” of voting inference in Twitter have been widely used —namely, tweet counts and lexicon-based sentiment analysis– and, thus, a number of papers have evaluated and compared both. (p. 3).”

This remains true today. In the time since, the methods and algorithms utilized by researchers have become more sophisticated with many, such as Bermingham’s model for the 2011 Irish election (2011), now employing a hybrid model combining both volume and sentiment.

Every analysis involves a number of common methods. In the case of volume/sentiment based models some of these shared characteristics include:

- A defined time period for collection.
- Defined keywords/hashtags/search terms.
- A method of data cleansing or de-noising.

Generally the overall purpose of a model is to quantify the amount of support a candidate/party is receiving within a particular time frame in a particular area. The manner in which this is accomplished varies wildly. Researchers may decide to only examine tweets which contain hashtags identified by the researcher, as Jungherr did when he examined the 2009 German election (2013), or they may decide to perform a broader search of keywords, such as the methodology proposed by Razzaq et al. to examine the 2013 Pakistan election (2014).

As social media platforms are global when attempting to predict regional opinion it is necessary for researchers to filter their data based on geographical location. Choy et al. (2012) utilized data from the location field for this purpose, developing an algorithm which searched only for tweets whose author had declared their location as the community which the researchers were examining.

We can examine the differences in methods within the context of a hypothetical scenario.

We have the candidate (C1), who is running for position (P1), under the parties (P2) banner.

In an analysis which only examines hashtags researchers are looking for direct conversation pertaining to the candidate, position, and party. They'll likely look for things such as #c1, #c1forP1, #P2, #c1p2, and any and all combination.

In a keyword analysis, a broader search is conducted. Researchers no longer look solely at hashtags, but also look at the content of the posts themselves. In this scenario the mere mention of a keyword such as c1 or any of its derivatives will be included within the model, rather than just the hashtagged versions. This method may be used as a means to uncover more of the peripheral conversation.

Period Of Collection

Period of collection refers to the specific time period from which data was created and retrieved. Researchers utilize platforms API's in order to accumulate vast quantities of data in real time.

In all referenced case studies the data was collected in real time. With most cases preceding an election. The only notable outlier being the work of Lampos et al. which was a study measuring the accuracy of opinion polling, and did not occur during an election period. (Lampos et al. 2013).

The length of the period of collection amongst the referenced case studies varies dramatically. Some case studies utilized data collected from a time period of days, while others collected data over the course of a number of years. Most studies collected tweets for a period which lasted up to a maximum of six months, as seen in Fig. 1.

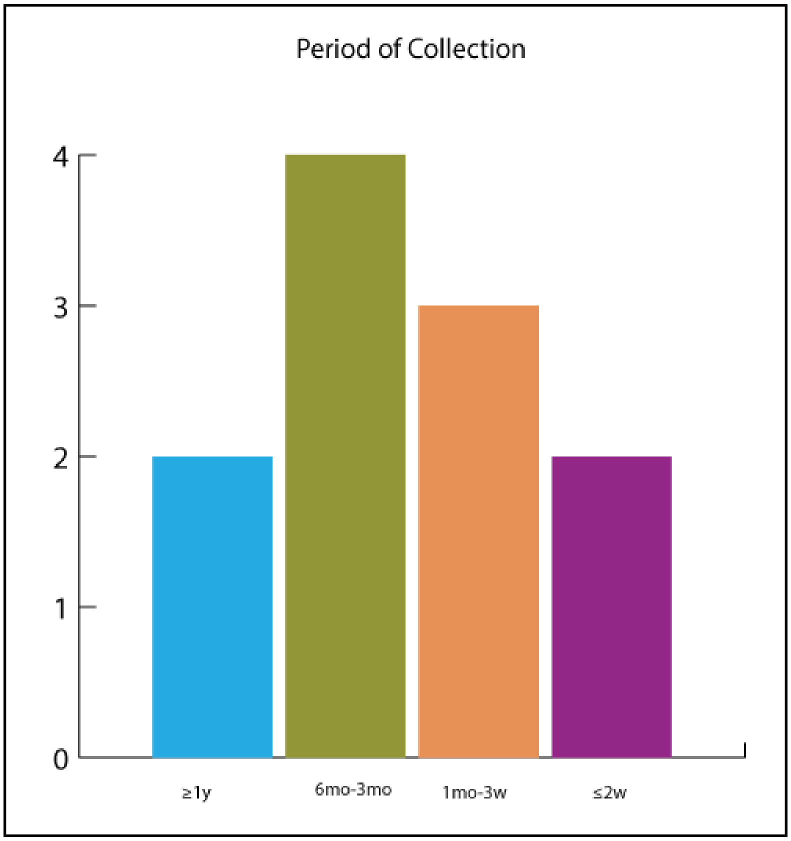

Figure 1. Examining the period of collection across referenced case studies.

In their work examining the 2012 US Presidential Primaries Lei Shi, Neeraj Agarwal, Ankur Agrawal, Rahul Garg, & Jacob Spoelstra collected 300 million tweets through daily requests between September 2011 and February 2012. (Shi et al. 2012.)

Conversely, in Eric Sanders & Antal van den Bosch's paper *Relating Political Party Mentions on Twitter with Polls and Election Results,* the collection period was a mere 10 days. (Sanders, Bosch, 2013)

It appears through taking a cursory glance at the results of the referenced case studies that there appears to be no direct link between length of data collection and accuracy of prediction. In fact the study with the longest collection period, the US

Primary investigation conducted by Shi et al., was of the worst performing. This doesn't preclude the notion that other studies with different models may benefit from a longer collection period, but there is as of yet no evidence to definitively say that they would.

CASE STUDIES

Authors	Election/Event	Period of Collection	Type of Analysis	Findings
Lei Shi, Neeraj Agarwal, Ankur Agrawal, Rahul Garg, Jacob Spoelstra	2012 US Primary Elections	September 2011 - February 2012	Volume Based, Sentiment Based, Lexicon Based Machine Learning.	Predictions based on volume with or without sentiment analysis are inaccurate. It is possible to predict elections with more sophisticated algorithms and machine learning techniques
Murphy Choy, Michelle Cheong, Ma Nang Laik, Koo Ping Shung	2012 US Presidential Election	August 2012 - October 2012	Census Corrected Sentiment Based.	Proposed model shows promise. Predicted results with a MAE of 2.60%
Muhammad Asif Razzaq, Ali Mustafa Qamar, Hafiz Syed Muhammad Bilal	2013 Pakistan Election	N/A	Sentiment Based	Difficult to correlate sentiment with election results.
Vasileios Lampos, Daniel Preot ̧iuc-Pietro and	General Opinion Polling in UK and Austria	March 2010 - February 2012 (UK)	Volume Based Corrected. Multi-task	+/- 1.5% MAE when it comes to predicting polls.

Trevor Cohn		January 2012 - December 2012 (Austria)	Learning.	
Adam Bermingham and Alan F. Smeaton	Irish General Election, 2011.	February 8th - February 25th 2011	Hybrid Sentiment & Volume.	+/- 3.7% MAE
Andreas Jungherr	2009 German Federal Election	Mid June - Early October 2009	Hashtag volume.	The volume of usage a particular hashtag receives is not predictive of a party's/candidates vote share.
Erik Tjong Kim Sang and Johan Bos	2011 Dutch Senate Elections	February 16th 2011	Volume Based. Sentiment Based.	A wide swing in terms of MAE. 17% in some cases.
Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe	2009 German Federal Election	August 13th - September 19th 2009	Sentiment Based. Volume Based.	Raw volume based model was able to predict election result with a MAE of +/- 1.65%

Marko Skoric, Nathaniel Poor, Palakorn Achananuparp, Ee-Peng Lim, Jing Jiang	2011 Singapore General Election	April 1st 2011 - May 7th 2011	Volume Based.	Not incredibly accurate with a MAE of +/- 5.3%
Eric Sanders, Antal van den Bosch	2012 Netherlands Parliamentary Elections	September 2nd 2012 - September 12th 2012	Volume Based.	MAE of cumulative tweets +/- 2.2%
Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams	2015 UK General Election	November 28th 2014 - March 9th 2014	Sentiment Based.	Ultimately their prediction was completely inaccurate.

Table 1

TYPES OF ANALYSIS

Volume

Volume based models range from rudimentary, simply counting the number of mentions a particular candidate/party has regardless of age, geography, or any other factor, to advanced, where all data is cleansed to account for demographic details such as age and location as well as to eliminate spam and false positives.

Gayo-Avello explains the popularity of the method as such (2012) “Such a method is appealing for many reasons: it is easy to implement, it can be applied in near real-time, and it can be used both to obtain aggregated vote rates and to infer voting intentions for individuals (i.e. the candidate a user is mentioning the most would be his or her chose).” (p.10)

In a basic volume centric predictive model if c1 receives 6 mentions and c2 receives 4, then c1 would be represented as 60% and c2 as 40%. While appearing to be quite rudimentary, this type of analysis has shown promising results. Andranik Tumasjan who used a volume centric approach in his prediction of the 2009 German election (2010), used a volume centric method with notable success (1.65% MAE).

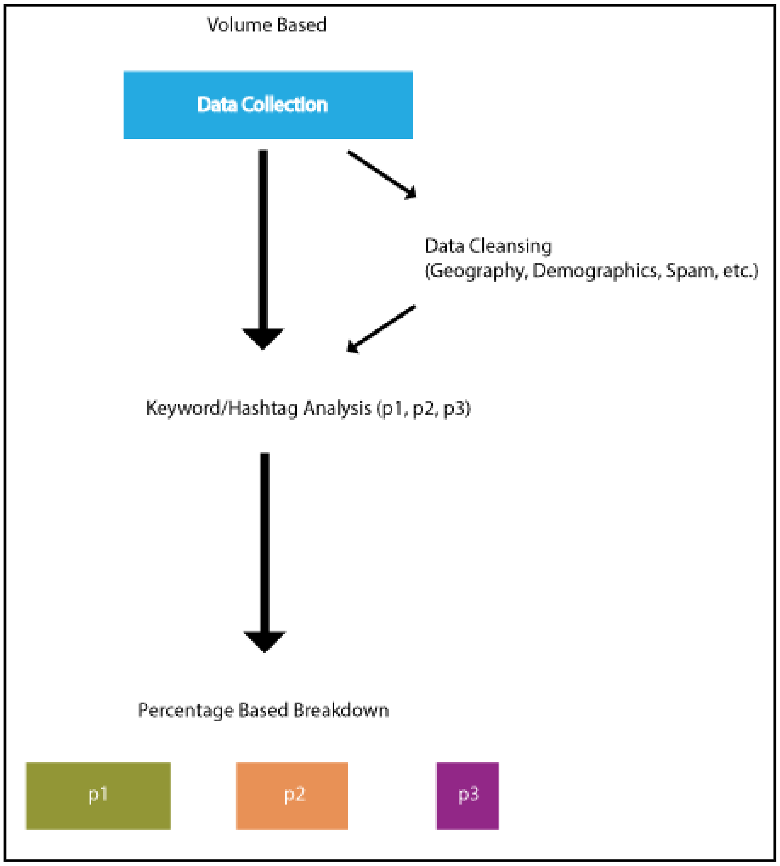

Figure 2. Basic flow-chart of a volume based prediction model.

Sanders and Bosch also used a volume centric method (2013) to predict the 2011 Netherlands elections. Their results were similar to that of Tumasjan, and they managed to predict the outcome within a 2.2% MAE. While this was still considerably worse than traditional polling (1.7% MAE), it adds to the evidence that merely counting mentions is enough to give a somewhat accurate view of the electorate.

Others, such as Skoric et al. have been unable to replicate this success (2012), instead their use of a volume centric model to predict the 2011 singapore elections resulted in a MAE of 5.3%.

One obvious weakness in a volume based novel is the inability to account for the intention of the user. Volume based models make an assumption that all mentions are equal as a means of prediction.

For instance, if there were two tweets with the first stating “c1 is the best ever”, and the second stating “c1 is the worst ever”, both would be equally weighted within the prediction.

Sentiment

Currently the most popular method of prediction is lexicon-based sentiment analysis. These models are inherently more complex than volume based models, attempting to categorize mentions by determining whether the underlying sentiment is positive or negative based upon predetermined lists (lexicons) of words categorized as such. In a sentiment based model if c1 was mentioned 10 times then their rating would be based on the ratio of positive to negative interactions. If 8 mentions were of a positive nature and 2 were of a negative then they would have a high ratio resulting in them having a higher chance of being elected, whereas if the inverse were true then even if the number of mentions overall was the same, but most were negative they would have a decidedly lower probability of being elected according to the model.

Pedro H. Calais Guerra, Adriano Veloso, Wagner Meira Jr., & Virgílio Almeida (2011) explain the history of this method, “sentiment analysis (also known as opinion mining) algorithms have been designed for static and well-controlled scenarios…” they go on to outline that within the confines of user reviews and other forms of static content traditional methods such as, “pre-defined lists of positive and negative words (i.e., lexicons) and traditional supervised machine learning techniques have been quite successful.” Furthermore they argue that “current state-of-the-art sentiment analysis strategies are not effective for mining opinions in this new, challenging environment.” The new challenging environment being real time social media feeds.

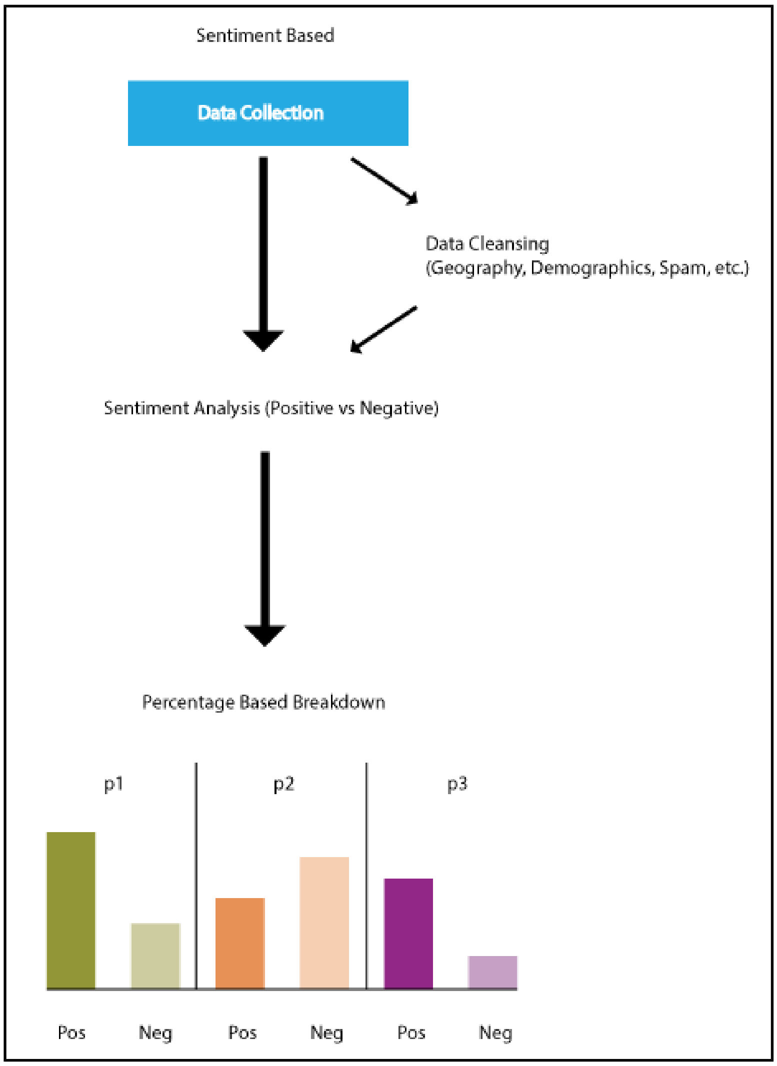

Figure 3. Basic flow-chart of a sentiment based prediction model.

In their analysis of the Irish election of 2011, Bermingham and Smeaton (2011) concluded that “Twitter does appear to display a predictive quality which is marginally augmented by the inclusion of sentiment analysis.” (p.8) When compared to other methods however, sentiment seemed to fall short, they explained that “volume is the single biggest predictive variable followed by inter-party sentiment. Given sufficient data, intra-party sentiment appears to be less valuable as a predictive measure.” (p.8)

A more recent example of sentiment based prediction would be the work of Razzaq et al. and their examination of the 2013 Pakistan election (2014). They make note of the challenges of such a work by stating that, “This work of forecasting elections for developing countries especially for Pakistan is of unique attempt” and “We also ignored on ground realities of not considering major part of population living in rural areas with tribal loyalties.” (p.703) These realities are not something most researchers have to take into account so they’re worth noting here. Ultimately their prediction was of questionable accuracy. While they were able to show that the ratio of positive to negative sentiment accurately predicted the place of the parties, they were largely unable to predict specific percentages or seats.

Discussion

Weaknesses of Methods

That each of these methods has weaknesses is indisputable. What these weaknesses are, however, largely depend on the context. For instance if you merely wish to know if a party is going to win an election then the sentiment based technique developed Razzaq et al. (2014) will likely provide you with that information, but if you wished to have a more detailed breakdown of the election results, including percentages and seat distributions, then this method would not be able to provide you with that without further development.

Some major weaknesses that stand out in regards to sentiment analysis are;

1. Inability to perceive intent of language (sarcasm).
2. Inability to understand language outside of lexicon.
3. Inability to provide accurate breakdown without further analysis.

As noted above, most of the major problems with sentiment analysis is its inability to understand and decipher the nuances of natural language. This is especially important when it comes to accounting for regional variations in popularity. If an algorithm is unable to decipher the subtle variations of regional dialects then it is unable to provide an accurate results.

For example my home province, Newfoundland & Labrador, features a dialect that is very heavy on sarcasm. One way in which I could write a tweet in that dialect to express dissatisfaction with a candidate might look something like “Yes b’y! c1 wants more turtles.”

Now any normal lexicon based sentiment analysis algorithm would likely classify that tweet as positive due to the inclusion of the word yes. Even if you were to include an operator which looked for the combination of Yes + b’y as an indicator of sarcasm the algorithm would still be wrong, because it can also act as a means of expressing agreement. Now this is only one example of one dialect. The added complexity of accounting for the linguistic diversity of an entire country makes creating an accurate sentiment analysis algorithm an extremely arduous task.

It’s also difficult to know when exactly to incorporate sentiment. Bermingham et al. (2011) found the following:

> It is difficult to say how best to incorporate sentiment. On the one hand, sentiment distribution in the tweets relevant to a single party is indicative of the sentiment towards that party. For example, if the majority of the mentions of a party contain negative sentiment, it is reasonable to assume that people are in general negatively disposed towards that party. However, this only considers a party in isolation. If this negative majority holds true for all parties, how do we differentiate public opinion towards them? In a closed system like an election, relative sentiment between the parties perhaps has as much of an influence. (p.4)

One large weakness of the field as a whole is the inability to accurately determine a user's location. Skoric et al. (2012) explain that "Given the complexity of electoral systems and arbitrariness of many electoral boundaries, paired with imprecise geolocation data supplied by Twitter users, it seems futile to make specific predictions." (p.2589) They go on to argue that "if certain conditions are met then the analysis of Twitter messages could represent an inexpensive, unobtrusive and reasonably accurate method for gauging political sentiment. However, suitable theoretical frameworks need to be developed in order to fully understand the processes behind public opinion formation on Twitter." (p.2590)

Defining Success

When it comes to using social media to predict elections how do we define success? Do we examine the results from these studies in a vacuum against each other? Or do we put it in the context of the larger ecosystem and compare the results against more traditional methods of prediction such as telephone based opinion polling.

Since one goal of this paper is to examine the strategic utility of these models, it would be useful to compare these models not only to each other in a vacuum but to also examine the way they fit into the broader strategic landscape by examining their efficacy and accuracy by comparing them with methods which have already found widespread adoption such as telephone and internet polling.

The 2012 US Presidential election will serve as our benchmark for measuring the efficacy of traiditional polling methods. The primary motivation for using this specific election is the vast amount of polls taken by a wide variety of organizations. This large sample size should be representative of the field as a whole. We'll examine polls from the weeks immediately preceding the election, as well as polls from months prior.

Our method of examining the efficacy of telephone polling will be to calculate the MAE of each telephone poll as compared to the actual results of the election. Because the MAE is a measure of deviation from the actual results we can use it to assess any sort of predictive model. Therefore it is perfect for comparing telephone polling to social media based models.

The results used to calculate the MAE of these telephone polls were; Obama 51.1%, and Romney 47.2%

Publication	Date	Prediction	MAE
NBC News/Wall Street Journal Survey	November 1st -3rd, 2012	Obama 48% Romney 47%	3.3%
Public Policy Polling	November 3rd, 2012	Obama 50% Romney 47%	1.3%
CNN/Opinion Research	November 2nd - 4th, 2012	Obama 49% Romney 49%	3.9%
Pew	October 31st - November 3rd	Obama 50% Romney 47%	1.3%

YouGov	October 31st - November 3rd	Obama 48.5% Romney 46.5%	3.3%
Associated Press	May 3rd - 7th, 2012	Obama 51% Romney 41%	6.3%
NBC News/Wall Street Journal Survey	May 16th - 20th, 2012	Obama 47% Romney 43%	8.3%
CNN/Opinion Research	April 13th - 15th, 2012	Obama 52% Romney 43%	5%

Table 2. List of 2012 US Presidential Election Phone Polls

By just taking a cursory glance at the results we can see that time has a significant effect on the accuracy of the prediction. In the random sampling of polls listed above those which were conducted in the weeks immediately preceding the election had a combined MAE of 2.62%. Conversely those which were conducted months prior to the election had a combined MAE of 6.53%.

The results show that in the weeks immediately preceding an election opinion polling is reflective of public sentiment, and is ultimately predictive in nature. The results also show that the accuracy of a poll is wholly dependant on when it was conducted. While every poll examined managed to accurately predict the winner, any poll which was conducted more than a month before the election was not remotely close to being able to predict the final result.

Based on these findings we can define success as a result which has a lower MAE than the MAE of the aggregated phone polling. Success can also be defined by predicting the correct result, but ultimately this is less helpful.

Can Social Media Predict Elections?

Taking the conditions set out in the last section we can begin to assess the case studies, and answer the question can social media predict elections?

First, we can make note of something quite important in regards to the traditional polls referenced in the previous section. While there was variation amongst traditional pollsters generally each were fairly accurate and the MAE was consistent amongst each other. This could be due to a number of reasons, but as a whole it can be attributed to the maturity of the practice and the pollsters ability to create a truly representative sample.

Anstead and O'Loughlin (2014) argue that the accuracy of these models can be compared to traditional polling methods, and that as of now they pale in comparison. They note that there are a number of reasons for this, but primarily it stems from the inability to create a representative sample from social media data. They are not pessimistic in their view though, and comment that they believe in the future social media based prediction will supplant traditional opinion polling as the de facto standard.

In comparison to the homogeneity found within the results of traditional polling methods the MAE amongst social media based predictions was all over the map, for example Tjong et al. ended up with a 17% MAE when they attempted to predict the 2011 Dutch elections (2012). While Tumasjan et al. recorded only a 1.65% MAE when investigating the German elections of 2009 (2010).

This lack of consistency is incredibly important. It speaks to the lack of maturation within the field. There also seems to be a lack of continued investigation into specific models. Using the same model over multiple elections would provide a considerable amount of information relevant to determining the strategic utility. In each of the case studies investigated as part of this research none used their model in a secondary scenario in order to further validate their results.

Metaxas and Mustafarj (2012) highlight the lack of standards as being one of the largest problems facing the field. They argue that researchers have been unable to set forth a protocol which is consistently accurate in its predictions, and that even things such as the period of data collection has yet to be standardized. They also allege that researchers have manipulated their data after the election in order to provide the appearance of accuracy. This is important because while you can use historical data as a means of training a model, ultimately the utility of the model is its ability to predict things in the future. The main purpose of constructing predictive models is to make accurate claims as to what is going to occur, now what already has occurred.

In their work *Limits of Electoral Predictions using Twitter,* Gayo-Avello, Metaxas, & Mustafaraj (2011), argued that their results showed “that data from Twitter did no better than chance in predicting results in the last US congressional elections.”(p.4) Furthermore, they argue that “Predicting elections is something that professional pollsters have been doing for the last 80 years, a mathematically proven application of correctly identifying likely voters and getting an un-biased representative sample of them. Today’s social media do not seem fit to do this.” (p.3)

As outlined in the previous section, save a few choice incidents (e.g., The Dewey Defeats Truman debacle.) professional pollsters have been reasonably accurate through the history of their craft. This is largely because of their ability to accurately create a representative sample. Traditional polling methods adjust their sample to reflect the most probable electorate. They factor age, race, income, and geography into their samples to ensure that their poll represents the most likely outcome. Social media based predictions do not account for this, and are largely unable to collect the data necessary for performing such a task.

Gayo Avello & Mustafarj (2011) explain that social media based models are likely to be unable to rival the complexity of traditional polling methods due to a number of factors. First, that traditional pollsters adjust their samples to fit within what is called a “likely voter model”. They do this by specifically targeting portions of the population in the proportions that they last voted. Hypothetically this could look something like 33% 18-35, 33% 36-50, 33% 50+. They note that obtaining this information through social media is next to impossible, and that this is one of the primary reasons why they believe that traditional polling will remain as the standard for the foreseeable future.

A large reason why traditional polling methods are still used is their ability to be used, with little adjustment, in any election period. While active, interview based, methods such as telephone polling may require a large investment in terms of human capital to perform the required interviews, passive, observation based, methods such as social media require significant investments of time and expertise to be able to adjust algorithms specifically for certain scenarios, and lack the consistency necessary to be trusted.

Sanders and Bosch (2013) argue that the difficulty in determining the authenticity of an account, and the potential for parties to create accounts to astroturf currently represent large problems. They go on to explain that by developing an automatic profiling system which integrates both machine learning and text classification techniques, one could theoretically determine the authenticity of a specific profile and then cleanse their data from the prediction.

Passive methods of prediction allow you to analyze your audience in the abstract. Political strategy however, depends on granular data in order to allocate resources in a manner that is strategically efficient. It seems that passive methods provide an overview of the general mood of the electorate, but this is not necessarily indicative of voting intentions.

This overview is useful for making decisions at a macro level, but does little to provide relevant information to local races. Skoric et al. argue “We suggest that Twitter data may be more suitable for making macro-level assessments of political sentiment than for predicting specific outcomes of local elections which are more volatile and more easily skewed by a few influential Twitterers.” (p.2589)

Essentially Skoric is arguing that social media based prediction fails to offer granular data. Instead it provides a broad overview which may or may not be useful depending on your need.

In an electoral system similar to that of the United States where you have a nationwide two party contest for the presidential a macro view assessment provides an opportunity to see broad trends emerge. In a first past the post parliamentary system similar to Canada’s where local candidates are elected and the governing party is determined by which party managed to win the most seats, the strategic implications of this broader view are dulled by both the regional nature of the electoral system, and the potential for strategic voting. Strategic voting being defined by a constituent voting for a party who is their second choice rather than their first in order to prevent another party from coming to power.

In his work, *A meta-analysis of state-of-the-art electoral prediction from Twitter data,* Gayo-Avello (2012) answered the question whether social media can predict elections by stating that in regards to volume based models “In the absence of further research showing that the method can consistently predict future results for a number of elections outperforming both the incumbency and the past-results baselines, we must conclude that there is no strong evidence to consider it a valid method of prediction.” (p.13) Furthermore, in regards to sentiment based models “In short, results are contradictory. However, taking into consideration that even naïve sentiment analysis seems to outperform the baseline it is clear that further research is needed in that line. This work concludes with some recommendations in that sense.” (p.13)

The same criticisms remain true today. Examining an array of new case studies in conjunction with those studied by Gayo-Avello has shown that little progress has been made in addressing issues such as the lack of consistency, the lack of best practices, or the inability to accurately model likely voters. While some results have shown promise, thus far researchers have been unable to replicate these results or to generate analysis in real time. While we’ve been able to conclusively show that traditional polling methods perform more accurately and with more consistency than social media based models, I would argue that there is evidence to support the claim that social media data can be used to predict elections. I would also argue that we have yet to establish exactly how to do this, but we are beginning to ask the right questions, and to consider new technological approaches.

Annotated Bibliography

Predicting US Primary Elections with Twitter

Lei Shi, Neeraj Agarwal, Ankur Agrawal, Rahul Garg, & Jacob Spoelstra

Shi et al. examined tweets leading up to the 2012 Republican Primaries. They proposed utilizing a sophisticated algorithm which encompasses more than just volume and sentiment as a means of supplanting traditional prediction methodologies.

US Presidential Election 2012 Prediction using Census Corrected Twitter Model

Murphy Choy, Michelle Cheong, Ma Nang Laik, & Koo Ping Shung

Choy et al. proposed a model for predicting elections based on adjusting data by geography based on census data. Their model proved to be accurate in result, but they acknowledged that there is much work to be done in terms of developing a model that is consistent.

Prediction and Analysis of Pakistan Election 2013 based on Sentiment Analysis

Muhammad Asif Razzaq, Ali Mustafa Qamar, & Hafiz Syed Muhammad Bilal

Razzaq et al. investigated the use of sentiment analysis to predict the 2013 Pakistan Election. They noted that while they were able to predict the winners based on the ratio of positive to negative sentiment that in a developing country such as Pakistan other factors, such as tribal loyalties, make it difficult to quantify political opinion.

A User-centric Model of Voting Intention from Social Media

Vasileios Lampos, Daniel Preot¸iuc-Pietro & Trevor Cohn

Lampos et al. conducted a long-term study examining the predictive power of social media in relation to political opinion polling in the UK and Austria. Their results were encouraging as they found that by using data from social media it is possible to accurately model and predict political opinion.

On Using Twitter to Monitor Political Sentiment and Predict Election Results

Adam Bermingham, and Alan F. Smeaton

Bermingham and Smeaton used the 2011 Irish Election as a case study to examine the potential to monitor political sentiment through social media. They discovered that they were able to monitor political sentiment, but that volume of tweets is perhaps the most accurate predictive indicator.

Tweets and Votes, a Special Relationship

Andreas Jungherr

Jungherr looked at the 2009 German Elections as a case study to examine the predictive power of Twitter hashtags. Jungherr found that hashtags were not useful predictive indicators.

Predicting the 2011 Dutch Senate Election Results with Twitter

Erik Tjong, Kim Sang, & Johan Bos

Tjong et al. examined the 2011 Dutch Senate elections and whether social media data would be useful as a predictive tool. They concluded that while social media was accurate to a degree, that further work was required to ensure consistent results and to raise the accuracy to that of current professional standards.

Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment

Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, & Isabell M. Welpe

Tumasjan et al. used the 2009 German election as a case study to examine the predictive power of social media data in a political context. They found that by simply measuring the number of tweets they were able to accurately model the result of the election.

Tweets and Votes: A Study of the 2011 Singapore General Election

Marko Skoric, Nathaniel Poor, Palakorn Achananuparp, Ee-Peng Lim, & Jing Jiang

Skoric et al. examined the predictive power of social media through an analysis of the 2011 Singapore election. They found that while social media data had the potential to be a useful predictive indicator at a macro level, that it did not accurately reflect the realities of local races.

Relating Political Party Mentions on Twitter with Polls and Election Results

Eric Sanders, Antal van den Bosch

Sanders and Bosch looked at the 2012 Dutch Parliamentary elections in order to determine the predictive utility of social media data and found that it was quite accurate. They used a small window preceding the election to collect data, but their results were well within the current accepted range for predictive methods.

140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election

Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, & Matthew Williams

Burnap et al. found that while their method to predict the 2015 UK election was generally methodologically sound there is still considerable work to be done in order to create a consistent and accurate model. This work revolves around geolocation and other sampling issues.

Conclusion and Future Work

In this piece of research I analyzed the current state of social media based electoral prediction. I examined the methods and techniques used to collect and analyze data, and compared their results against both each other and other methods of prediction such as telephone polling.

Through this I found a field that is still in its infancy. Much work remains if social media based predictions are seen in the same light as the methods that are currently employed. Greater focus needs to be paid into accounting for sampling bias and demographics, and a considerable amount of work needs to be completed before a common set of best practices are developed and agreed upon.

As of now social media based predictive models can accurately predict election results at a high level, but questions remain regarding their consistency and their ability to predict low level numbers. Traditional methods such as telephone polling continue to be the most effective and accurate means of gauging user support. At this point any social media based prediction is best used in conjunction with a broader analytical strategy, and not to be depended on as the sole source of information.

The potential for real time granular information pertaining to political opinion is an understandably intriguing prospect to researchers and strategists alike. The ability to target and study specific niches of voters is the next evolution of political strategy, but it’s an evolution that is as of yet in the early stages.

For the future it'll be worth watching other fields for related advances; financial markets, crime detection, and epidemiology offer novel approaches to prediction, while dealing with similar issues pertaining to granular data. Any advance within those fields should be closely analyzed for its potential to work within the context of electoral prediction.

Advances in sentiment analysis will be the key to evolving the field. The development of a contextual sentiment analysis tool, or any advancement in natural language processing, has the potential to radically alter the methods used to predict elections. The next step in research will be to closely investigate the potential of integrating more advanced machine learning techniques, and to consider new ways of defining popularity rather than pos v neg.

Closer attention will have to be paid to the supporting theory of these analysis in order to better define success. We must know why these models work or fail, and we need to be able to replicate results. Consistency should be the focus going forward, as it seems we are beginning to understand what tools should be used, but just not exactly why we are using them.

References

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S., & Liu, B. (2011). Predicting Flu Trends using Twitter data. *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*.

Bermingham, A., & Smeaton, A. (2013). On Using Twitter to Monitor Political Sentiment and Predict Election Results.

Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2015). 140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election. *SSRN Electronic Journal SSRN Journal*.

Choy, Murphy, Michelle Cheong, Ma Nang Laik, and Koo Ping Shung. "US Presidential Election 2012 Prediction Using Census Corrected Twitter Model." (n.d.): n. pag. 11 Nov. 2012. Web.

Garcia, D., Mendez, F., Serdült, U., & Schweitzer, F. (2012). Political polarization and popularity in online participatory media. *Proceedings of the First Edition Workshop on Politics, Elections and Data - PLEAD '12*.

Gayo-Avello, D. (2013). A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. *Social Science Computer Review,* 649-679.

Gayo-Avello, D., Metaxas, P., & Mustafaraj, E. (2011). Limits of Electoral Predictions using Twitter.

Gerber, M. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems,* 115-125.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting Personality From Twitter. *2011 IEEE Third Inernational Conference on Social Computing (SocialCom),* 149-156.

Jungherr, A. (2013). Tweets and votes, a special relationship. *Proceedings of the 2nd Workshop on Politics, Elections and Data - PLEAD '13*.

Lampos, V., Preot¸iuc-Pietro, D., & Cohn, T. (2013). A User-centric Model of Voting Intention from Social Media. *In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL),* 993-1003.

Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data. *PLoS ONE*.

Prick, R. (2014). Prediction of product life cycles using Twitter data.

Razzaq, M., Qamar, A., & Bilal, H. (2014). Prediction and analysis of Pakistan election 2013 based on sentiment analysis. *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014),* 700-703.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science,* 1063-1064.

Sanders, E., & Bosch, A. (2013). Relating Political Party Mentions on Twitter with Polls and Election Results. *Proceedings of the 13th Dutch-Belgian Workshop on Information Retrieval,* 68-71.

Schoen, H., Gayo-Avello, D., Metaxas, P., Mustafaraj, E., & Strohmaier, M. (2013). The power of prediciton with social media.

Shi, L., Agarwal, N., Agarwal, A., Garg, R., & Spoelstra, J. (2012). Predicting US Primary Elections with Twitter.

Skoric, M., Poor, N., Achananuparp, P., Lim, E., & Jiang, J. (2012). Tweets and Votes: A Study of the 2011 Singapore General Election. *2012 45th Hawaii International Conference on System Sciences,* 2583-2591.

Stieglitz, S., & Dang-Xuan, L. (2012). Social media and political communication: A social media analytics framework. *Soc. Netw. Anal. Min. Social Network Analysis and Mining,* 1277-1291.

Tjong, E., Sang, K., & Bos, J. (2012). Predicting the 2011 dutch senate election results with Twitter. *Proceedings of the Workshop on Semantic Analysis in Social Media*.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.

Wang, X., Gerber, M., & Brown, D. (2012). Automatic Crime Prediction Using Events Extracted from Twitter Posts. *Social Computing, Behavioral - Cultural Modeling and Prediction Lecture Notes in Computer Science,* 231-238.

Weerkamp, W., & Rijke, M. (2012). Activity Prediction: A Twitter-based Exploration.

Zhang, W., & Skiena, S. (2010). Trading Strategies to Exploit Blog and News Sentiment.