The University of Texas at Austin
Center for Identity

Using Twitter to Predict the United States Presidential Election

T. Huang
Razieh Nokhbeh Zaeem
K. Suzanne Barber

UTCID Report #20-09

June 2020

Empty Vessels Make the Most Noise: Using Twitter to Predict the United States Presidential Election

Social media has become an essential aspect of our life, and we are used to expressing our thoughts on these platforms. Using social media as an opinion finder has become a popular measure. For any topic that the public opinion matters, there is the potential of using social media to evaluate the problem. Presidential election definitely falls into this category. Previous researches have proven the effectiveness of using social media such as Twitter to predict the outcome of elections. Nevertheless, the composition of social media users can never be the same as the real demographic. What makes things worse is the existence of malicious users who intend to manipulate the public's tendencies toward candidates or parties. In this paper, we aim to increase the predicting precision under the premise that the extracted tweets are noisy. By taking an individual's trustworthiness, participation bias and the influence into account, we propose a novel method to forecast the U.S. presidential election.

CCS Concepts: • **Information systems** → *Sentiment analysis*; *Web and social media search*; *Retrieval effectiveness.*

Additional Key Words and Phrases: Election prediction, social media, sentiment analysis, participation bias

ACM Reference Format:
. 2018. Empty Vessels Make the Most Noise: Using Twitter to Predict the United States Presidential Election. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Since the invention of social media, the way people communicate with each others has been altered drastically. The social media nowadays functions like a mixture of letter, podium, phones, billboard and even provides virtual gathering spaces.The characteristics of low-cost but easily-spreading advertising effects soon attracted people's attention. Naturally, election campaigns quickly embrace this new trend with open arms. At the same time, researchers have excavated much of the potential of social media as an important public opinion source.

The 2016 presidential election brought social media under the spotlight. Especially when one of the candidates at the time, Donald Trump, is famous for his fondness for intense Twitter usage. Many of the Trump campaign slogans went viral on social media, such as #MakeAmericaGreatAgain or #MAGA. As a response, Clinton camp brought up #ImWithHer and #StrongerTogether. Besides the battle between two candidates, the 2016 presidential election was also severely influenced by malicious users such as zombie accounts controlled by hackers or organizations. There was even a suspicion that Russian agencies play a role on Twitter in their attempt to influence the presidential election [11]. These users spread tons of tweets trying to manipulate the election, which makes the attempt in predicting the election through Twitter become even harder.

With all the chaos in mind, we aim to develop a methodology to effectively forecast the election. When performing prediction methods, we found some unconventional characteristics of the election-related tweets and the user behaviors. This might influence the public opinion if someone relies mostly on social media for election-related information retrieval. In this work, we applied the calibration and trustworthiness differentiation of the users to mitigate the effect of these characteristics and increase the prediction precision. In the meantime, we gain more understanding of what happened on social media during the time of the election campaign.

The main contributions of this work are as follows. Firstly, we use percentage of users instead of number of tweets, which prevents users who post large number of tweets distorting the prediction results. Second, we apply trust filters which was originally used in different domains to evaluate the influences of considered trustworthy users. Last and the most importantly, We propose a novel calibration method to mitigate the influence of the participation bias or demographic differences of election-related Twitter users. To make the calibration possible, we categorize users based on their geographic locations which solves the difficulty of the lack of demographic information.

The remaining sections of this paper are organized as follows. Section 2 introduces methods used to predict the elections and the difficulties when using social media to predict the election, which brought out many questions which we aim to solve. Section 3 describes all procedures from retrieving Twitter data to generate prediction results. Section 4 compares prediction performances between different methods. Lastly, section 5 concludes the findings in this work and provides suggestions on election prediction by social media.

2 RELATED WORK

As the emergence of social media, numerous people start sharing their daily lives on the Internet. From posting a memorable moment, expressing an opinion to support a social issue. Social media has become so ubiquitous that it can be seen as a miniature of real-world social behavior. Soon enough researchers found its potential of being an expedited way to extract thoughts of the public.

Many researchers have used social media as a tool of opinion finder. From disease, disaster, finance, entertainment to politics, any domain that the public opinion matters have the potential of using social media as a poll platform. Twitter especially, due to its limitation to 140 characters in a post, force its users to express their opinion in a most concise way. This characteristic gave researchers a perfect opportunity to identify important information from billions of tweets. Using Twitter to predict the election outcome was first introduced by Tumasjan et al. [17], and soon after various methods trying to extract the âĂIJtrueâĂİ public opinion from Twitter has been used to examine the effectiveness of elections around the world. From German election in 2009 [17], Spain election in 2011 [15], Indonesia election in 2014 [9], India election in 2016 [14] to French election in 2017 [18], regional to national, Twitter has been used to predict various elections.

Tumasjan analyzed the share of Twitter traffic, i.e., the number of tweets that supported different parties, to predict the German election. It shows an astonishing result that the MAE (mean absolute error) for all 6 parties is merely 1.65%. Compared to other sources like election polls and considering the simplicity of the method, using social media to predict the election had soon caught the attention of people.

2.1 different methods to predict the election

2.1.1 Number of tweets. Many of the earlier works like [15, 17] use the number of tweets which mention the supporting parties or candidates as an indicator. However, this method may result in a higher error rate because not all tweets mentioning the parties or candidates possess a positive sentiment. One candidate could have a high exposure on social media while most of the comments are negative.

2.1.2 *Sentiments of tweets.*

To further improve the accuracy of the forecast, sentiment analysis became popular on top of simply counting the number of tweets among the researchers. Chung et al. [4] categorize each tweet as positive, negative or neutral, then counting the sum of supporting tweets and objecting tweets to another side. Burnap et al. [3] apply sentiment scores (+5 to -1) on tweets and sum the scores up. Different sentiment analysis methods are also applied in [9, 14, 18].

2.1.3 *Hashtags as a predicting attribute.*

Bovet et al. [2] applied hashtags as the opinion finder which are used to train a machine learning classifier. Four clusters has been classified as pro-Trump, anti-Clinton, pro-Clinton and anti-Trump which show a clear boundary between the usage of hashtags. They first considered only the strongly connected giant components (SCGC), which is formed by the users that are part of interaction loops and are the most involved in discussions. From the distribution of supporters, they pointed out there exists a huge gap between the number of tweets having hashtags exclusively in the Trump supporters and in the Clinton supporters. Even referring to the number of users, Trump supporters are still much more than Clinton supporters (538,720 for Trump versus 393,829 for Clinton) compared to the actual popular vote ratio of 48.89% for Trump and 51.11% for Clinton. They then used the same collection of hashtags to calculate the whole Twitter dataset and found the situation reversed - Clinton supporters became the majority of the users. This is due to a huge number of Trump supporters belonging to the SCGC. This paper shows a huge potential of using hashtags as a predicting attributes. Nevertheless, their work was mainly used hashtags as a predictor of the poll and did not provide the statewide prediction.

2.1.4 *Hybrid methods/Machine learning algorithm.*

Tsakalidis et al. [16] collect several Twitter-based potential features which originate from the number of posted tweets, positive or negative tweets proportion and the proportion of Twitter users as well. In this research, a poll-based feature is also taken into account. Utilizing the above features as inputs, they have tested several data mining algorithms such as linear regression, Gaussian process and sequential minimal optimization for regression.

2.2 Difficulties in Twitter derived election prediction

Even though using Twitter to predict the election seems to be promising and is convenient compared to the traditional polls, there are questions brought up by other researchers which cannot be ignored. In [12], some suggestions on how to correctly predict the election are given. First, you cannot actually âĂIJpredictâĂİ the election retroactively, so anyone who intends to predict the election should choose the methods or words carefully. Second, social media is fundamentally different from real society - there is more likely to exist spammers and propagandists on the Internet than the real world. Therefore, researchers should consider the credibility of tweets prior to taking all tweets into account. In section 3.3, we applied trust scores in attempt to evaluate the importance of trustworthiness of Twitter users. Third, a successful forecast should be able to explain why and in what condition it predicts. Otherwise, it might be pure luck or the file-drawer effect. Since the 2016 presidential election was over, we can only give best of our knowledge to explain how we chose the methods and why they predict. We will also set foot in the upcoming presidential election in the future research.

Another literature survey paper [6] suggests that âĂIJNot everybody is using Twitter, yet not every Twitter user tweets politics.âĂİ It also possesses a similar view as [12] that not all tweets are true, so it might be required to filter out the untrustworthy tweets before the main process. Gayo-Avello [7] also have insight for using Twitter to predict the election. He thinks among the prediction related researches, many of the sentiment analysis is applied as black box and with naivete. Most of the time, the sentiment-based classifiers perform slightly better than the random classifiers.

He also points out that the demographics are often neglected. Therefore, the researchers cannot consider the Twitter environment as a totally representative and unbiased sample of the voting population. Needless to say, there are a considerable amount of malicious users or spammers spreading misleading information on Twitter. Another important issue is that self-selection bias is usually ignored in the past research. Self-selection bias, or participation bias, may lead a significant influence on the constitution of the tweets. The guarantee the effectiveness of the prediction results, we applied three different election-related attributes to compare with. As it mentioned several times in the previous researches, the demographics of Twitter users and the composition of election-related users should be considered as important effects when we use social media to predict the election. Therefore, we implement a calibration process before the prediction in Sec. 3.4.

In [2] shows a notable property of Trump supporters, which is the majority of strongly connected giant components (SCGC, which is mentioned in Sec. 2.2) which composed the social-connection graph are Trump supporters. In other words, there are more Twitter users who tweet lots of election-related topics as Trump supporters, and many of them are highly connected with each other. This phenomenon distorts the classification of tweets and makes the prediction even harder. Consequently, we applied a similar calculating method which counts the number of users instead of the number of tweets. This is also more consistent with the spirit of the election - One person, one vote.

To mitigate the above mentioned weakness on Twitter-based election prediction, we introduce a user-oriented trust enhancement prediction algorithm and a calibration method for the participation bias.

2.3 Trust Filters on Twitter Users

To better understand the role of trustworthiness on election-related tweets, we have applied a trust scoring method called trust filter [10]. This method has been proved effective in the stock price prediction [8], which relies only on the opinion of Twitter users. In this paper, users are weighted based on their trust scores calculated by trust filters. Therefore, a more trustworthy user would have a higher contribution to the stock price prediction. We would like to apply the idea of trust filter to the election domain, since the election result is only decided by the public opinion. If a trustworthy user can represent the majority of the public or have better insight of the candidate's popularity, trust filters can therefore improve the prediction performance.

3 METHODOLOGY

In this section, we will explain how we retrieve the Twitter data, extract required information, generate trust scores and election-related attributes, calibrate the influence of participation bias and predict the election.

3.1 Twitter Data Acquisition

There is a "spritzer" version of Twitter data collections available on Internet Archive, which is a non-profit digital library. This data set has been fully examined to be consistent with the Tweet2013 collection [13]. Currently, the data sets contain tweets collected from 2011 to early 2020, an approximately 1% sample of public posts, which provides us sufficient quantity and length for research purposes. Its sampling method is collecting all the tweets of a particular time slot (length of 1/100 second) in every second, which guarantees the sampling rate of the number of tweets could be around 1%.

The goal in this paper is to âĂIJpredictâĂİ the outcome of the 2016 United States presidential election, which was held on November 8th, 2016. Therefore, we download the Twitter data from October 1st to November 7th, total 38 days, to ensure the closeness and completeness of the real public opinion. The raw decompressed data set was 513 GB.

Table 1. Top 15 most frequently used hashtags among all election-related tweets (Tweets mentioned candidates) and the counts.

Most frequently used hashtags on tweets mentioned Trump		Most frequently used hashtags on tweets mentioned Clinton	
Hashtag	Count	Hashtag	Count
draintheswamp	10,566,998	imwithher	4,114,818
maga	9,593,476	maga	1,968,479
makeamericagreatagain	2,829,238	spiritcooking	1,855,153
trumptrain	2,421,967	draintheswamp	1,518,953
icymi	2,331,581	trump	1,143,091
trumppence16	2,279,631	flythew	1,012,650
trump	2,150,848	podestaemails25	1,008,920
crookedhillary	1,927,314	neverhillary	957,344
maga3x	1,228,923	lockherup	899,070
6days	1,207,518	election2016	848,288
americafirst	1,205,336	trumppence16	776,753
imwithyou	1,003,915	hillaryindictment	743,956
trump2016	992,726	strongertogether	693,142
nevertrump	936,353	benghazi	683,214
imwithher	852,788	game7	642,018

3.2 Data Extraction and Preprocessing

In this step, we have to extract only the essential data to accelerate following steps. First, a thorough tweet data scan to generate an election-related user list. An election-related user is defined as any Twitter user who had ever posted at least a tweet mentioned the presidential candidates (i.e. Donald Trump or Hillary Clinton) during the observing period. To avoid ambiguity, We defined the âĂIJmentionedâĂİ as referencing the candidateâĂŹs Twitter identity (i.e. realDonaldTrump or HillaryClinton). After the extraction, there are 187,030 election-related users who posted XXX tweets.

Then we extract all tweets posted by election-related users, whether the tweets are election-related or not. Containing all tweets posted by election-related users is crucial for generating trust scores in the next step.

Hashtag is an important indicator in this work, but unlike calculating the number of tweets or the sentiment analysis, there is no trivial way or tool to find which hashtag is popular among political tweets for a certain group. In this work we calculate the most frequently used hashtags in tweets that mentioned any of the candidates. Table 1 shows the partial hashtags and the counts. From the top 50 most frequently used hashtags of both sides, we manually picked hashtags that can be directly linked to support or opposing specific candidates, which is called candidate-related hashtags. First, we selected the main campaign slogans advocated by each side, such as #MAGA, #IamWithHer or #StrongerTogether. Next step is applying the Apriori algorithm to find highly-related hashtags which are on the top 50 list. Excluding the hashtags which stand in a blurry position, we manually identified the pro-Trump hashtags and pro-Clinton hashtags from the highly-related hashtags. The final candidate-related hashtags are shown in Table 2.

3.3 User-oriented Trust Scores and Election-related Attributes

As the difficulties mentioned in Sec. 2, there exists spammers or propagandists spreading misleading tweets which severely interfere with the correctness of the judgements. Therefore, the core concept in this work is instead of predicting

Table 2. Pro-candidate hashtags selected from top 50 most frequently used hashtags of both sides. Pro-Trump hashtags are actually consist of pro-Trump and anti-Clinton hashtags, and similar for pro-Clinton hashtags. $*withher$ hashtags include [im, iam, were, weare, hes, whyim]withher.

Pro-Trump Hashtags	Pro-Clinton Hashtags
teamtrump	∗withher
draintheswamp	strongertogether
trumptrain	getoutthevote
trumppence	nevertrump
crookedhillary	rememberwhentrump
maga	hillary2016
americafirst	lovetrumpshate
trump2016	1uwomen
neverhillary	4hillary
votetrump	1uvote
podestaemails	gohillary
lockherup	stoptrump
hillaryindictment	votehillary
hillaryforprison	uniteblue
fbisongs	dumptrump
voterfraud	fuckfacevonclownstick
trumpforceone	trumptapes
trumprally	gohillary
guns4us	
indicthillary	
hillarycarefail	
obamacarefail	
supportstrump	
paytoplay	
fortrump	
hillaryforprision	
indicthillary	
fbimwithher	

the election simply based on tweets, the prediction should be based on users who posted the tweets. Moreover, we can implement the trust filters introduced in Sec. 2 to enhance the differences on trustworthiness among Twitter users.

We have applied four trust filters in this work, which are Expertise, Experience, Authority and Reputation. Three more userâĂŹs attributes, number of followers, number of friends and average number of words per tweet (Avg_word_tweet), are also taken into consideration for the comparison purpose. For each user, a set of trust scores is generated based on the history of the social interconnection, types of tweets posted and the characteristics of the tweets.

In this work, the definitions of trust scores are as follows.

(1) *Expertise score*: Ratio of election-related tweets to all posted tweets for a single user during the sampling period.
(2) *Experiences score*: Absolute differences on Expertise score between a single user and all users' average.
(3) *Reputation score*: PageRank derived from PageRank algorithm.

(4) *Authority score*: Authority score derives from HITS algorithm. Both Reputation and Authority is based on the directed graph which is constructed by interconnection of the social network [5]. In this work, the interconnection is generated by ”quotes” and ”re-tweets” between Twitter users.

The interactions between users (e.g. re-tweets or quotes) are considered in the entire network of stock-related users and tweets, not only limited to election-related tweets. By doing that, we can expand the social network graph to the most and hence provide more information. This is based on the assumption that if a user is trustworthy as a person, then his/her tweets should be more trustworthy regardless of the subject.

The election-related attributes are the main indicator of the prediction. It shows the tendency that a user is supporting or opposing a specific candidate. We choose three attributes in this work, which are

- *Candidate-related tweets*: Number of candidate-related tweets is counted by tweets which mentioned either @RealDonaldTrump or @HillaryClinton, the Twitter identity of the presidential candidates.
- *Sentiment analysis of the tweets*: We applied SentiWordNet 3.0 [1] as the sentiment analysis dictionary in this work. SentiwordNet is freely available for non-profit research purpose and includes 33,763 non-neutral words (when the words with equal possibility of positive or negative sentiment are excluded). A tweet is first determined which candidate is mentioned, which is the same method to calculate candidate-related tweets, then we sum up the sentiment scores of words that appeared in the dictionary to generate the sentiment of the tweet.
- *Candidate-related hashtags*: Candidate-related hashtags are divided into two groups. Pro-Trump hashtags could be supporting Trump or opposing Clinton and similar to Pro-Clinton hashtags, as shown in Table 2. The assumption is if a tweet contains more Pro-Trump hashtags than Pro-Clinton hashtags, then we can consider this tweet as supporting Donald Trump. Therefore, this attribute only needs to calculate the number of candidate-related hashtags regardless which candidate is mentioned in the tweet.

Unlike the conventional way of simply summing up all attributes, in this work we calculate attributes independently for each user. To be consistent with the presidential election system of the US, we categorize users based on their location, to be specific - which state they lived, based on the location recorded in the user file.

3.4 Calibration and Prediction

Once all trust scores and user attributes are ready, the prediction algorithm can then proceed. The prediction is based on the concept that the social network can be seen as a virtual society. When we collect tweets posted by a particular user and analyze the tweets to determine which candidate the user might be prone to, this procedure would act like taking a poll to the individual. If a sufficient quantity of users is considered, it should be representative as a type of poll.

As we mentioned in Section 2.2, one of the difficulties we might face when extracting information from tweets is the participation bias. If participation bias exists in the sample, then there will be systematic error interfering with the final results. However, what if we can find the ratio of participation bias and calibrate the sample? The assumption is if the participation bias exists, then it should be relatively fixed to a certain ratio between states. For example, if there are X% of TrumpâĂŹs potential supporters posting pro-Trump tweets and Y% of Clinton’s potential supporters posting pro-Clinton tweets in one state, then X/Y should be close to constant no matter which the state is.

Fig. 1(a) to (c) show this trend for all three election-related attributes. R^2, which is a statistical measure of how close the data are to the fitted line, is pretty high in all three charts. The number of Twitter users as supporters is proportionally related to the actual vote counts, but differences exist on the slope of trend lines among supporters between different candidates. Even for states having closed vote counts the number of users as supporters still exists

Table 3. Support rate of the presidential candidates based on candidate-related hashtag-derived users and actual votes before calibration. Three states (Michigan, Minnesota and Wisconsin) had tight races, California and Kentucky represented as the states that one of the candidates is overwhelming.

State	Donald Trump		Hillary Clinton		Support Rate of Trump (Users)	Actual Support Rate of Trump
	Tweet Users	Actual Votes	Tweet Users	Actual Votes		
Michigan	397	2,279,543	124	2,268,839	76.2%	50.1%
Minnesota	142	1,322,951	64	1,367,716	68.9%	49.2%
Wisconsin	171	1,405,284	60	1,382,536	74%	50.4%
California	1853	4,483,810	597	8,753,788	75.6%	33.9%
Kentucky	232	1,202,971	44	628,854	84.1%	65.7%

considerable differences. Table 3 shows the gap no matter the election outcome of the state was a tie or overwhelming. From the above observation, we can conclude that all three election-related attributes are highly connected to actual votes. However, there exists bias between Trump supporters and Clinton supporters.

Even though we know the gap exists, it is impossible to reference actual votes to calibrate the result. Otherwise, it could not be claimed as a prediction and could not be applied to the elections in the future. Therefore, we have to compensate for the calibration and use the population of states instead of actual votes as the calibration references. To be more precise, we can retrieve the registered voters per state before the election as an estimation of the vote count. From Fig. 1(d) we can see the linearity between users as supporters and registered voters still exists and R^2 keeps almost the same as in Fig. 1(c). Table 4 and 5 show the support rate of Donald Trump for all three election-related attributes before and after the calibration.

To sum up, the prediction procedures are as follows:

(1) Determine which candidate is supported by a user based on election-related attributes of the user.
(2) Each user is counted as âĂIJone voteâĂİ to the candidate. To strengthen the credibility of Twitter users, each vote count is multiplied by one of the userâĂŹs trust scores before adding to the candidate's vote count.
(3) Sum up all weighted votes and derive the preliminary prediction results. Vote count is separately calculated by the state detected in the user profile.
(4) Calibrate the results based on the state supporting rate differences.

4 RESULTS

There are three different election-related attributes multiplying seven user-oriented trust scores and one without using trust score, which could generate 24 prediction results. To mitigate calculation efforts, we first compare the linearity of the comparison between users and registered voters like the chart in Fig. 1 (d). By comparing R^2 values, we can select some of the most effective combinations before applying the following calculation. From Table 6, we can find only Expertise and the number of followers have improved R^2 values. Therefore, we only discuss Expertise and the number of followers as trust scores in this section.

Among all trust scores, Authority, Reputation and number of friends performed the worst which did not increase R^2 values but decreased them a lot. Because the social interconnection is defined by re-tweets and quotes in Authority and Reputation filters, someone would receive a high score being re-tweeted or quoted by lots of users. Therefore, one possible explanation for the poor performance of Authority and Reputation filters is in the case of campaign, the

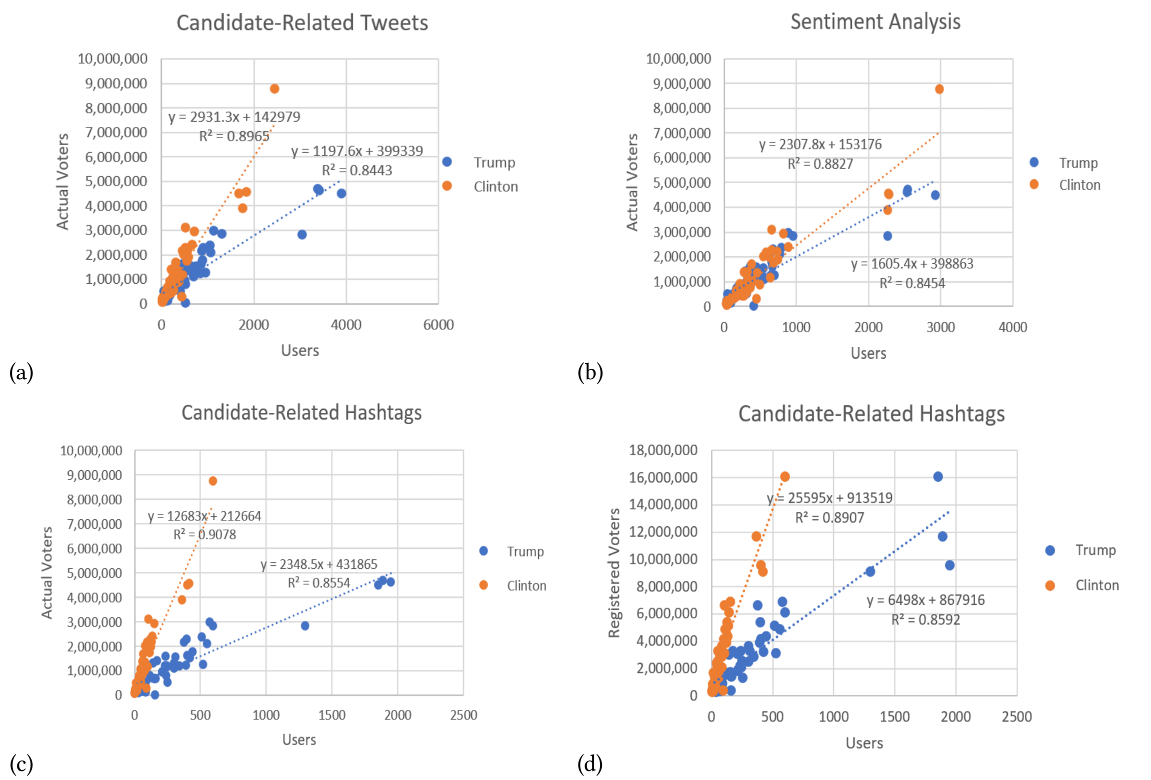

Fig. 1. (a-c) Comparison of actual votes and users derived from (a) Candidate-related tweets. (b) Sentiment analysis. (c) Candidate-related hashtags. (d) Comparison of registered voters and users derived from candidate-related hashtags. Each dot represents one state.

information contained in the tweet does not necessarily need to be accurate but catering to people's favorite to spread widely. As for the huge gap between number of friends and followers, it could be due to the functional differences of having a friend and having a follower. The definition of a "friend" is the account you have followed and the definition of a "follower" is someone who has followed you. One can gain more friends by submitting lots of "follow" requests while it could not be done so for the opposite.

Fig. 2 shows the prediction compared to the actual election outcome. Since the prediction is separated by the states, we can clearly see how the prediction changes between different methods. We can also notice that states with less populations have higher variation and inevitable higher error. This could be due to the lower number of tweets collected from states with lower populations. Because the popular vote results are pretty tight in many states, even a 1 to 2% difference in support rate might overturn the outcome. The predicted approval ratings for candidate-related hashtags in Fig. 2(c) ranges from 40% to 65%. As a comparison, the predicted approval ratings for candidate-related tweets and sentiment analysis in Fig. 2(a) and Fig. 2(b) are only within the range of 45% to 55%, which might increase the possibility of misprediction. However, we can still observe that the slopes of the trendlines for all four charts are pretty close to 1, which means the calibration made in Sec 3.4 did compensate for the quantity inequality between the attributes of Trump supporters and Clinton supporters.

The performance of the prediction is shown in Table 7 which lists all combinations of three election-related attributes and two selected trust scores. If only considering the nationwide popular votes, all approaches have predicted close to the actual result (1.51% to -3.74%). However, this only gives us a blurry picture of the election. The US presidential election is way more complicated and uses Electoral votes to decide the final winner. Therefore, the predicted approval rating per state is transformed into Electoral votes to generate the election outcome in reality. However, compared to the majority of states that the winner of the plurality of the statewide vote receives all of the state's electors, Maine and Nebraska are partially based on this manner and partially based on the plurality of the votes in the congressional district. Due to the limitation of the information extracted from tweets, the identification of users' location is limited to state level. Therefore, the two states are still assigned as the "winner-take-all" system in the prediction. We can find all approaches except one predict Trump could win the election. This is consistent with the 2016 presidential election that Hillary Clinton beat Donald Trump in popular votes while Clinton lost in the electoral votes.

Last thing to examine is the quantified error rate. Mean absolute error is applied here, which does not just calculate errors nationwide but accumulating errors per state. The precise equation is defined as

$$MAE_{state} = \frac{\Sigma_{i \in state} \Sigma_{j \in candidate} |PredictedVotes_{ij} - ActualVotes_{ij}|}{N \times TotalVotes} \tag{1}$$

where N is the number of candidates, which is only considered as 2 in this work.

Among three election-related attributes, candidate-related hashtags possess lowest MAE and are closest to actual results in terms of nationwide popular votes. Candidate-related hashtags plus Expertise scores possess the lowest MAE among all combinations, but it is only slightly better than using candidate-related hashtags alone. Furthermore, Expertise scores do not lower the MAE when combining with candidate-related tweets for sentiment analysis.

5 CONCLUSION

In this paper, we introduce a user-oriented method to calculate the number of supporters for candidates. Compared to simply based on the quantity of tweets or sentiment extracted from tweets, this work applies to a number of users as an elementary unit, which mitigates the problem that some users might intentionally influence public attitude toward certain candidate or party. Like the fact found in [2], there exists a higher proportion of Twitter users who posted lots of tweets from Trump's supporters than Clinton's supporters. This might misguide the prediction using tweet counts to a wrong conclusion.

Three election-related attributes (Candidate-related tweets, sentiment analysis and candidate-related hashtags) have been applied to extract the opinion of the Twitter users. Seven user-related attributes (Expertise, Experience, Authority, Reputation, number of friends, number of followers and average number of words per tweet) are used as the enhancement of the users' trustworthiness and two of them (Expertise and number of followers) are selected to the final prediction comparison. We observed a gap of users between Trump supporters and Clinton supporters. The participation bias could result from different proportions of supporters who would post tweets expressing their political opinions or as described in [2], that the extracted tweets include more strongly connected users who are mostly the supporters on one side. Considering the participation bias which leads to high inequality of election-related attributes for each candidate, we proposed a calibration method without beforehand knowing the actual voters. By doing this, we can still retrieve the supporter quantities per state with relatively high precision. One thing worthy to mention is the calibration method does not require the demographic information of Twitter users.

WWW.IDENTITY.UTEXAS.EDU

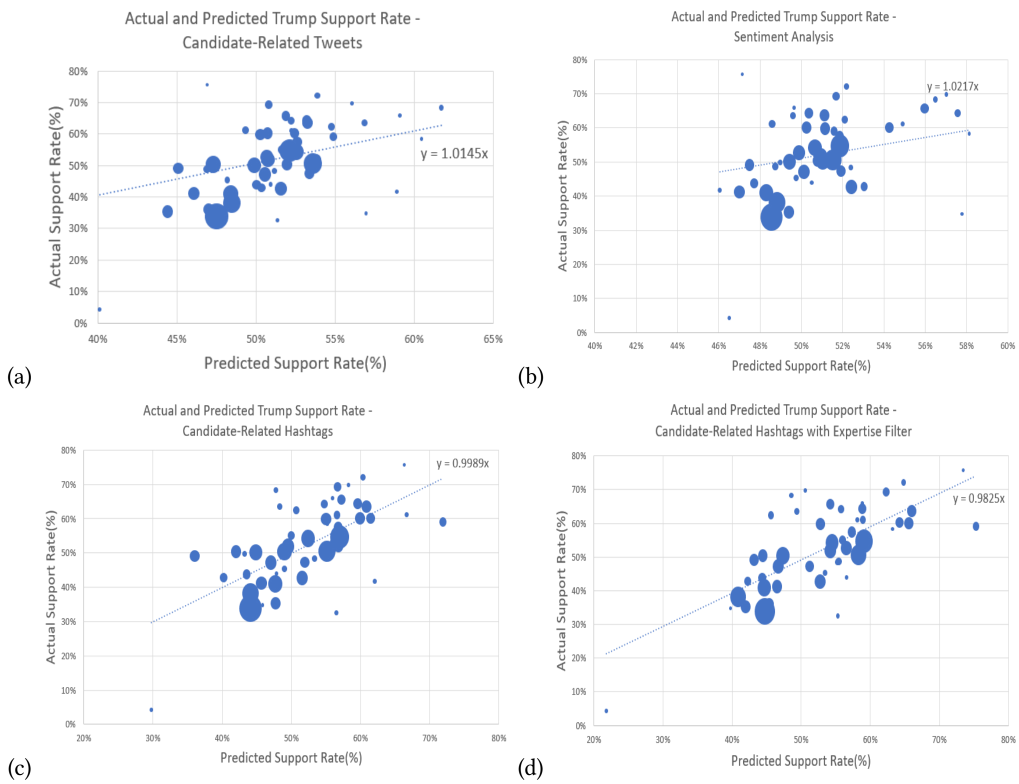

Fig. 2. Comparison of predicted approval rating of Donald Trump and actual Trump's voting rate. Each bubble represents one state and the area of the bubble is proportional to the registered voters per state. If the bubble is closer to the line $Y = X$, the prediction error of the represented state is lower. Four different prediction approaches are shown here. (a) Candidate-related tweets alone. (b) Sentiment analysis alone. (c) Candidate-related hashtags alone. (d) Candidate-related hashtags plus Expertise scores.

Candidate-related hashtags possess best prediction performance among the election-related attributes. The combination of hashtags and Expertise scores shows the lowest MAE, but adding the Expertise score only exhibits limited improvement compared to using hashtags alone. Nevertheless, Expertise score still increases the linearity of the comparison of registered voters and users as supporters in Table 6. There might exist better indicators for the trust scores which can more successfully distinguish the credible users and malicious users such as spammers.

REFERENCES

[1] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.. In *Lrec*, Vol. 10. 2200–2204.

[2] Alexandre Bovet, Flaviano Morone, and Hernán A Makse. 2018. Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific reports* 8, 1 (2018), 1–16.

[3] Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams. 2016. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies* 41 (2016), 230–233.

[4] Jessica Elan Chung and Eni Mustafaraj. 2011. Can collective sentiment expressed on twitter predict political elections?. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*.

[5] Pooja Devi, Ashlesha Gupta, and Ashutosh Dixit. 2014. Comparative study of hits and pagerank link based ranking algorithms. *International Journal of Advanced Research in Computer and Communication Engineering* 3, 2 (2014), 5749–5754.

[6] Daniel Gayo-Avello. 2012. " I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper"–A Balanced Survey on Election Prediction using Twitter Data. *arXiv preprint arXiv:1204.6441* (2012).

[7] Daniel Gayo-Avello. 2012. No, you cannot predict elections with Twitter. *IEEE Internet Computing* 16, 6 (2012), 91–94.

[8] Teng-Chieh Huang, Razieh Nokhbeh Zaeem, and K Suzanne Barber. 2019. It Is an Equal Failing to Trust Everybody and to Trust Nobody: Stock Price Prediction Using Trust Filters and Enhanced User Sentiment on Twitter. *ACM Transactions on Internet Technology (TOIT)* 19, 4 (2019), 1–20.

[9] Mochamad Ibrahim, Omar Abdillah, Alfan F Wicaksono, and Mirna Adriani. 2015. Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In *2015 IEEE international conference on data mining workshop (ICDMW)*. IEEE, 1348–1353.

[10] G. Lin, R. N. Zaeem, H. Sun, and K. S. Barber. 2017. Trust filter for disease surveillance: Identity. In *2017 Intelligent Systems Conference (IntelliSys)*. 1059–1066. https://doi.org/10.1109/IntelliSys.2017.8324259

[11] Darren L Linvill, Brandon C Boatwright, Will J Grant, and Patrick L Warren. 2019. âĂIJTHE RUSSIANS ARE HACKING MY BRAIN!âĂİ investigating Russia's internet research agency twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior* 99 (2019), 292–300.

[12] Panagiotis T Metaxas, Eni Mustafaraj, and Dani Gayo-Avello. 2011. How (not) to predict elections. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 165–171.

[13] Royal Sequiera and Jimmy Lin. 2017. Finally, a downloadable test collection of tweets. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1225–1228.

[14] Parul Sharma and Teng-Sheng Moh. 2016. Prediction of Indian election using sentiment analysis on Hindi Twitter. In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 1966–1971.

[15] Juan M Soler, Fernando Cuartero, and Manuel Roblizo. 2012. Twitter as a tool for predicting elections results. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 1194–1200.

[16] Adam Tsakalidis, Symeon Papadopoulos, Alexandra I Cristea, and Yiannis Kompatsiaris. 2015. Predicting elections for multiple countries using Twitter and polls. *IEEE Intelligent Systems* 30, 2 (2015), 10–17.

[17] Andranik Tumasjan, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*.

[18] Lei Wang and John Q Gan. 2017. Prediction of the 2017 French election based on Twitter data analysis. In *2017 9th Computer Science and Electronic Engineering (CEEC)*. IEEE, 89–93.

Table 4. Support rate of Donald Trump using three election-related attributes (Candidate-related tweets, Sentiment analysis and candidate-related hashtags) before and after calibration.

	Support Rate of Trump Before Calibration			Support Rate of Trump After Calibration			
State	Tweets	SA	Hashtags	Tweets	SA	Hashtags	Actual Support Rate of Trump
Alabama	66.71%	51.38%	85.31%	53.19%	50.37%	59.59%	64.37%
Alaska	72.96%	59.12%	82.89%	60.47%	58.13%	55.16%	58.39%
Arizona	67.25%	51.86%	83.81%	53.80%	50.84%	56.80%	51.89%
Arkansas	65.85%	58.55%	82.65%	52.23%	57.56%	54.74%	64.29%
California	61.48%	49.56%	75.63%	47.51%	48.54%	44.07%	33.87%
Colorado	66.86%	52.93%	81.00%	53.36%	51.91%	51.97%	47.32%
Connecticut	64.14%	54.05%	72.60%	50.35%	53.04%	40.21%	42.86%
Delaware	64.67%	51.52%	78.33%	50.93%	50.50%	47.86%	44.00%
D.C.	54.16%	47.53%	62.50%	40.12%	46.52%	29.73%	4.30%
Florida	67.07%	52.54%	82.87%	53.60%	51.52%	55.11%	50.62%
Georgia	64.45%	50.91%	83.58%	50.69%	49.90%	56.38%	52.66%
Hawaii	65.07%	49.75%	83.65%	51.37%	48.74%	56.51%	32.56%
Idaho	73.97%	57.49%	78.26%	61.71%	56.49%	47.75%	68.31%
Illinois	62.32%	49.31%	78.22%	48.40%	48.29%	47.69%	40.98%
Indiana	64.49%	51.26%	85.48%	50.73%	50.24%	59.91%	60.12%
Iowa	65.29%	52.76%	79.73%	51.61%	51.74%	49.96%	55.06%
Kansas	63.19%	49.60%	83.70%	49.32%	48.58%	56.58%	61.11%
Kentucky	65.53%	56.97%	84.06%	51.88%	55.97%	57.24%	65.67%
Louisiana	66.03%	55.28%	86.27%	52.44%	54.27%	61.46%	60.17%
Maine	64.88%	53.41%	81.82%	51.17%	52.39%	53.32%	48.40%
Maryland	61.01%	49.67%	75.90%	47.02%	48.65%	44.43%	35.98%
Massachusetts	58.48%	50.41%	78.24%	44.41%	49.39%	47.72%	35.35%
Michigan	63.72%	50.45%	76.20%	49.90%	49.43%	44.84%	50.12%
Minnesota	59.14%	48.51%	68.93%	45.08%	47.49%	36.03%	49.17%
Mississippi	68.21%	52.60%	90.97%	54.89%	51.58%	71.90%	59.09%
Missouri	64.06%	52.16%	82.81%	50.27%	51.15%	55.01%	59.82%
Montana	65.87%	55.92%	88.73%	52.25%	54.92%	66.66%	61.11%
Nebraska	69.92%	50.60%	78.65%	56.87%	49.59%	48.33%	63.55%
Nevada	60.88%	49.77%	82.74%	46.88%	48.75%	54.89%	48.71%
New Hampshire	63.56%	50.00%	75.00%	49.73%	48.98%	43.23%	49.80%
New Jersey	65.27%	53.44%	80.72%	51.59%	52.43%	51.53%	42.72%
New Mexico	62.11%	50.77%	79.09%	48.17%	49.75%	48.99%	45.35%
New York	62.38%	49.84%	75.66%	48.46%	48.82%	44.10%	38.23%
North Carolina	64.50%	51.95%	79.44%	50.75%	50.94%	49.52%	51.90%
North Dakota	69.23%	58.02%	84.62%	56.06%	57.03%	58.27%	69.81%
Ohio	66.13%	51.67%	81.28%	52.55%	50.65%	52.43%	54.27%
Oklahoma	64.56%	52.69%	83.73%	50.81%	51.68%	56.65%	69.30%
Oregon	63.85%	48.74%	75.23%	50.04%	47.72%	43.54%	43.84%
Pennsylvania	61.30%	52.04%	79.09%	47.32%	51.03%	48.99%	50.38%
Rhode Island	71.66%	47.06%	86.57%	58.91%	46.05%	62.07%	41.69%
South Carolina	66.19%	52.85%	83.78%	52.61%	51.84%	56.74%	57.46%
South Dakota	71.79%	50.67%	83.33%	59.07%	49.65%	55.94%	65.97%

Table 5. Continued from Table 4

State	Support Rate of Trump Before Calibration Tweets	SA	Hashtags	Support Rate of Trump After Calibration Tweets	SA	Hashtags	Actual Support Rate of Trump
Tennessee	66.76%	52.14%	85.98%	53.24%	51.13%	60.88%	63.62%
Texas	65.76%	52.87%	83.92%	52.13%	51.85%	56.99%	54.71%
Utah	68.09%	53.11%	80.18%	54.75%	52.09%	50.67%	62.38%
Vermont	70.00%	58.76%	76.92%	56.96%	57.77%	45.84%	34.81%
Virginia	64.35%	51.15%	77.76%	50.59%	50.14%	47.02%	47.17%
Washington	60.08%	48.01%	76.82%	46.05%	46.99%	45.69%	41.21%
West Virginia	67.33%	53.19%	85.71%	53.89%	52.18%	60.37%	72.16%
Wisconsin	65.59%	51.79%	74.03%	51.95%	50.78%	41.98%	50.41%
Wyoming	60.92%	48.15%	88.57%	46.92%	47.13%	66.30%	75.71%

Table 6. Average R^2 values of the trendline from comparison chart of actual votes and users derived from election-related attributes, as it shows in Fig. 1 (a) to Fig. 1 (c). The table shows the influences of user-oriented trust scores. The highest values are marked in red and second highest values are marked in blue.

User-Oriented Trust Scores	Election-Related Attributes: Candidate-Related Tweets	Sentiment Analysis	Candidate-Related Hashtags
Original	0.8704	0.8641	0.8816
Expertise	0.8789	0.8739	0.8888
Experience	0.7855	0.8171	0.8084
Authority	0.5058	0.5029	0.5172
Reputation	0.5941	0.6018	0.5106
Friends	0.5548	0.5622	0.6426
Followers	0.8707	0.8631	0.8817
Avg_word_tweet	0.8639	0.8577	0.8753

Table 7. Prediction results for all selected combinations of election-related attributes (Candidate-related tweets, sentiment analysis (SA) and candidate-related hashtags) and trust scores (Expertise and number of followers). Popular votes show the support rate of Trump when considering only two candidates, Donald Trump and Hillary Clinton. Electoral votes show the difference between two candidates. Positive means Trump wins more and vice versa. Mean absolute error (MAE) is calculated by the sum of absolute error per state and divided by all votes, which might be different from the definition of other works. The results most close to the election result or having lowest error are marked in red.

	Tweets	Tweets + Expertise	Tweets + Followers	SA	SA + Expertise	SA + Followers	Hashtags	Hashtags+ Expertise	Hashtags+ Followers	Election Result
Popular Votes	45.15%	46.07%	45.45%	49.52%	50.03%	50.40%	48.89%	48.24%	47.78%	48.89%
Electoral Votes	117	-29	111	81	77	91	1	47	31	77
MAE (State)	6.64%	6.82%	6.52%	6.76%	6.86%	6.74%	5.76%	5.64%	5.67%	0%