12-2012

Are Twitter Users Equal in Predicting Elections? A Study of User Groups in Predicting 2012 U.S. Republican Primaries

Lu Chen
Wright State University - Main Campus

Wenbo Wang
Wright State University - Main Campus

Amit P. Sheth
Wright State University - Main Campus, amit.sheth@wright.edu

Are Twitter Users Equal in Predicting Elections? A Study of User Groups in Predicting 2012 U.S. Republican Presidential Primaries

Lu Chen, Wenbo Wang, and Amit P. Sheth

Kno.e.sis Center, Wright State University, Dayton, OH 45435 USA
{chen,wenbo,amit}@knoesis.org

Abstract. Existing studies on predicting election results are under the assumption that all the users should be treated equally. However, recent work [14] shows that social media users from different groups (e.g., "silent majority" vs. "vocal minority") have significant differences in the generated content and tweeting behavior. The effect of these differences on predicting election results has not been exploited yet. In this paper, we study the spectrum of Twitter users who participate in the on-line discussion of 2012 U.S. Republican Presidential Primaries, and examine the predictive power of different user groups (e.g., highly engaged users vs. lowly engaged users, right-leaning users vs. left-leaning users) against Super Tuesday primaries in 10 states. Specifically, we characterize users across four dimensions, including three dimensions of user participation measured by tweet-based properties (engagement degree, tweet mode, and content type) and one dimension of users' political preference. We study different groups of users in each dimension and compare them on the task of electoral prediction. The insights gained in this study can shed light on improving the social media based prediction from the user sampling perspective and more.

Keywords: Electoral Prediction, Twitter Analytics, Social Intelligence, User Categorization, Engagement Degree, Tweet Mode, Content Type, Political Preference

1 Introduction

Over 80% of Americans use at least one social network, and people spend nearly 23% of their online time on social networks[1]. Among those popular social network sites, Twitter has over 140 million active users, generating over 340 millions tweets per day[2]. The topics being discussed in social networks cover almost every aspect of our lives. On one hand, researchers are making every effort to make sense of the social data to understand what is going on in the world. On the other hand, there is a surge of interest in building systems that harness the power of social data to predict what is about to happen. It has been reported

[1] http://www.socialmediaexaminer.com/26-promising-social-media-stats-for-small-businesses/
[2] http://blog.twitter.com/2012/03/twitter-turns-six.html

that social data is used to predict box-office revenues [1, 13], stock market [3, 9, 18], and election outcomes [2, 11, 15–17], etc.

Existing studies using social data to predict election results have focused on obtaining the measures/indicators (e.g., mention counts or sentiment of a party or candidate) from social data to perform the prediction. They treat all the users equally, and ignore the fact that social media users engage in the elections in different ways and with different levels of involvement. A recent study [14] has shown that significant differences exist between silent majority (users who tweeted once) and vocal minority (users who tweet very often) in the generated content and tweeting behavior in the context of political elections. However, whether and how such differences will affect the prediction results still remains unexplored. For example, in our study, 56.07% of Twitter users who participate in the discussion of 2012 U.S. Republican Primaries post only one tweet. The identification of the voting intent of these users could be more challenging than that of the users who post more tweets. *Will such differences lead to different prediction performance?* Furthermore, the users participating in the discussion may have different political preference. *Is it the case that the prediction based on the right-leaning users will be more accurate than that based on the left-leaning users, since it is the Republican Primaries?* Exploring these questions can expand our understanding of social media based prediction, and shed light on using user sampling to further improve the prediction performance.

In this paper, we investigate above questions by studying different groups of social media users who engage in the discussions of elections, and comparing the predictive power among these user groups. Specifically, we chose the 2012 U.S. Republican Presidential Primaries on Super Tuesday[3] among four candidates: Newt Gingrich, Ron Paul, Mitt Romney and Rick Santorum. We collected 6,008,062 tweets from 933,343 users talking about these four candidates in an eight week period before the elections. All the users are characterized across four dimensions: engagement degree, tweet mode, content type, and political preference. We first investigated the user categorization on each dimension, and then compared different groups of users with the task of predicting the results of Super Tuesday races in 10 states. Instead of using tweet volume or the overall sentiment of tweet corpus as the predictor, we estimated the "vote" of each user by analyzing his/her tweets, and predicted the results based on "vote-counting". The results were evaluated in two ways: (1) the accuracy of predicting winners, and (2) the error rate between the predicted votes and the actual votes for each candidate.

The main contributions of this paper are as follows. (1) We group social media users based on their participation (engagement degree, tweet mode, and content type) as well as political preference, and study the participation behaviors of different user groups, (2) we present a method to predict the "vote" of a user based on the analysis of his/her tweets, and count the votes of users to predict the election result, and (3) we examine the predictive power of different user groups in predicting the results of Super Tuesday races in 10 states.

[3] http://en.wikipedia.org/wiki/Super_Tuesday

2 Related Work

Using social media data for electoral prediction has attracted increasing interest in recent years. Gayo-Avello [7] provided a comprehensive summary of literature on election prediction with Twitter data. Here, we focus on the literature which is most relevant to our task.

O'Connor et al. [15] discovered correlations between public opinion derived from presidential job approval polls and sentiment based on analysis of Twitter messages. Tumasjan et al. [17] used the number of tweets mentioning a party or candidate to accurately predict the 2009 German federal elections. Sang et al. [16] showed that merely counting the tweets is not sufficient for electoral predictions, and the prediction could be improved by improving the quality of data collection and performing sentiment analysis. Bermingham and Smeaton [2] used both sentiment-based and volume-based measures to predict results of the 2011 Irish General Election. They found that social analytics using both measures were predictive, and volume was a stronger indicator than sentiment.

Meanwhile, some researchers argue that the predictive power of social media might be exaggerated, and the challenges of building the predictive models based on social data have been underestimated, especially for the electoral predictions. Gayo-Avello [8] showed that simple approaches based on mention counts and polarity lexicons failed in predicting the result of 2008 U.S. Presidential Elections. In another study [12], the authors found that the social data did only slightly better than chance in predicting the 2010 U.S. Congressional elections. They pointed out the need of obtaining a random sample of likely voters in order to achieve accurate electoral predictions.

To summarize, existing studies on electoral prediction have focused on exploring the measures and indicators (e.g., tweet volume or sentiment) to predict the election results, and left the problem that whether all the users and their tweets should be treated equally unexplored. Previous research [14] has shown that different groups of users could be very different in tweeting behavior and generated content. Should a user who posts one tweet be handled in the same way as another user who posts 100 tweets in predicating the election? Should a democrat be treated equally as a republican in predicting the republican primaries? We focus on exploring such questions in this paper.

3 User Categorization

Using Twitter Streaming API, we collected tweets that contain the words "gingrich, "romney", "ron paul", or "santorum" from January 10th 2012 to March 5th 2012 (Super Tuesday was March 6th). Totally, the dataset comprises 6,008,062 tweets from 933,343 users. The data used for this study is collected as part of a social web application – Twitris[4], which provides real-time monitoring and multi-faceted analysis of social signals surrounding an event (e.g., the 2012 U.S. Presidential Election). In this section, we discuss user categorization on four dimensions, and study the participation behaviors of different user groups.

[4] http://twitris.knoesis.org/

Table 1: User Groups with Different Engagement Degrees

Engagement Degree	Very Low	Low	Medium	High	Very High	Total
Tweets per User	1	[2, 10]	[11, 50]	[51, 300]	>300	
User Volume	56.07%	35.93%	6.19%	1.58%	0.23%	100%
Tweet Volume	8.71%	20.31%	20.42%	26.83%	23.73%	100%

3.1 Categorizing Users by Engagement Degree

We use the number of tweets posted by a user to measure his/her engagement degree. The less tweets a user posts, the more challenging the user's voting intent can be predicted. An extreme example is to predict the voting intent of a user who posted only one tweet. Thus, we want to examine the predictive power of different user groups with various engagement degrees.

Specifically, we divided users into the following five groups: the users who post only one tweet (*very low*), 2-10 tweets (*low*), 11-50 tweets (*medium*), 51-300 tweets (*high*), and more than 300 tweets (*very high*). Table 1 shows the distribution of users and tweets over five engagement categories. We found that more than half of the users in the dataset belong to the *very low* group, which contributes only 8.71% of the tweet volume, while the very highly engaged group contributes 23.73% of the tweet volume with only 0.23% of all the users. It raises the question of whether the tweet volume is a proper predictor, given that a small group of users can produce a large amount of tweets.

To further study the behaviors of the users on different engagement levels, we examined the usage of hashtags and URLs in different user groups (see Table 2). We found that the users who are more engaged in the discussion use more hashtags and URLs in their tweets. Since hashtags and URLs are frequently used in Twitter as ways of promotion, e.g, hashtags can be used to create trending topics, the usage of hashtags and URLs reflects the users' intent to attract people's attention on the topic they discuss. **The more engaged users show stronger such intent and are more involved in the election event.** Specifically, only 22.95% of all tweets created by very lowly engaged users contain hashtags, this proportion increases to 39.45% in the *very high* engagement group. In addition, the average number of hashtags per tweet (among the tweets that contain hashtags) is 1.43 in the *very low* engagement group, while this number is 2.68 for the very highly engaged users. The users who are more engaged also use more URLs, and generate less tweets that are only text (not containing any hashtag or URL). We will see whether and how such differences among user engagement groups will lead to varied results in predicting the elections later.

3.2 Categorizing Users by Tweet Mode

There are two main ways of producing a tweet, i.e., creating the tweet by the user himself/herself (original tweet) or forwarding another user's tweet (retweet). Original tweets are considered to reflect the users' attitude, however, the reason for retweeting can be varied, e.g., to inform or entertain the users' followers, to

Table 2: Usage of Hashtags and URLs by Different User Groups

Engagement Degree	Very Low	Low	Medium	High	Very High
Tweets with Hashtags	22.95%	26.98%	30.58%	32.85%	39.45%
Hashtags per tweet	1.43	1.58	1.95	2.14	2.68
Tweets with URLs	33.44%	40.16%	49.02%	53.88%	59.89%
Only Text	50.93%	43.11%	34.19%	29.35%	25.31%

be friendly to the one who created the tweet, etc., thus retweets do not necessarily reflect the users' thoughts. It may lead to different prediction performance between the users who post more original tweets and the users who have more retweets, since the voting intent of the latter is more difficult to recognize.

According to users' preference on generating their tweets, i.e., tweet mode, we classified the users as *original tweet-dominant*, *original tweet-prone*, *balanced*, *retweet-prone* and *retweet-dominant*. A user is classified as *original tweet-dominant* if less than 20% of all his/her tweets are retweets. Each user from *retweet-dominant* group has more than 80% of all his/her tweets that are retweets. In Table 3, we illustrate the categorization, the user distribution over the five categories, and the tweet mode of users in different engagement groups.

Table 3: User Distribution over Categorization of Tweet Mode

Tweet Mode	Orig. Tweet-Dom.	Orig. Tweet-Prone	Balanced	RT-Prone	RT-Dom.	Total
Retweet	<20%	[20%, 40%)	[40%, 60%)	[60%, 80%)	>=80%	
All Users	49.04%	4.76%	7.22%	4.27%	34.71%	100%
Very Low	55.32%	0.00%	0.00%	0.00%	44.68%	100%
Low	41.04%	9.83%	16.70%	8.81%	23.62%	100%
Medium	42.01%	15.41%	14.78%	13.21%	14.59%	100%
High	38.44%	15.21%	16.62%	15.39%	14.35%	100%
Very High	31.89%	13.88%	17.03%	17.73%	19.47%	100%

It is interesting to find that the *original tweet-dominant* group accounts for the biggest proportion of users in every user engagement group, and this proportion declines with the increasing degree of user engagement (55.32% of very lowly engaged users are original tweet-dominant, while only 31.89% of very highly engaged users are original tweet-dominant). It is also worth noting that **a significant number of users (34.71% of all the users) belong to the *retweet -dominant* group, whose voting intent might be difficult to detect**.

3.3 Categorizing Users by Content Type

Based on content, tweets can be classified into two classes – opinion and information (i.e., subjective and objective). Studying the difference between the users who post more information and the users who are keen to express their opinions could provide us with another perspective in understanding the effect of using these two types of content in electoral prediction.

We first identified whether a tweet represents positive or negative opinion about an election candidate. We used the approach proposed in [4] to learn a candidate-specific sentiment lexicon from the tweet collection. This lexicon contained sentiment words and phrases which were used to express positive or negative opinions about the candidates. Totally, this lexicon comprised 1674 positive words/phrases and 1842 negative words/phrases, which was applied to recognize the opinions about each candidate in tweets. If a tweet contained more positive (negative) words than negative (positive) words about a candidate, e.g., Mitt Romney, it was annotated as "positive(negative)_Mitt_Romney". If there were no sentiment words found in a tweet about a candidate, e.g., Mitt Romney, it was annotated as "neutral_Mitt_Romney". Thus, every tweet has four sentiment labels (one for each candidate). "I want Romney to win over Santorum but you must be careful in your negative ads." was labeled as "neutral_Newt_Gingrich", 'neutral_Ron_Paul", "positive_Mitt_Romney", and "negative_Rick_Santorum".

The tweets that are positive or negative about any candidate are considered *opinion* tweets, and the tweets that are neutral about all the candidates are considered *information* tweets. We also used a five-point scale to classify the users based on whether they post more opinion or information with their tweets: *opinion-dominant*, *opinion-prone*, *balanced*, *information-prone* and *information-dominant*. Table 4 shows the user distribution among all the users, and the users in different engagement groups categorized by content type.

The users from *very low* engagement group have only one tweet, so they either belong to *opinion-dominant* (39%) or *information dominant* (61%). With users' engagement increasing from low to very high, the proportions of *opinion-dominant*, *opinion-prone* and *information-dominant* users dramatically decrease from 11.09% to 0.05%, 11.75% to 0.42%, and 27.40% to 0.66%, respectively. In contrast, the proportions of *balanced* and *information-prone* users grow. In *high* and *very high* engagement groups, the *balanced* and *information-prone* users together accounted for more than 95% of all users. It shows the tendency that **more engaged users post a mixture of content, with similar proportion of opinion and information, or larger proportion of information**.

3.4 Identifying Users' Political Preference

Since we focused on the Republican Presidential Primaries, it should be interesting to compare two groups of users with different political preferences – left-leaning and right-leaning. Some efforts [6, 10, 5] have been made to address the problem of predicting the political preference/orientation of Twitter users in recent years. In our study, we use a simple but effective method to identify the left-leaning and right-leaning users.

We collected a set of Twitter users with known political preference from Twellow[5]. Specifically, we acquired 10,324 users who are labeled as Republican, conservative, Libertarian or Tea Party as right-leaning users, and 9,545 users who are labeled as Democrat, liberal or progressive as left-leaning users. We denote the top 1000 left-leaning users and top 1000 right-leaning users who have

[5] http://www.twellow.com/

Table 4: User Distribution over Categorization of Content Type

Content Type	Opinion-Dom.	Opinion-Prone	Balanced	Info.-Prone	Info.-Dom.	Total
Opinion	>=80%	[60%, 80%)	[40%, 60%)	[20%, 40%)	<20%	
All Users	25.89%	4.74%	14.75%	9.92%	44.70%	100%
Very Low	39.00%	0.00%	0.00%	0.00%	61.00%	100%
Low	11.09%	11.75%	30.92%	18.84%	27.40%	100%
Medium	0.59%	8.02%	42.85%	38.60%	9.94%	100%
High	0.22%	1.43%	53.84%	42.06%	2.45%	100%
Very High	0.05%	0.42%	58.98%	39.89%	0.66%	100%

the most followers as L_I and R_I, respectively. Among the remaining users that are not contained in L_I or R_I, there are 1,169 left-leaning users and 2,172 right-leaning users included in our dataset, and these 3,341 users are denoted as T.

The intuitive idea is that a user tends to follow others who share the same political preference as his/hers. The more right-leaning users one follows, the more likely that he/she belongs to the right-leaning group. Among all the users that a user is following, let N_l be the number of left-leaning users from L_I and N_r be the number of right-leaning users from R_I. We estimated the probability that the user is left-leaning as $\frac{N_l}{N_l+N_r}$, and the probability that the user is right-leaning as $\frac{N_r}{N_l+N_r}$. The user is labeled as left-leaning (right-leaning) if the probability that he/she is left-leaning (right-leaning) is more than a threshold τ. Empirically, we set $\tau = 0.6$ in our study. We tested this method on the labeled dataset T and the result shows that this method correctly identified the political preferences of 3,088 users out of all 3,341 users (with an accuracy of 0.9243).

Totally, this method identified the political preferences of 83,934 users from all of the 933,343 users in our dataset. Other users may not follow any of the users in L_I or R_I, or follow similar numbers of left-leaning and right-leaning users, thus their political preferences could not be identified. Table 5 shows the comparison of left-leaning and right-leaning users in our dataset. We found that **right-leaning users were more involved in this election event in several ways**. Specifically, the number of right-leaning users was two times more than that of left-leaning users, and the right-leaning users generated 2.65 times the number of tweets as the left-leaning users. Compared with the left-leaning users, the right-leaning users tended to create more original tweets and used more hashtags and URLs in their tweets. This result is quite reasonable since it was the Republican election, with which the right-leaning users are supposed to be more concerned than the left-leaning users.

4 Electoral Prediction with Different User Groups

In this section, we examine the predictive power of different user groups in predicting the Super Tuesday election results in 10 states. We first recognized the users from each state. There are two types of location information from

Table 5: Comparison between Left-leaning and Right-leaning Users

Political Preference	Left-Leaning	Right-Leaning
# of Tweets	702,178	1,863,186
# of Users	27,586	56,348
Tweets per User	25.5	33.1
Original Tweets	48.46%	56.09%
Retweets	51.54%	43.91%
Tweets with Hashtags	33.02%	37.99%
Hashtags per Tweet	1.68	1.93
Tweets with URLs	45.95%	52.75%
Only Text	34.57%	30.19%
Opinion	41.31%	41.47%

Twitter – the geographic location of a tweet, and the user location in the profile. We utilized the background knowledge from LinkedGeoData[6] to identify the states from user location information[7]. If the user's state could not be inferred from his/her location information, we utilized the geographic locations of his/her tweets. A user was recognized as from a state if his/her tweets were from that state. Table 6 illustrates the distribution of users and tweets among the 10 Super Tuesday states. We also compared the number of users and tweets in each state to its population. The Pearson's r for the correlation between the number of users/tweets and the population is 0.9459/0.9667 ($p < .0001$). In the following of this section, we first describe how we estimated a user's vote, and next report the prediction results, followed by a discussion of the results.

4.1 Estimating a User's Vote

To answer the question that for whom a user will vote, we need to find for which candidate the user shows the most support. We think there are two indicators that can be extracted from a user's tweets of one candidate – mention and sentiment. Intuitively, people show their support for celebrities through frequently talking about them and expressing positive sentiments about them.

As described in Section 3.3, we have analyzed each user's tweets, identified which candidate is mentioned, and whether a positive or negative opinion is expressed towards a candidate in a tweet. For each user, let N be the number of all his/her tweets, $N_m(c)$ be the number of tweets in which he/she mentioned a candidate c, $N_{pos}(c)$ be the number of positive tweets about c from the user, $N_{neg}(c)$ be the number of negative tweets about c from the user. We define the user's support score for c as:

$$\begin{cases} (1 - \frac{N_{neg}(c)}{N_{pos}(c)+\beta}) \times \frac{N_m(c)}{N} & \text{if } N_{pos}(c) + N_{neg}(c) > 0 \\ \gamma \times \frac{N_m(c)}{N} & \text{otherwise} \end{cases}$$

[6] http://linkedgeodata.org/About

[7] Since geographical analysis is not the focus of this paper, we did not verify if the users are actually from the locations specified in their profiles.

Table 6: Distribution of Tweets and Users over 10 Super Tuesday States

U.S. State	Alaska	Georgia	Idaho	Massachusetts	North Dakota
# of Tweets	7,633	88,555	17,331	89,842	3,763
# of Users	736	13,210	1,830	15,009	661
Population	722,718	9,815,210	1,584,985	6,587,536	683,932
	Ohio	Okalahoma	Tennessee	Vermont	Virginia
# of Tweets	102,880	27,747	58,384	5,525	73,172
# of Users	18,066	3,965	7,980	1,183	9,796
Population	11,544,951	3,791,508	6,403,353	626,431	8,096,604

where β ($0 < \beta < 1$) is a smoothing parameter, and γ ($0 < \gamma < 1$) is used to discount the score when the user does not express any opinion towards c ($N_{pos}(c) = N_{neg}(c) = 0$). We used $\beta = \gamma = 0.5$ in our study. According to this definition, the more positive tweets (less negative tweets) are posted about c, and the more c is mentioned, the higher the user's support score for c is. After calculating a user's support score for every candidate, we selected the candidate who received the highest score as the one that the user will vote for.

4.2 Prediction Results

In this section, we report the comparison of different user groups in predicting Super Tuesday races, and discuss our findings.

To predict the election results in a state, we used only the collection of users who are identified from that state. Then we further divided each user collection of one state over four dimensions – engagement degree, tweet mode, content type, and political preference. In order to get enough users in one group, we used a more coarse-grained classification instead of the five-point scales described in the section of User Categorization. To be specific, we classified users as three different groups according to their engagement degree: *very low*, *low*, and *high**. The *very low* and *low* engagement groups are the same as what we have defined in Section 3.1. The *high** engagement group comprises the users who post more than 10 tweets (i.e., the aggregation of the medium, high and very high groups defined previously). Based on the tweet mode, the users were divided into two groups: *original tweet-prone** and *retweet-prone**, depending on whether they post more original tweets or more retweets. Similarly, the users were classified as *opinion-prone** or *information-prone** according to whether they post more opinions or more information. The *right-leaning* users and *left-leaning* users were also identified from the user collection of each state. In all, for each state, there were nine user groups over four different dimensions.

We also considered users in different time windows. Our dataset contains the users and their tweets discussing the election in 8 weeks prior to the election day. We wanted to see whether it will make any difference to use the data in different time windows. Here we examined four time windows – *7 days*, *14 days*, *28 days* or *56 days* prior to the election day. For example, the *7 days* window is from February 28th to March 5th. In a specific time window, we assessed a user's vote using only the set of tweets he/she creates during this time[8].

With each group of users in a specific state and a specific time window, we counted the users' votes for each candidate, and the one who received the most votes was predicted as the winner of the election in that state. The performance of a prediction was evaluated in two ways: (1) whether the predicted winner is the actual winner, and (2) comparing the predicted percentage of votes for each candidate with his actual percentage of votes, and getting the mean absolute error (MAE) of the four candidates.

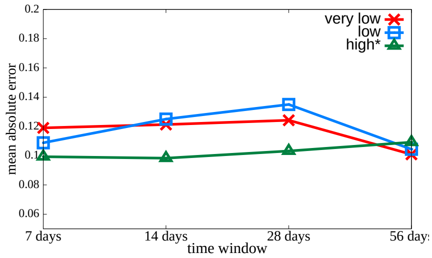
Table 7 shows the accuracy of winner prediction by different user groups in different time windows. The accuracy was calculated as $\frac{N_{true}^{state}}{N^{state}}$, in which N_{true}^{state} was the number of states where the winner was correctly predicted, and $N^{state}(=10)$ was the number of all Super Tuesday states. Figure 1 illustrates the average MAE of the predictions in 10 states by different user groups in different time windows. From Table 7 and Figure 1, we do see that different user groups on each dimension show varied prediction performance.

Table 7: The Accuracy of Winner Prediction by Different User Groups

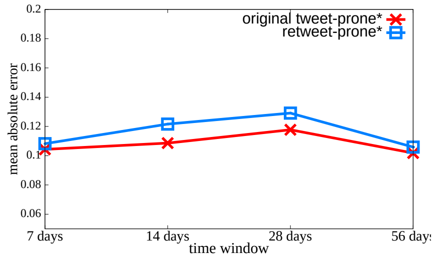
	7 Days	14 Days	28 Days	56 Days
Engagement Degree				
Very Low	0.5	0.4	0.3	0.6
Low	0.7	0.3	0.3	0.6
High*	0.5	0.8	0.5	0.6
Tweet Mode				
Original Tweet-Prone*	0.7	0.4	0.4	0.6
Retweet-Prone*	0.6	0.3	0.3	0.6
Content Type				
Opinion-Prone*	0.5	0.5	0.4	0.6
Information-Prone*	0.6	0.4	0.3	0.7
Political Preference				
Left-Leaning	0.5	0.2	0.3	0.6
Right-Leaning	0.5	0.7	0.7	0.8

As shown in Table 7, the *high** engagement group correctly predicted the winners of 5 states in 7 day, 8 in 14, 5 in 28 and 6 in 56 day time windows, respectively, which is slightly better than the average performance of *very low* and *low* engagement groups. In addition, the average prediction error of *high** engagement group is smaller than that of *very low* and *low* engagement groups in three out of the four time windows (see Figure 1a). Comparing two user groups over the tweet mode dimension, *original tweet-prone** group beat the *retweet-prone** group by achieving better accuracy on winner prediction and smaller prediction error in almost all the time windows (see Figure 1b). The two user groups categorized by content type also show differences in predicting the

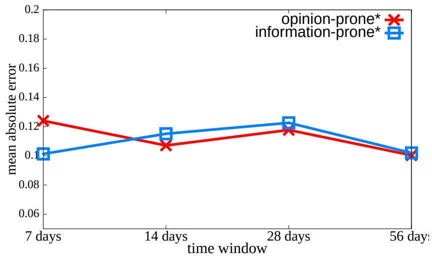
[8] A user's vote might be varied in different time windows, since we used different sets of tweets for the assessment.



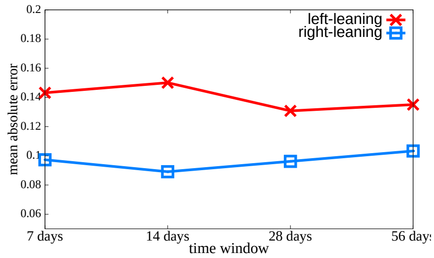
(a) User Groups Categorized by Engagement Degree



(b) User Groups Categorized by Tweet Mode



(c) User Groups Categorized by Content Type



(d) User Groups Categorized by Political Preference

Fig. 1: The Mean Absolute Error (MAE) by Different User Groups in 10 States

elections, but the difference is not as clear as that of user groups on other dimensions. On winner prediction, the *opinion-prone** group achieved better accuracy in 14 day and 28 day time windows, and *information-prone** group achieved better accuracy in 7 day and 56 day time windows. Although the prediction error of *opinion-prone** group was smaller than that of *information-prone** group in three time windows, the gap was quite small (see Figure 1c).

It is interesting to find that, among all the user groups, the *right-leaning* group achieved the best prediction results. In Table 7 and Figure 1d, *right-leaning* group correctly predicted the winners of 5, 7, 7 and 8 states (out of 10 states) in 7 day, 14 day, 28 day and 56 day time windows, respectively. Furthermore, it also showed the smallest prediction error (<0.1) in three out of four time windows among all the user groups. In contrast, the prediction by the *left-leaning* group was the least accurate. In the worst case, it correctly predicted the winners in only 2 states (in the 14 day time window), and its prediction error was over 0.15.

To further verify our observation, we looked at the average prediction error of four time windows for each state, and applied paired t-test to find whether the difference of the average prediction errors in 10 states between a pair of user groups was statistically significant. The test showed that the difference between *right-leaning* and *left-leaning* user groups is statistically highly significant ($p < .001$). The difference between *low* and *high** engagement user groups was also found statistically significant ($p < .01$). However, the difference between *original tweet-prone** and *retweet-prone**, or between *opinion-prone** and *information-prone** was not significant.

In addition, we also compared our results with random predictions. From the winner prediction perspective, all the user groups except the *left-leaning* one beat the random baseline (25% accuracy) in all the time windows. The random baseline showed a mean prediction error (of vote percentage) over 0.13, which is higher than that of all the user groups except the *left-leaning* one.

4.3 Discussion

There are at least two factors that could affect the accuracy of electoral prediction. Firstly, whether the prediction of users' votes is accurate. Secondly, whether the users' opinion is representative of the actual voters' opinion. We interpret the varied prediction results with different user groups based on these two factors.

In our study, the *high** engagement user group achieved better prediction results than *very low* and *low* engagement groups. It may be due to two reasons. Firstly, high engagement users posted more tweets. Since our prediction of a user's vote is based on the analysis of his/her tweets, it should be more reliable to make the prediction using more tweets. Secondly, according to our analysis, more engaged users showed stronger intent and were more involved in the election event. It might suggest that users in the *high** engagement group were more likely to vote, compared with the users in the *very low* and *low* engagement groups.

However, the *low* engagement group did not show better performance compared with the *very low* engagement group. One possible explanation might be that the users from these two groups are not that different. A more fine-grained classification of users with different engagement degrees might provide more insight. Since the prediction is state-based, we could not get enough users in each group (especially the groups of highly engaged users) if we divided users into more groups. It is worth noting that more than 90% of all the users in our dataset belonged to *very low* and *low* engagement groups. Accurately predicting the votes of these users is one of the biggest challenges in electoral prediction.

The results also show that the prediction based on users who post more original tweets is slightly more accurate than that based on users who retweet more, although the difference is not significant. It may be due to the difficulty of identifying users' voting intent from retweets. In most of the current prediction studies, original tweets and retweets are treated equally with the same method. Further studies are needed to compare these two types of tweets in prediction, and a different method might be needed for identifying users' intent from retweets. In addition, a more fine-grained classification of users according to their tweet mode could provide more insight.

No significant difference is found between the *opinion-prone** and the *information-prone** user groups in prediction. It suggests that the likely voters cannot be identified based on whether users post more opinions or more information. It also reveals that the prediction of users' votes based on more opinion tweets is not necessarily more accurate than the prediction using more information tweets.

The *right-leaning* user group provides the most accurate prediction result, which is significantly better than that of the *left-leaning* group. In the best case (56 day time window), the right-leaning user group correctly predict the winners in 8 out of 10 states (Alaska, Georgia, Idaho, Massachusetts, Ohio, Oklahoma, Vermont, and Virginia). It is worth noting that this result is significantly better than the prediction result of the same elections based on Twitter analysis reported in the news article[9], in which the winners are correctly predicted in only 5 out of 10 states(Georgia, Idaho, Massachusetts, Ohio, Virginia). Since the elections being predicted were Republican primaries, the attitude of *right-leaning* users could be more representative of the voters' attitude. To some extent, it demonstrates the importance of identifying likely voters in electoral prediction.

This study can be further improved from several aspects. First, more effort could be made to investigate the possible data biases (e.g., spam tweets and political campaign tweets) and how they might affect the results. Second, we estimated the vote intent of each Twitter user in our dataset and aggregated them to predict the election results. However, these users are not necessarily the actual voters. Identification of the actual voters from social media is also an interesting problem to explore. In addition, our work examined the predictive power of different user groups in republican primaries, thus some of our findings may not apply to other elections of different natures, e.g., general elections. However, we believe the general principle that Twitter users are not equal in predictions is common for all elections.

5 Conclusion

In this paper, we studied the spectrum of Twitter users in the context of the 2012 U.S. Republican Presidential Primaries, and examined the predictive power of different user groups in predicting the results from the 10 states that held Republican primaries on Super Tuesday. We divided users into different groups on four dimensions – engagement degree, tweet mode, content type, and political preference. To predict the election results, we first predicted each user's vote based on analyzing the mentions and sentiments of the candidates in the user's tweets, and then counted the votes received by each candidate from every user group. Comparing the prediction results obtained by different user groups, we found the result achieved by right-leaning users was significantly better than that achieved by left-leaning users. The prediction based on highly engaged users was better than that based on lowly engaged users. The users who posted more original tweets provided slightly higher accuracy in the prediction than the users who retweeted more did. To some extent, these findings demonstrate the importance of identifying likely voters and user sampling in electoral predictions.

Acknowledgment

We are grateful to Ashutosh Jadhav, Hemant Purohit, Pavan Kapanipathi and Pramod Anantharam for helpful discussions, Sarasi Lalithsena and Sujan Udayanga for insightful comments. We also thank the anonymous reviewers for their useful comments. This research was supported by US NSF grant IIS-1111182: SoCS: Social Media Enhanced Organizational Sensemaking in Emergency Response.

[9] http://www.usatoday.com/tech/news/story/2012-03-07/election-social-media/53402838/1

References

1. Asur, S. and Huberman, B.A.: Predicting the future with social media. Arxiv preprint arXiv:1003.5699. (2010) http://arxiv.org/abs/1003.5699
2. Bermingham, A. and Smeaton, A.F.: On using Twitter to monitor political sentiment and predict election results. In: Proceedings of the Sentiment Analysis where AI meets Psychology Workshop at IJCNLP. (2011)
3. Bollen, J. and Mao, H. and Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science. (2011)
4. Chen, L. and Wang, W. and Nagarajan, M. and Wang, S. and Sheth, A.P.: Extracting Diverse Sentiment Expressions with Target-dependent Polarity from Twitter. In: Proceedings of ICWSM. (2012)
5. Conover, M.D. and Gonçalves, B. and Ratkiewicz, J. and Flammini, A. and Menczer, F.: Predicting the political alignment of twitter users. In: Proceedings of the IEEE 3rd International Confernece on Social Computing, pp. 192–199 (2011)
6. Conover, M.D. and Ratkiewicz, J. and Francisco, M. and Goncalves, B. and Flammini, A. and Menczer, F.: Political polarization on twitter. In: Proceedings of ICWSM, pp. 89–96 (2011)
7. Gayo-Avello, D.: I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper. Arxiv preprint arXiv:1204.6441. (2012) http://arxiv.org/abs/1204.6441
8. Gayo-Avello, D.: Don't turn social media into another 'Literary Digest' poll. Communications of the ACM, 54(10) pp. 121–128 (2011)
9. Gilbert, E. and Karahalios, K.: Widespread worry and the stock market. In: Proceedings of ICWSM. pp. 229–247 (2010)
10. Golbeck, J. and Hansen, D.: Computing political preference among twitter followers. In: Proceedings of the annual conference on Human factors in computing systems, pp. 1105–1108 (2011)
11. Livne, A. and Simmons, M.P. and Adar, E. and Adamic, L.A.: The party is over here: Structure and content in the 2010 election. In: Proceedings of ICWSM. (2011)
12. Metaxas, P.T., Mustafaraj, E. and Gayo-Avello, D.: How (Not) to predict elections. In: Proceedings of the IEEE 3rd International Confernece on Social Computing, pp. 165–171 (2011)
13. Mishne, G. and Glance, N.: Predicting movie sales from blogger sentiment. In: AAAI Symposium on Computational Approaches to Analysing Weblogs. (2006)
14. Mustafaraj, E. and Finn, S. and Whitlock, C. and Metaxas, P.T.: Vocal minority versus silent majority: Discovering the opionions of the long tail. In: Proceedings of the IEEE 3rd International Confernece on Social Computing, pp. 103–110 (2011)
15. OConnor, B. and Balasubramanyan, R. and Routledge, B.R. and Smith, N.A.: From tweets to polls: Linking text sentiment to public opinion time series. In: Proceedings of ICWSM, pp. 122–129 (2011)
16. Sang, E.T.K. and Bos, J.: Predicting the 2011 Dutch Senate Election Results with Twitter. In: Proceedings of SASN 2012, the EACL 2012 Workshop on Semantic Analysis in Social Networks, pp. 53–60 (2012)
17. Tumasjan, A. and Sprenger, T.O. and Sandner, P.G. and Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: Proceedings of ICWSM, pp. 178–185 (2010)
18. Zhang, X. and Fuehres, H. and Gloor, P.A.: Predicting Stock Market Indicators Through Twitter I hope it is not as bad as I fear. Procedia-Social and Behavioral Sciences, 26 pp. 55–62 (2011)