International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 03, Volume 4 (March 2017) Special Issue www.ijirae.com

Quantifying Political Leaning from Tweets, Retweets, and Retweeters

Sudhakar.K[1]
HOD /Assistant Professor
Computer Science and Engineering
Sengunthar College of Engineering, Tiruchengode, India

Manoja.R[2]
Computer Science and Engineering
Sengunthar College of Engineering
Tiruchengode, India

Abstract* — The widespread use of online social networks (OSNs) to disseminate information and exchange opinions, by the general public, news media and political actors alike, has enabled new avenues of research in computational political science. In this paper, we study the problem of quantifying and inferring the political leaning of Twitter users. We formulate political leaning inference as a convex optimization problem that incorporates two ideas: (a) users are consistent in their actions of tweeting and retweeting about political issues, and (b) similar users tend to be retweeted by similar audience. We then apply our inference technique to 119 million election-related tweets collected in seven months during the 2012 U.S. presidential election campaign. On a set of frequently retweeted sources, our technique achieves 94% accuracy and high rank correlation as compared with manually created labels. By studying the political leaning of 1,000 frequently retweeted sources, 232,000 ordinary users who retweeted them, and the hashtags used by these sources, our quantitative study sheds light on the political demographics of the Twitter population, and the temporal dynamics of political polarization as events unfold.

***Index Terms* —Twitter, political science, data analytics, inference, convex programming, signal processing**

I. INTRODUCTION

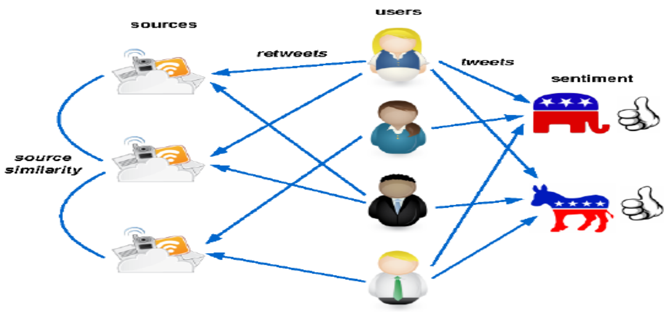
In recent years, big online social media data have found many applications in the intersection of political and computer science. Examples include answering questions in political and social science (e.g., proving/disproving the existence of media bias [3, 30] and the "echo chamber" effect [1, 5]), using online social media to predict election outcomes [46, 31], and personalizing social media feeds so as to provide a fair and balanced view of people's opinions on controversial issues [36]. A prerequisite for answering the above research questions is the ability to accurately estimate the political leaning of the population involved. If it is not met, either the conclusion will be invalid, the prediction will perform poorly [35, 37] due to a skew towards highly vocal individuals [33], or user experience will suffer. In the context of Twitter, accurate political leaning estimation poses two key challenges: (a) Is it possible to assign meaningful numerical scores to tweeters of their position in the political spectrum? (b) How can we devise a method that leverages the scale of Twitter data while respecting the rate limits imposed by the Twitter API? Focusing on "popular" Twitter users who have been retweeted many times, we propose a new approach that incorporates the following two sets of information to infer their political leaning.

***Tweets* and *retweets*:** the target users' temporal patterns of being retweeted, and the tweets published by their retweeters. The insight is that a user's tweet contents should be consistent with who they retweet, e.g., if a user tweets a lot during a political event, she is expected to also retweet a lot at the same time. This is the "time series" aspect of the data.

***Retweeters*:** The identities of the users who retweeted the target users. The insight is similar users get followed and retweeted by similar audience due to the homophily principle. This is the "network" aspect of the data. Our technical contribution is to frame political leaning inference as a convex optimization problem that jointly maximizes tweet-retweet agreement with an error term, and user similarity agreement with a regularization term which is constructed to also account for heterogeneity in data.

International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 03, Volume 4 (March 2017) Special Issue www.ijirae.com

Our technique requires only a steady stream of tweets but not the Twitter social network, and the computed scores have a simple interpretation of "averaging," i.e., a score is the average number of positive/negative tweets expressed when retweeting the target user. Using a set of 119 million tweets on the U.S. Presidental election of 2012 collected over seven months, we extensively evaluate our method to show that it outperforms several standard algorithms and is robust with respect to variations to the algorithm. The second part of this paper presents a quantitative study on our collected tweets from the 2012 election, by first (a) quantifying the political leaning of 1,000 frequently-retweeted Twitter users, and then (b) using their political leaning, infer the leaning of 232,000 ordinary Twitter users.



II. FORMULATION

2.1 MOTIVATION AND SUMMARY

To motivate our approach in using retweets for political leaning inference, we present two examples to highlight the existence of useful signals from retweet information. From our dataset on the 2012 presidential election (see details in Section 4), we identify the Twitter accounts of two major media sources, one with liberal and the other with conservative leaning. In Figure 2 we plot their retweet popularity (their columns in matrix A, see Section 3.2) during the12 events in the dataset (see Table 1). We observe negative correlation ($\rho = -0.246$) between the two sources' pat-terns of being retweeted, especially during events 6 and 7.[3] This can be explained by Democrat/Republican supporters enthusiastically retweeting Romney/Obama-bashing tweets published by the media outlets during the corresponding events.

This example leads us to conjecture that: (a) the number of retweets received by a retweet source during an event can be a signal of her political leaning. In particular, one would expect a politically inclined tweeter to receive more retweets during a favorable event. (b) The action of retweeting carries implicit sentiment of the retweeter. This is true even if the original tweet does not carry any sentiment itself. The intuition is that users tend to follow and retweet those who share similar political views, e.g., a user is more likely to retweet a newspaper to which it subscribes than any random newspaper, a manifestation of the homophily principle.

2.2 DEFINITIONS

Consider two political parties running for an election. During the election campaign there have been E *events* which attracted considerable attention in Twitter. We are interested in quantifying the *liberal-conservative*[5] political leaning of N prominent retweet *sources*, e.g., media outlets' Twitter accounts and celebrities. For event i, let U_i be the set of users who tweeted about the event, and T_{iu} be the set of tweets sent by user $u \in U_i$ about the event. Also define each tweet t to carry a score $s_t \in [-1,1]$, such that it is 1 if the tweet shows full support for one candidate, or -1 if full support is for the other. Then for user u her approval score is$\sum$

$$s_t|T_{iu}|\, t \in Tiu$$

Averaging over all users in U_i, the *average tweet leaning* y_i of event i is[6]

$$y_i = |U_i| \quad \sum \sum s_t |T_{iu}|\, u \in Ui\ t \in Tiu \quad \text{minimize}$$
$$x \in R^N \|Ax - y\|^2 2.$$

2.3 REGULARIZATION

In statistical inference, solving ill-posed problems requires us to incorporate prior knowledge of the problem to rule out undesirable solutions. One such common approach is regularization, and we can change the objective function in Problem (5), $\|Ax - y\|^2 2$, to $\|Ax - y\|^2 2 +^{\lambda f(x),}$ where $\lambda > 0$ is a regularization parameter, and f(x) quantifies the"fitness" of a solution such that undesirable solutions have higher f (x) values. For example, Tikhonov regularization for least-squares uses $f(x) = \|x\|^2$

We propose a regularization term that favors political leaning assignments x with x_i being close to x_j if sources i and j have similar retweet responses. Let W_{ij} be a regularization weight between sources i and j such that $W_{ij} \geq 0$ and $W_{ij} = W_{ji}$. Furthermore, let W be the weight matrix whose elements are W_{ij}. Then we set

$$f(x) = \sum_{i=1,j=1}\sum W_{ij}(x_i - x_j)^2,$$

So that if W_{ij} is large(sources i and j are similar),then x_i should be close to x_j to minimize $W_{ij}(x_i - x_j)^2$.

III. DATASET

In this section we describe the collection and processing of our Twitter dataset of the U.S. presidential election of 2012. Our dataset was collected over a time span of seven months, covering from the initial phases to the climax of the campaign.

Data Collection: From April 8 to November 10 2012, we used the Twitter streaming API[10] to collect 119 million tweets which contain any one of the following keyword phrases: "obama", "romney", "barack", "mitt", "paul ryan", "joe biden", "presidential", "gop", "dems", "republican" and "democrat" (string matching is case-insensitive).

Event Identification: By inspecting the time series of tweet counts in Figure 6, we manually identified 12 events as listed in Table 1. We defined the dates of an event as fol-lows: the start date was identified based on our knowledge of the event, e.g., the start time of a presidential debate, and the end date was defined as the day when the number of tweets reached a local minimum or dropped below that of the start date. After the events were identified, we extracted all tweets in the specified time interval[11] without additional filtering, assuming all tweets are relevant to the event and those outside are irrelevant.

Extracting Tweet Sentiment: We applied SentiStrength a lexicon-based sentiment analysis package,to extract the sentiment of tweets. We adjusted the provided lexicon by compiling a high-frequency tweet-word list per event, and then removing words[13] that we consider to not carry sentiment in the context of elections. Sentiment analysis was done as a ternary (positive, negative, neutral) classification. For each tweet t, we set its score $s_t = -1$ if either (a) it mentions solely the Democrat camp (has "obama", "biden" etc. in text) and is classified to have positive sentiment, or (b) it mentions solely the Republican camp ("romney", "ryan" etc.) and has negative sentiment. We set $s_t = 1$ if the opposite criterion is satisfied. If both criteria are not satisfied, we set $s_t = 0$.

IV. EVALUATION

4.1 GROUND TRUTH CONSTRUCTION

We compare the political leaning scores learnt by our technique with "ground truth" constructed by human evaluation. First, 100 sources are randomly selected from the 1,000 most popular retweet sources.[14] Then we ask 12 human judges with sufficient knowledge of American politics to classify each of the 100 sources as "L" (Liberal, if she is expected to vote Obama), "C" (Conservative, if expected to vote Romney), or "N" (Neutral, if she cannot decide), supposing each source is one voter who would vote in the presidential election. For each source, a judge is presented with (a) the source's user profile, including screen name, full name, self description, location and etc., and (b) ten random tweets published by the source. Given the set of labels, we compute our ground truth political leaning scores $\{x_i\}$ as follows: for each label of L/N/C, assign a score of −1/0/+1, then the score of source i, call it x_i, as the average of her labels. Assumption of Twitter political leaning being the perceived leaning by a source's retweeters.

If source i has x_i with extreme values (−1 or +1), then it is unambiguously liberal/conservative, but if x_i takes intermediate values, then some human judges may be confused with the source's leaning, and the general Twitter population is likely to have similar confusion, which suggests that the "correct" x_i should also take intermediate values. Defining $\{x_i\}$ this way also allows us to understand the usefulness of quantifying political leaning with a continuous score. Obviously, if all x_i are either −1 or +1, a simple binary classification of the sources is enough, but as we see in Figure 7, $\{x_i\}$ is evenly spread across the range of allowed values [−1, 1], so characterizing sources with simple binary, or even ternary, classification appears too coarse. To further support our claim, we also compute the inter-rater agreement of our manual labels as Fleiss' $\kappa = 0.430$ [17], a moderate level of agreement [28].

TABLE 1 - SUMMARY OF EVENTS IDENTIFIED IN THE DATASET.

ID	Dates[12]	Description	# tweets (m)	# non-RT tweets (m)
1	May 9 - 12	Obama supports same-sex marriage	2.10	1.35
2	Jun 28 - 30	Supreme court upholds health care law	1.21	0.78
3	Aug 11 - 12	Paul Ryan selected as Republican VP candidate	1.62	0.96
4	Aug 28 - Sep 1	Republican National Convention	4.32	2.80
5	Sep 4 - 8	Democratic National Convention	5.81	3.61
6	Sep 18 - 22	Romney's 47 percent comment	4.10	2.55
7	Oct 4 - 5	First presidential debate	3.49	2.19
8	Oct 12 - 13	Vice presidential debate	1.92	1.19
9	Oct 17 - 19	Second presidential debate	4.38	2.67
10	Oct 23 - 26	Third presidential debate	5.62	3.35
11	Nov 4 - 6	Elections (before Obama projected to win)	7.50	4.40
12	Nov 7 - 9	Elections (after Obama projected to win)	6.86	4.43
Total			48.90	30.28

Finally, we also manually classify each of the 1,000 most popular sources into four classes:

- *Parody*: role-playing and joke accounts created for entertainment purposes (example joke tweet: "Icooked Romney noodles Obama self," a pun on "I cooked ramen noodles all by myself")
- *Political*: candidates of the current election and ac-counts of political organizations
- *Media*: outlets for distributing information in an objective manner, setting aside media bias issues
- *Others*: personal accounts, including those of celebrities, pundits, reporters, bloggers and politicians (excluding election candidates).

4.2 PERFORMANCE METRICS

The quality of political leaning scores is measured under two criteria. Classification: One should be able to directly infer the liberal/conservative stance of a source i from her sign of x_i, i.e., it is liberal if $x_i < 0$, or conservative if $x_i > 0$. Taking $\{x_i\}$ as ground truth, we say source i is correctly classified if the signs of x_i and x_i agree.[15] Classification performance is measured using the standard metrics of accuracy, precision, recall and F1 score. Rank correlation: The set of scores $\{x_i\}$ induce a ranking of the sources by their political leaning. This ranking should be close to that induced by the ground truth scores $\{x_i\}$. We measure this aspect of performance using Kendall's τ , which varies from −1 (perfect disagreement) to 1 (perfect agreement).

4.3 RESULTS

We solve Problem (10) with A_L = {Obama2012} and A_C = {MittRomney} and compare the results with those from a number of algorithms:

- *PCA: we run Principal Components Analysis on A with each column being the feature vector of a source, with or without the columns being standardized, and take the first component as $\{x_i\}$. This is the baseline when we use only the information from A (retweet counts).*
- *Eigenvector: we compute the second smallest eigen-vector of L, with L becoming computed from S being either the cosine or Jaccard matrix. This is a technique commonly seen in spectral graph partitioning [15], and is the standard approach when only the information from S (retweeters) is available. Note that the x computed this way is equivalent to solving the optimization problem: minimize $x^T Lx$, subject to $\|x\|_2 = 1$, $x^T 1 = 0$.*
- *Sentiment analysis: we take x_i as the average sentiment of the tweets published by source i, using the same methdology in computing y [45]. This is the baseline when only tweets are used.*
- *SVM on hashtags: following [12], for each source we compute its feature vector as the term frequencies of the 23,794 hashtags used by the top 1,000 sources. We then train an SVM classifier (linear kernel, standardized features) using the 900 of the top 1,000 sources that are not labeled by 12 human judge as training data.[16] Hashtags have been suggested to contain more information than raw tweet text and a better source of features .*
- *Retweet network analysis: also following [12], we con-struct an undirected graph of the 25,000 Twitter users with highest retweet activity. The edge between two users is weighted as the number of times either one has retweeted the other.*

International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 03, Volume 4 (March 2017) Special Issue www.ijirae.com

Then we apply majority voting-based label prop-agation [42] with initial conditions (label assignments) from the leading eigenvector method for modularity maximization [38]. Given the modularity maximization method used here is analogous to the above eigenvector baseline. We treat its output as political leaning scores and report its performance. We also experimented with synchronous soft label propagation [54], similar to the algorithm in [19], but it did not produce better results.

Table 2 reports the evaluation results. Our algorithm, in combining information from A, S and y, performs signifi-cantly better than all other algorithms in terms of Kendall's τ, F1 score and accuracy. We also observe that if no matrix scaling is applied in constructing W, the algorithm tends to assign all $\{x_i\}$ (except those of anchors) to the same sign, resulting in poor classification performance. In the remaining of this paper, we focus on the political leaning scores computed using cosine similarity.

TABLE 2- PERFORMANCE OF OUR ALGORITHM COMPARED TO OTHERS. BEST TWO RESULTS (ALMOST ALWAYS DUE TO OUR METHOD) ARE HIGHLIGHTED IN BOLD.

Algorithm	Kendall's τ	Precision, L	Recall, L	Precision, C	Recall, C	F1 score, L	F1 score, C	Accuracy
Ours, cosine matrix	0.652	0.942	0.970	0.935	0.935	0.955	0.935	0.94
Ours, Jaccard matrix	0.654	0.940	0.940	0.879	0.935	0.940	0.906	0.92
Ours, cosine w/o scaling	0.649	0.670	1	0	0	0.80	0	0.67
Ours, Jaccard w/o scaling	0.641	0	0	0.31	1	0	0.473	0.31
PCA	0.002	0.663	0.791	0.300	0.194	0.721	0.235	0.59
PCA, standardized columns	0.011	0.750	0.224	0.325	0.839	0.345	0.468	0.41
Eigenvector, cosine matrix	0.297	0.667	0.985	0	0	0.795	0	0.66
Eigenvector, Jaccard matrix	0.308	0.663	0.970	0	0	0.787	0	0.65
Sentiment analysis	0.511	0.926	0.746	0.700	0.903	0.826	0.789	0.78
SVM on hashtags	0.436	0.863	0.851	0.840	0.677	0.857	0.750	0.78
Modularity maximization	0.510	0.934	0.851	0.763	0.935	0.891	0.841	0.86
Label propagation + modularity	—	0.940	0.940	0.935	0.935	0.940	0.935	0.92

V. QUANTITATIVE STUDY

5.1 QUANTIFYING PROMINENT RETWEET SOURCES

We study the properties of the political leaning of the 1,000 most popular retweet sources. Similar to that in the score histogram on the full set has a bimodal distribution. We note that by incorporating retweeter information, our algorithm is able to correctly position "difficult" sources that were highly retweeted during events unfavorable to the candidate they support, e.g., JoeBiden, CBSNews and all accounts related to Big Bird, an improvement over the preliminary version of this paper [51]. We also find that WSJ is assigned a slightly liberal score. This is consistent with findings reported in prior studies [25, 32] explained by the separation between WSJ's news and editorial sections. We also study the score distributions of sources grouped by account type. Figure 10 shows noticeable differences among the different groups. Parody sources are skewed towards the liberal side. Political sources are strongly polarized with no sources having neutral (close to zero) scores. Media sources are less polarized with a more even spread of scores. Sources in the "Others" class have a score distribution close to that of political sources, but also includes a few neutral scores, which can be attributed to celebrities with no clear political stance. To check the differences are not due to artifacts of our algorithm, e.g., a possibility of biasing sources with more available data towards the ends of the spectrum, we plot the relationship between the number of times a source is retweeted and its score polarity in and find no significant correlation between the variables ($\rho = 0.0153$, p-value= .628).

5.2 QUANTIFYING ORDINARY TWITTER USERS

Given the political leaning of 1,000 retweet sources, we can use them to infer the political leaning of ordinary Twitter users who have retweeted the sources. We consider the set of users seen in our dataset who have retweeted the sources at least ten times, including retweets made during non- event time periods. In total there are 232,000 such users. We caution this set of users is not necessarily representative of the general Twitter population, or even the full population of our dataset (9.92 million users in total), but we believe it is possible to "propagate" score estimates from these 232,000 users to everyone else, which remains as future work.

VI. CONCLUSIONS AND FUTURE WORK

Scoring individuals by their political leaning is a fundamental research question in computational political science. From roll calls to newspapers, and then to blogs and microblogs, researchers have been exploring ways to use bigger and bigger data for political leaning inference. But new challenges arise in how one can exploit the structure of the data, because bigger often means noisier and sparser. In this paper, we assume: (a) Twitter users tend to tweet and retweet consistently, and (b) similar Twitter users tend to be retweeted by similar sets of audience, to develop a con-vex optimization-based political leaning inference technique that is simple, efficient and intuitive. Our method is evaluated on a large dataset of 119 million U.S. election-related tweets collected over seven months, and using manually constructed ground truth labels, we found it to outperform many baseline algorithms. With its reliability validated, we applied it to quantify a set of prominent retweet sources, and then propagated their political leaning to a larger set of ordinary Twitter users and hashtags. The temporal dynamics of political leaning and polarization were also studied. We believe this is the first systematic step in this type of approaches in quantifying Twitter users' behaviour. The retweet matrix and retweet average scores can be used to develop new models and algorithms to analyze more complex tweet-and-retweet features. Our optimization frame-work can readily be adapted to incorporate other types of information. The y vector does not need to be computed from sentiment analysis of tweets, but can be built from exogenous information (e.g., poll results) to match the opinions of the retweet population. Similarly, the A matrix, currently built with each row corresponding to one event, can be made to correspond to other groupings of tweets, such as by economic or diplomatic issues. The W matrix can be constructed from other types of network data or similarity measures. Our methodology is also applicable to other OSNs with retweet-like endorsement mechanisms, such as Facebook and YouTube with "like" functionality.

REFERENCES

[1]. S. Ansolabehere, R. Lessem, and J. M. Snyder, "The ori-entation of newspaper endorsements in U.S. elections," Quarterly Journal of Political Science, vol. 1, no. 4, pp. 393- 404, 2006.
[2]. D. boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet:Conversational aspects of retweeting on Twitter," in Proc. HICSS, 2010.
[3]. M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," in Proc. ICWSM, 2010.
[4]. S.Finn, E.Mustafaraj, and P.T.Metaxas,"The co- retweeted network and its applications for measuring the perceived political polarization," in Proc. WEBIST, 2014.
[5]. J. Golbeck and D. Hansen, "A method for computing political preference among Twitter followers," Social Networks,vol. 36, pp. 177-184, 2014.
[6]. F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang, "Quantifying political leaning from tweets and retweets," in Proc.ICWSM, 2013.
[7]. I. Weber, V. R. K. Garimella, and A. Teka, "Political hashtag trends," in Proc. ECIR, 2013.
[8]. A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M.Welpe, "Predicting elections with Twitter: What 140 characters reveal about political sentiment," in Proc. ICWSM, 2010.
[9]. Felix Ming Fai Wong, Member, IEEE, Chee Wei Tan, Senior Member, IEEE, Soumya Sen, Senior Member, IEEE, Mung Chiang, Fellow, IEEE, "Quantifying Political Leaning from Tweets, Retweets, and Retweeters", IEEE Transactions on Knowledge and Data Engineering, 2016.