**Quantifying Political Leaning from Tweets, Retweets, and Retweeters**

**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.

**EXISTING SYSTEM:**

* A variety of methods have been proposed to quantify the extent of bias in traditional news media. Indirect methods involve linking media outlets to reference points with known political positions. For example, Lott and Hassett linked the sentiment of newspaper headlines to economic indicators.
* Groseclose and Milyo linked media outlets to Congress members by co-citation of think tanks, and then assigned political bias scores to media outlets based on the Americans for Democratic Action (ADA) scores of Congress members.
* Gentzkow and Shapiro performed an automated analysis of text content in newspaper articles, and quantified media slant as the tendency of a newspaper to use phrases more commonly used by Republican or Democrat members of the Congress.

**DISADVANTAGES OF EXISTING SYSTEM:**

* In the context of Twitter, accurate political leaning estimation poses two key challenges:
* Is it possible to assign meaningful numerical scores to tweeters of their position in the political spectrum?
* How can we devise a method that leverages the scale of Twitter data while respecting the rate limits imposed by the Twitter API?
* A more fundamental problem is data scarcity. Because the amount of data available for analysis is limited by how fast the media sources publish, researchers may need to aggregate data created over long periods of time, often years, to perform reliable analysis. Analyzing media sources through their OSN outlets offers many unprecedented opportunities with high volume data from interaction with their audience.

**PROPOSED SYSTEM:**

* 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.
* 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.
* Liberals dominate the population of less vocal Twitter users with less retweet activity, but for highly vocal populations, the liberal-convservative split is balanced. Partisanship also increases with vocalness of the population.
* Hashtag usage patterns change significantly as political events unfold.
* As an event is happening, the influx of Twitter users participating in the discussion makes the active population more liberal and less polarized.

**ADVANTAGES OF PROPOSED SYSTEM:**

* 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.
* Our optimization framework can readily be adapted to incorporate other types of information.

**SYSTEM ARCHITECTURE:**



**MODULES:**

* System Construction
* Tweets, Retweets and Retweeters
* Lexicon-Based Sentiment Analysis
* Quantitative Study

**MODULES DESCSRIPTION:**

**System Construction Module**

* In the first module we develop the System Construction module, to Quantifying Political Leaning from Tweets, Retweets, and Retweeters. For this purpose we develop User entities. In User entity, a user can search information about political tweets in Twitter OSN.a user can able to search other user tweets about political.
* A user can mining the informations in big data like Twitter OSN about the political tweets. It minimize the time and increase accuracy of results to know about the political tweets. A user can also view the other user id who tweet about the politics. A user can able to search tweets tweeted by users about particular political leader .
* 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.

**Tweets, Retweets and Retweeters:**

* In this module we saw about Tweets Retweets Retweeters and tweet retweet agreement with an error term. 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.
* 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.
* 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. 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 .

**Lexicon-Based Sentiment Analysis:**

* In this module 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 words13 that we consider to not carry sentiment in the context of elections. Sentiment analysis was done as a ternary (positive, negative, neutral) classification.
* 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.
* For each tweet t, we set its score st = −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 st = 1 if the opposite criterion is satisfied. If both criteria are not satisfied, we set st = 0.
* This module is very useful for know about positive and negative tweets about political leaders by using a Lexicon-Based Sentiment Analysis.

**Quantitative Study:**

* In this module we study the properties of the political leaning of the 1,000 most popular retweet sources.The score histogram on the full set has a bimodal distribution. We note that by incorporating retweeter infor- mation, 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.
* In a 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 nonevent 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.
* A Quantitative Study on our collected tweets from the 2012 election, by first (a) quantifying the political leaning of 1,000 frequentlyretweeted Twitter users, and then (b) using their political leaning, infer the leaning of 232,000 ordinary Twitter users. We make a number of findings:
* Parody Twitter accounts have a higher tendency to be liberal as compared to other account types.They also tend to be temporally less stable.
* Liberals dominate the population of less vocal Twitter users with less retweet activity, but for highly vocal populations, the liberal conservative split is balanced. Partnership also increases with vocalness of the population .
* Hashtag usage patterns change significantly as political events unfold.
* As an event is happening , the influx of Twitter users participating in the discussion makes the active population more liberal and less polarized.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : JAVA/J2EE
* Tool : Netbeans 7.2.1
* Database : MYSQL

**REFERENCE:**

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.**