

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.

Technofist,

YES Complex, 19/3&4, 2nd Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph: 080-40969981, Website: www.technofist.com. E-mail: technofist.projects@gmail.com

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 and the “echo chamber” effect using online social media to predict election outcomes), and personalizing social media feeds so as to provide a fair and balanced view of people’s opinions on controversial issues. 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 due to a skew towards highly vocal individuals, 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? 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.

Technofist,

YES Complex, 19/3&4, 2nd Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph: 080-40969981, Website: www.technofist.com. E-mail: technofist.projects@gmail.com

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. See Figure 1 for an illustration. Using a set of 119 million tweets on the U.S. presidential 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.

Technofist,

YES Complex, 19/3&4, 2nd Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph: 080-40969981, Website: www.technofist.com. E-mail: technofist.projects@gmail.com