

Scalable News Slant Measurement Using Twitter

Huyen Le, Zubair Shafiq, Padmini Srinivasan

Department of Computer Science

The University of Iowa

Email: {huyen-t-le, zubair-shafiq, padmini-srinivasan}@uiowa.edu

Abstract

Prior research has reported evidence of polarization in the consumption and distribution of political news in online social media. Methods to effectively measure slant of news articles at a large scale are lacking in prior literature. To fill this gap, we propose a method to measure slant of individual news articles by observing their sharing patterns on Twitter. Our method monitors tweets about a news article and a set of landmark Democrat and Republican accounts on Twitter to estimate its slant. Our results show that the slant of news articles estimated by our method matches the slant estimated by crowdsourced workers.

Introduction

People increasingly rely on online social media to consume, share, and discuss news. According to the Pew Research Center's Journalism Project, more than 50% of Facebook and Twitter users consume news on social media sites (Mitchell 2015). Ideological polarization of news outlets is extensively reported in prior literature (Iyengar and Hahn 2009). Recent studies on Facebook (Bakshy, Messing, and Adamic 2015) and Twitter (Conover et al. 2011) have revealed polarization in the consumption and distribution of political news on social media sites. Political news polarization in online social media is partly explained by the selective exposure theory (Sears and Freedman 1967) which suggests that users tend to be attracted to, consume, and share news that conform with their ideological beliefs. It is also, in part, explained by user preference for homophilous social interactions (Knoke 1990).

Set in this context, we present a method to measure ideological slant of individual news articles by monitoring their consumption on online social media. Specifically, our proposed method measures an article's slant by analyzing the users tweeting about the article and analyzing their connectivity to a set of Republican and Democratic landmark users. Landmark users are well-recognized individuals whose slants (or political affiliations) are known with high confidence. A news article is assessed to have Republican slant if it is shared more by users who follow Republican landmarks than users who follow Democratic landmarks.

Similarly, a news article is assessed to have Democratic slant if it is shared more by users who follow Democratic landmarks than users who follow Republican landmarks. Our method exploits the massive amount of news consumption data available on online social media to measure news slant at a large scale. This approach—while seemingly simple—is powerful enough to effectively gauge article slant while overcoming scalability limitations of prior research.

Prior work used crowdsourcing to label political slant of news articles (Budak, Goel, and Rao 2015). However, large-scale news slant estimation using crowdsourcing requires the availability and monetary compensation for a large number of politically informed crowd workers. We compare our news article slant estimates with prior work (Budak, Goel, and Rao 2015) that relies on crowdsourcing. The comparison shows that our method can accurately estimate article-level slant for 80% Democratic-leaning news articles and 76% Republican-leaning news articles.

Related Work

Prior methods for estimating slant can be broadly divided into two categories: content-based methods and audience-based methods.

Content-based methods, as the name implies, address the problem more directly by analyzing the content of news articles. For example, (Groseclose and Milyo 2005) measured media slant by monitoring the relative citation frequency of various policy groups by news outlets and the members of Congress. As another example, (Gentzkow and Shapiro 2010) parsed congressional speeches to identify partisan phrases that are more commonly used by Republican or Democratic members of Congress. They then analyzed text content in newspaper articles and quantified each newspaper's political slant by measuring the relative use of partisan phrases. More recently, researchers have used crowdsourcing to estimate slant of individual news articles. (Budak, Goel, and Rao 2015) used crowd workers on Amazon Mechanical Turk (AMT) to identify slant of individual news articles. Unfortunately, content-based methods that utilize crowdsourcing for slant measurement do not scale well because they require non-trivial monetary compensation to crowd workers for large-scale labeling. In contrast, our proposed method does not require explicit user participation to measure slant of news articles.

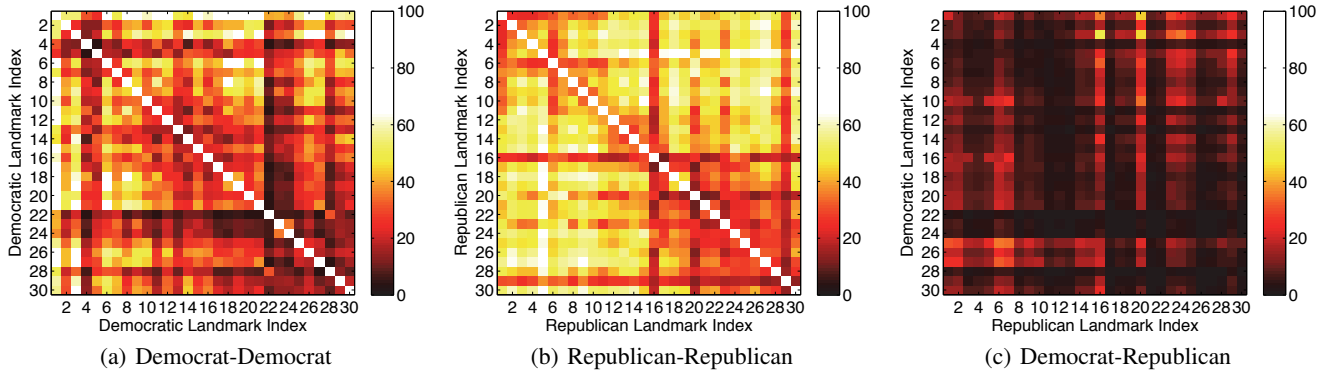


Figure 1: Visualization of follower graph similarity among Democratic and Republican landmark users. The value of a cell is the percentage overlap between followers of two landmarks. Brighter colors represent more similarity and darker colors represent less similarity.

Audience-based methods, in contrast, rely on the idea that news readers have their own ideological biases which are reflected in their news consumption and sharing behavior (Sears and Freedman 1967). These methods analyze data about news consumption and sharing to indirectly measure slant of news articles. The widespread usage of online social media allows audience-based approaches to operate at a large scale. Our proposed slant measurement method falls in this latter category. Prior audience-based methods (Gentzkow and Shapiro 2010; An et al. 2012; Wong et al. 2013; Morgan, Lampe, and Shafiq 2013; Golbeck and Hansen 2014; Le et al. 2017) aim to measure slant of individual users or news outlets in aggregate. Unlike prior audience-based methods which are dependent on manually labeled ground truth (Zhou, Resnick, and Mei 2011) or self-reported user political alignments (Bakshy, Messing, and Adamic 2015), our method requires only a small set of pre-labeled landmarks. Our proposed method leverages a small number of landmarks to accurately estimate slant of individual news articles at scale.

Method

We want to measure the political slant of news articles (towards Republicans or Democrats) by observing the patterns of sharing amongst users in online social media and the connectivity of these users to a set of landmark individuals. At a high level, our idea is that if a news article is tweeted/retweeted by more Democrats than Republicans, it is likely to have a Democratic slant. It is the opposite for articles likely to have a Republican slant. This strategy is similar to (Bakshy, Messing, and Adamic 2015), who used the ratio of number of Democrats to Republicans for estimating slant on Facebook. The key difference is that they relied on self-reported political affiliations. However, only 9% of adult Facebook users in the U.S. self-report their political affiliation. Since such data is very limited, as explained next, we propose to estimate user affiliation or slant with a different strategy involving landmark users.

Landmark Selection

A landmark user is one whose slant, here political party affiliation, is well recognized. We manually identify a set of 30 popular Democrats (e.g., Rachel Maddow) and 30 popular Republicans (e.g., Sean Hannity) on Twitter as our “landmark” users. We curated these landmarks in consultation with political scientists. We chose several journalists as landmarks because they have large followings on Twitter. On average, each Democratic landmark user has 223,656 followers and each Republican landmark user has 277,671 followers. The large following on Twitter provides us sufficient coverage to quantify slant of news articles. Our coverage analysis showed that approximately 95% of news articles in our data set have more than 10 connections to one of the 60 landmarks (median is 115). Thus, our analysis shows that 60 landmarks (each with a large Twitter following) provide us reasonable coverage.

Next, to validate our selection of landmarks, we examine the overlap in followers for pairs of landmarks. We compute the follower graph similarity (i.e., percentage overlap between their followers) for all pairs of landmarks. Figure 1(a) shows the percentage follower overlap between Democratic landmarks, Figure 1(b) shows the percentage follower overlap between Republican landmarks, and Figure 1(c) shows the percentage follower overlap across Democratic and Republican landmarks. The colorbar denotes the percentage follower overlap between two landmarks. Overall, we note that Figures 1(a) and (b) are much brighter (more overlap) than Figure 1(c). More specifically, the average percentage of follower overlap amongst Democrats is 32.4% and amongst Republicans is 43.6%. In contrast, the average percentage of follower overlap across Democrats and Republicans is only 8.8%. This pattern demonstrates political polarization on Twitter and is consistent with (Conover et al. 2011), i.e., Democrats tend to make connections with other Democrats and Republicans tend to make connections with other Republicans, while tending to not do so across party lines. Most importantly for us these results demonstrate that selective exposure theory and homophily in social interactions are exhibited in the follower connections of our land-

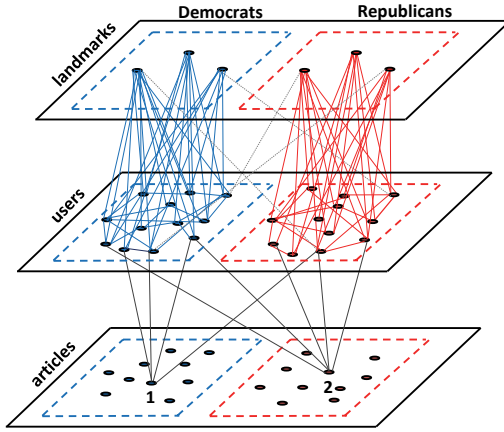


Figure 2: Our method to measure slant of news articles.

marks. This result also reaffirms our confidence in the selection of Democratic and Republican landmarks.

Slant Estimation

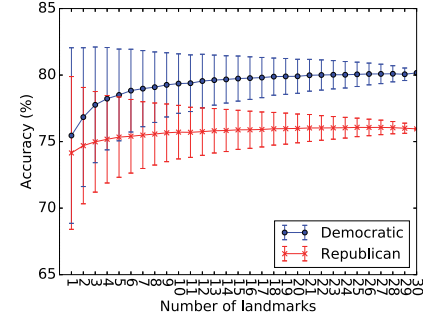
Figure 2 illustrates our proposed news slant estimation method. The top tier contains the Democratic and Republican landmarks. The middle tier contains Twitter users. The bottom tier represents news articles. Each article-user link indicates that a user tweeted/retweeted a news article. Based on the selective exposure theory (Sears and Freedman 1967), we expect users to consume and share news that conform with their ideological beliefs. Each user-landmark link indicates that a user follows a landmark. Again due to the natural preference for homophilous social interactions (Knoke 1990), we expect a large number of links from users to landmarks of the same political affiliation and only a few links to the opposite political affiliation. Thus, we can compute a news article’s political slant by monitoring its sharing activity on Twitter and user connectivity to the landmarks.

To implement our proposed method, we first collect tweets which mention a news article using Twitter’s streaming API. From this sample of tweets, we get a list of all users who have tweeted about the news article. We then count the number of landmark Democrats and Republicans which each user follows. We use Twitter’s REST API to collect the follower lists of landmark Democrats and Republicans. Using the counts of landmark Democrats and Republicans that all users follow, we quantify the political slant of a news article as: $Slant = \frac{\#Republicans - \#Democrats}{\#Republicans + \#Democrats}$. We quantify the slant of a news article in the range of -1 to 1, where -1 indicates Democratic slant and +1 indicates Republican slant.

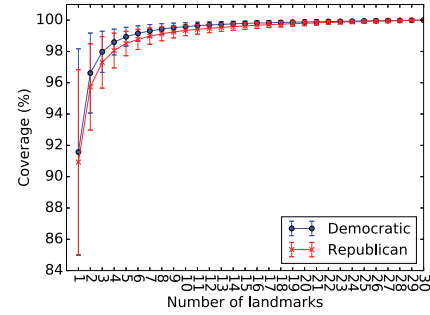
Results

Evaluation

Below we compare our article-level slant estimation results with (Budak, Goel, and Rao 2015). They used crowdsourcing to label political slant of 10,500 news articles as



(a) Accuracy



(b) Coverage

Figure 3: Impact of landmark selection on slant measurement accuracy and coverage.

Republican-leaning, Center, or Democratic-leaning.¹ They recruited two crowd workers to evaluate the political slant of each news article on a five-point scale (“Positive”, “Somewhat Positive”, “Neutral”, “Somewhat Negative”, “Negative”) for both Democrats and Republicans. To mitigate noise and increase reliability of slant labels by crowd workers, we only consider the articles for which the evaluations by two crowd workers match each other. After this filtering, the data set contains 605 Democratic-leaning, 653 Republican-leaning, and 3,837 Neutral news articles. For comparison, we compute the slant score for these Democratic- and Republican-leaning news articles using our proposed method. The results show that our estimated slant scores are well correlated with the article set’s political affiliation. The average estimated slant score for the Democratic-leaning article set is strongly Democratic-leaning (-0.40), while the average estimated slant score for the Republican-leaning set is strongly Republican-leaning (0.44). Furthermore, 80% of Democratic-leaning articles have slant scores smaller than 0 and 76% of Republican-leaning articles have slant scores larger than 0.

Impact of Landmark Selection

We next examine the impact of the selected landmarks on the accuracy (with respect to crowdsourcing) and cover-

¹We thank (Budak, Goel, and Rao 2015) for sharing their crowd labeled data set.

age (number of news articles) of slant measurement. We compute the accuracy for the sets of Democratic- and Republican-leaning articles respectively when the number of chosen landmarks is gradually increasing from 1 to 30. For this experiment, we select landmarks randomly from the full set of 30 landmarks for Democrats and Republicans. Figure 3 shows the results for the average and standard deviation of the accuracy and coverage of our slant measurements for 1,000 independent runs. Overall, we note that both average accuracy and coverage increase (while their standard deviations decrease) as we include more landmarks. It is noteworthy that the accuracy and coverage start to plateau when we reach around 10 landmarks. These results demonstrate the diminishing returns on average slant accuracy and coverage as we include more landmarks. We conclude that our full set of 30 landmarks are sufficient for achieving good accuracy and coverage for both Democratic- and Republican-leaning news articles.

Slant of Neutral Articles

Previously we showed that our estimated slant scores are well correlated with each article set’s political affiliation labeled using crowdsourcing. We now want to categorize news articles into 3 categories of Democratic, Republican, and Neutral according to their measured slant scores. To do so, we need to decide suitable thresholds a and b such that $-1 \leq a < b \leq 1$, and label the news articles with slant scores lying within (a, b) as Neutral, the news articles with slant scores in the range $[-1, a]$ as Democratic-leaning, and the news articles with slant scores in the range $[b, 1]$ as Republican-leaning. We experiment with multiple choices for the thresholds a and b . For each choice, we compute a confusion matrix of the match between crowdsourced news article slant by (Budak, Goel, and Rao 2015) and our measured slant scores. Table 1 depicts the confusion matrix for symmetric thresholds $b = -a = 0.37$ which maximize the sum along the diagonal of the confusion matrix. Table 2 depicts that confusion matrix for asymmetric threshold $a = -0.48, b = 0.28$ which maximize the sum along the diagonal of the confusion matrix.

Conclusion

We presented a method to scalably measure the political slant of news articles. We estimated the slant of individual news articles by analyzing the social connectivity of users who tweet about them. Our proposed method relies on a small set of Democratic and Republican landmarks on Twitter. The results showed that our method accurately identifies slant of individual news articles.

References

An, J.; Cha, M.; Gummadi, K. P.; Crowcroft, J.; and Quercia, D. 2012. Visualizing Media Bias through Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.
 Bakshy, E.; Messing, S.; and Adamic, L. 2015. Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348(6239):1130–1132.
 Budak, C.; Goel, S.; and Rao, J. M. 2015. Fair and Balanced? Quantifying Media Bias through Crowdsourcing Content Analysis.

Crowdsourced Slant Label	Democratic	Neutral	Republican
Democratic-leaning	62.5%	26.8%	10.7%
Neutral	34.1%	46.0%	19.8%
Republican-leaning	10.7%	23.6%	65.7%

Table 1: Confusion matrix of the match between crowdsourced news slant measured by (Budak, Goel, and Rao 2015) and our measured slant scores when the threshold is symmetric ($b = -a = 0.37$).

Crowdsourced Slant Label	Democratic	Neutral	Republican
Democratic-leaning	55.9%	31.6%	12.5%
Neutral	26.5%	50.5%	23.0%
Republican-leaning	8.4%	22.8%	68.8%

Table 2: Confusion matrix of the match between crowdsourced news slant measured by (Budak, Goel, and Rao 2015) and our measured slant scores when the threshold is asymmetric ($a = -0.48, b = 0.28$).

In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

Conover, M. D.; Ratkiewicz, J.; Francisco, M.; Goncalves, B.; Menczer, F.; and Flammini, A. 2011. Political Polarization on Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

Gentzkow, M., and Shapiro, J. M. 2010. What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica* 78(1):35–71.

Golbeck, J., and Hansen, D. 2014. A method for computing political preference among twitter followers. *Social Networks*.

Groseclose, T., and Milyo, J. 2005. A Measure of Media Bias. *The Quarterly Journal of Economics* 120(4):1191–1237.

Iyengar, S., and Hahn, K. S. 2009. Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication* 59.

Knoke, D. 1990. Networks of Political Action: Toward Theory Construction. *Social Forces* 68(4):1041–1063.

Le, H.; Boynton, G.; Mejova, Y.; Shafiq, Z.; and Srinivasan, P. 2017. Revisiting The American Voter on Twitter. In *ACM Conference on Human Factors in Computing Systems (CHI)*.

Mitchell, A. 2015. State of the News Media 2015. <http://www.journalism.org/2015/04/29/state-of-the-news-media-2015>.

Morgan, J. S.; Lampe, C.; and Shafiq, M. Z. 2013. Is News Sharing on Twitter Ideologically Biased? In *ACM Conference on Computer-Supported Cooperative Work (CSCW)*.

Sears, D. O., and Freedman, J. L. 1967. Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly* 31(2):194–213.

Wong, F. M. F.; Tan, C. W.; Sen, S.; and Chiang, M. 2013. Quantifying Political Leaning from Tweets and Retweets. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.

Zhou, D. X.; Resnick, P.; and Mei, Q. 2011. Classifying the Political Leaning of News Articles and Users from User Votes. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*.