Exploring Language in News Sources: Does Political Leaning Affect the Sentiment of Words?

Team SaaS
apetteru, sdsouza1, slow, arothber

HYPOTHESIS

In light of the upcoming 2020 election, we wanted to take a deeper look into how our political leanings affect the types of words we use. Articles in our exploratory research brought forward the notion that conservatives tend to be more negative because they are less accepting of change and more affected by jarring realities. We tested whether **conservative news sources have a greater proportion of articles with negative words such as "fail", "hate", "critic", and "scare" than liberal news sources.** We chose the words by looking at the 5,000 words that occur most frequently across the articles and choosing those with a negative sentiment.

DATA

We found a political news dataset containing 87,157 English news articles dealing with politics, from which we randomly generated a sample of 5,000 articles. Our sample contained articles from 10 news sources, and we determined how conservative or liberal they were using a media bias chart and assigning each a "bias score" from -2 (extremely liberal) to +2 (extremely conservative).

FINDINGS

Claim #1: There is no evidence to suggest that conservative news sources have a greater proportion of articles with negative words.

Support for Claim #1: We performed a multivariable regression using the StatsModels OLS package. One reason we chose this type of analysis is that there seemed to be more liberal articles in our dataset than conservative articles, and rather than "throw" them out, it allowed us to account for this bias in the data with the constant term in the OLS regression. While we did not want to count multiple instances of the word "hate" in an article, for example, we did think that the existence of multiple target words suggests a stronger indication of article tone. Regarding the effect magnitude and direction as shown in Figure 1, the direction of the effects are exactly opposite to what our alternative hypotheses predicted. The negative coefficients for our test words indicated that the presence of these words was correlated with liberal articles. The coefficients for the variables hover right around -0.3. However, the domain of the dependent variable bias_score ranges from -2.0 to 2.0 inclusive. This means that an effect size of magnitude 0.3 was not a particularly compelling value.

Dep. Variable:	bias_score	**R-squared:**	0.029
Model:	OLS	**Adj. R-squared:**	0.027
Method:	Least Squares	**F-statistic:**	11.37
Date:	Tue, 05 May 2020	**Prob (F-statistic):**	7.85e-11
Time:	12:58:54	**Log-Likelihood:**	-2660.9
No. Observations:	1878	**AIC:**	5334.
Df Residuals:	1872	**BIC:**	5367.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.3103	0.025	-12.180	0.000	-0.360	-0.260
hate	-0.3378	0.098	-3.430	0.001	-0.531	-0.145
fail	-0.1578	0.061	-2.590	0.010	-0.277	-0.038
critic	-0.2040	0.040	-5.054	0.000	-0.283	-0.125
scare	-0.3177	0.179	-1.775	0.076	-0.669	0.033

Figure 1: *OLS Regression Test Results*

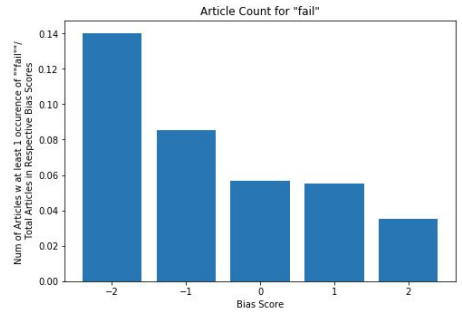

Figure 2: *Bias distribution for the word "fail"*

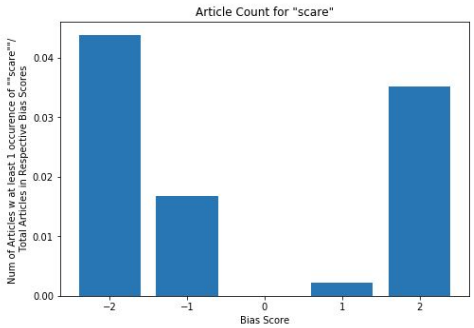

Figure 3: *Bias distribution for the word "scare"*

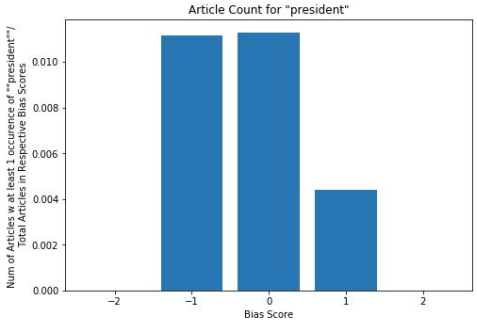

Figure 4: *Bias distribution for control word "president"*

This led us to fail to reject the null hypothesis in favour of the alternative hypothesis.

Claim #2: Liberal articles have a more negative tone than conservative articles

Support for Claim #2: We performed a sanity check test using the NLTK Vader SentimentIntensityAnalyzer. We identified the net positivity (positive sentiment - negative sentiment) of each political article, multiplied this term by the article's bias score, and took the mean of all liberal articles, all conservative articles, and all articles for this score. This functioned as a sanity check, not meant to directly test the hypothesis or claim that negativity is a direct proxy for it. We expected that net_positivity would be negatively correlated with anger and fear, and so based on our exploratory research we expected net_positivity to be positive for liberal articles and negative for conservative articles. Given that the bias score assigned to liberal articles was negative and conservative articles, positive, a negative net_positivity * bias score would have supported our exploratory research.

Across the dataset, the overall bias_score * net positivity is positive, with both liberal and conservative article groups having positive means. Thus, in our data, liberal articles had a net negative tone, and conservative articles had a net positive tone.

	Net positivity * bias score
Liberal articles	0.14
Conservative articles	0.03