

Please submit ≥ 1 paragraph of an update of what you've tried on your project so far. The goal of this assignment is to prompt some teams to meet with the teaching staff, and to help give directions or guidance. It is not meant to be a strict evaluation or to adversarially impact any grades. Your grade on this assignment will be 0/1/2: 0 = not submitted, 2 = submitted something reasonable, 1 = minimal things done so far, or the update was very uninformative.

Since our proposal our group has continued to do exploratory analysis on the climate change media mentions using Time Series Analysis, Topic Analysis, and Sentiment Analysis.

Time Series Analysis

For the time series analysis, we plotted the instances of climate mentions and datasets on U.S. natural disasters from 2010 to 2019 in order to compare the two. First, we compared annual mentions of climate change to an annual dataset of billion dollar disasters. Then we compared it to another dataset with all US Natural disaster declarations, that has data at a more granular timescale. This enabled us to investigate monthly disaster rates, both the total disasters and disasters segmented by type, against monthly CNN network references to counts of climate change mentions. We noticed that the trends did not appear to overlap, in that monthly disaster data from 2010 to 2019 did not mimic the trend of monthly climate references on CNN. For our next step, we will investigate a granular daily trend in Fox, CNN, and MSNBC data against granular disaster rates. We will investigate what steps are necessary to generate stationary data and remove possible yearly seasonality from the data.

Topic Analysis

Regarding the topic analysis, we use coherence scores to rate the quality of a range of numbers of topics; the number of topics that yields the highest coherence is 10. We then fix our number of topics and group the text snippets by month and by channel. We quantify monthly topic relevance by counting the number of times that the top 5 words associated to each topic appear in the monthly snippet, then divide by the total number of words to produce a frequency. We have plotted by both topic and by channel and noted a few interesting phenomena. First, we note that CNN has an order of magnitude spike in mentions of a topic characterized by the words “talk, scientist, sensor, road, tree” around 2013. Other than this, topic mentions are highly correlated across channels, with few exceptions. We plan to test the hypothesis that certain topics, such as one seemingly related to American leadership (characterized by words such as “people want american lot like”) have increased in frequency before versus after the change in administration.

Sentiment Analysis

_____ We begin our sentiment analysis module by breaking up the entire corpus by TV network (i.e. CNN, Fox News, MSNBC). We are interested in understanding the sentiment polarity and subjectivity surrounding climate change discussion based on TV network. We note that there are more snippets for Fox News and MSNBC than CNN, so we use the ratio of the number of positive sentiment snippets / negative sentiment snippets to compare them appropriately. We

use a pre-trained Naive Bayes Classifier offered by the [TextBlob NLP Python package](#) to classify the snippets across the entire corpus into Positive, Neutral, and Negative classes. We find the respective Positive/Negative ratios to be: **CNN: 3.582; Fox News: 2.416; MSNBC: 3.343**. Our main finding from this is that across the networks, Fox News had the lowest positive / negative ratio indicating that as a network, they had relatively the most negative commentary surrounding climate change topics.

Additionally, we used a pre-trained Naive Bayes subjectivity classifier also provided by TextBlob to classify the snippets as being in objective or subjective classes. We report the Objective / Subjective ratios for each of the networks across the entire corpus:

CNN: 1.331; Fox News: 1.418; MSNBC: 1.44. The difference between these ratios isn't as large as was the case with sentiment, but we notice the ratio of objective snippets / subjective snippets is lower for CNN than it is for Fox News and MSNBC suggesting that CNN has relatively more subjective snippets than the other networks.

Due to the fact that our snippets are not independent from one another (dependence over time), we will not use hypothesis testing as was originally stated in the project proposal, although we are still interested in understanding the change in sentiment and subjectivity over time around climate change coverage so we will instead apply time series analysis techniques to understand this phenomenon.

Time Series Analysis

- Investigated monthly disaster rates (total and by disaster type) against monthly CNN network reference counts of climate change mentions (<--will do Fox, MSNBC, should be same code)
- Compiled Annual Counts of Climate Change mentions and graphed against annual instances of billion dollar disasters (by total disasters and by disaster type)
- The trends did not appear to overlap (interestingly enough)
- Hypothesis: top news networks in American English-speaking TV are too focused on political associations/topics that come up in reference to climate change, rather than associating climate change communication with meteorology/natural disasters
 - This tracks with none of the top 10 topics in the topic analysis having to do with weather/disasters
 - The climate change mentions may also track with disasters on a smaller timescale (monthly rather than annual) so we are planning to look at a separate dataset
- Save for later: For now, paired with our topic analysis finding that none of the top 10 topics are related to weather or natural disasters, we hypothesize that it's possible that top news networks in American English-speaking TV are too focused on political associations/topics that come up in reference to climate change, rather than associating climate change communication with meteorology/natural disasters. Further investigation into more granular data will allow us to probe this hypothesis.