Project Proposal: Quantify the Political Narrative Framing in News Media Reporting

PPOL 6801, Spring 2024
Prof. Tiago Ventura

Zhiqiang Ji

Introduction

In the age where the term “fake news” has entered common parlance and “alternative truths” obscures the line between fact and fiction, public trust in news media stands on precarious ground. While prestigious news media platforms have been putting great effort to avoid distortion and mistake about facts, they may craft narratives with subtle political biases, conveying different sentiments out of the same facts. The ideological discrepancy among news media has become a widely recognized fact. Disbelieving in the objectivity of news media and indulged by the ease to get free news headlines, the audience adapt to care more about how the news stories make they feel rather than the authenticity of the facts behind the reporting to consume more content from outlets that align with their political leanings. Consequentially, news media may be more driven to align their ideological stance with their separate audience bases.

We have reason to be concerned that the expected authenticity, objectivity, neutrality, and comprehensiveness of news media will be further undermined, ultimately harming the democratic system. To address this tendency, this study tries to answer the following question. Assuming most prestigious news media platforms are covering similar major news stories at given period of time, how do they consistently attach their political leaning to the reporting of facts, maintain their ideological images, and frame their narratives? Can such framing be quantitatively portrayed? More importantly, has this politically-loaded narrative framing intensifying in sync with the gradual polarization of the political climate in the past decades?

This study is inspired by (Rozado & al-Gharbi, 2022) and will employ word embedding methods to quantify how news reports from different media outlets associate positive or negative sentiment words with political terms related to different parties, such as prominent party figures and key issues, in order to construct their ideologically-oriented discourse. By calculating the projections of sentiment words onto semantic axes representing political orientations in the embedding space, this study aims to reveal the latent political biases embedded in the language used by various media organizations.

Furthermore, this study will segment the corpus into different time periods to investigate how these sentiment associations change over time. By training separate embedding models on news articles from different decades and comparing the political sentiment associations across these models, we can observe how the political divergence among major media outlets has evolved over the past forty years up to the present day. This diachronic analysis will provide valuable insights into the long-term trends of media polarization and how the ideological landscape of the media ecosystem has shifted in response to major socio-political events and transformations.

Data

This study will start with utilizing the preprocessed dataset (Rozado & al-Gharbi, 2021) curated for their paper. This dataset consists of a vast corpus of news and opinion articles from 47 prominent U.S. media outlets, spanning from 1975 to 2019. The dataset, comprising over 27 million articles, was collected from the outlets' online domains and public cache repositories, such as Google Cache, the Internet Wayback Machine, and Common Crawl. The articles have been cleaned and processed to extract the relevant textual content, including headlines and main text, while removing extraneous elements and noise.

To extend the temporal scope of the analysis and include more recent data, this study will additionally acquire a dataset of cable news transcripts from six major networks: MSNBC, Fox News Network, CNN, NPR, CBS, and CNBC, covering the period from 2019 to the present day. These transcripts will be obtained from databases such as LexisNexis or Factiva, which provide comprehensive collections of news media content. The inclusion of these additional sources will allow for an up-to-date examination of political sentiment associations in the rapidly evolving media landscape. These two databases are accessible via Georgetown University's Library website.

Method

This study will follow the methodology of Rozado & al-Gharbi (2022) to quantify media bias by constructing semantic axes representing political orientations and projecting sentiment words onto these axes. Word2vec embedding models will be trained on the news articles and transcripts, and vector arithmetic will create axes for political parties, figures, and issues. Cosine similarity will measure associations between sentiments and political orientations.

A dictionary of political issue topics will be compiled to investigate their projections onto party affiliation axes, testing the hypothesis that media outlets allocate attention differently when framing ideological narratives. Supervised topic modeling will validate these findings.

Fine-grained sentiment classification, using lexicons like LIWC or NRC, will examine the differential application of emotions (anger, fear, sadness) by political parties and their associations with specific issues.

If time allows, a pre-trained NER model (Raza, 2023) will detect biased language in a sample of transcripts, exploring whether certain outlets employ a characteristic vocabulary contributing to bias.

By combining these approaches, this study aims to provide a comprehensive investigation of media bias and polarization, revealing the interplay between sentiment, political orientation, issue framing, and linguistic choices in shaping ideological narratives.

Expected Results

1. Word embedding analysis will reveal significant associations between sentiment words and political orientations in news media discourse, with left and right-leaning outlets showing opposite patterns.
2. Projection of political issues onto party affiliation axes will show differences in attention allocation, supporting the hypothesis that media outlets selectively emphasize issues to construct ideological narratives.

3. Fine-grained sentiment analysis will uncover nuanced differences in the use of emotions by political parties and their associations with issues, providing insight into emotional framing and its role in shaping public opinion.
4. If conducted, exploratory NER analysis may identify biased language patterns characteristic of certain outlets, further supporting the notion of active construction of ideological narratives through linguistic choices.

Reference

David, & Musa. (2021). Data for manuscript: "Using Word Embeddings to Probe Sentiment Associations of Politically Loaded Terms in News and Opinion Articles from News Media Outlets" [dataset]. Zenodo. https://doi.org/10.5281/zenodo.4797464

Raza, S. (2023). Navigating News Narratives: A Media Bias Analysis Dataset (arXiv:2312.00168). arXiv. http://arxiv.org/abs/2312.00168

Rozado, D., & al-Gharbi, M. (2022). Using word embeddings to probe sentiment associations of politically loaded terms in news and opinion articles from news media outlets. Journal of Computational Social Science, 5(1), 427–448. https://doi.org/10.1007/s42001-021-00130-y