Multidimensional Analysis on the Political Spectrum of Media Ideologies

Yilun Liu, Ke Chen

Problem & Motivation

The contemporary field of information and communication has taken on a landscape in which many have become accustomed to being exposed only to ideas they already agree with, which has made news media, along with their target readerships turning greatly polarized and biased for a long time. Once such filter bubbles form, people tend to get caught up in it unconsciously, becoming limited in their ability to understand each other and easily manipulated into thinking and voting.

To liberate people from filter bubbles, making the opinions of different media transparent could greatly help the general public to acess a wider variety of information, to easily identify different perspectives, to avoid being influenced by ideology biases, to better understand each other and to make wiser decisions.
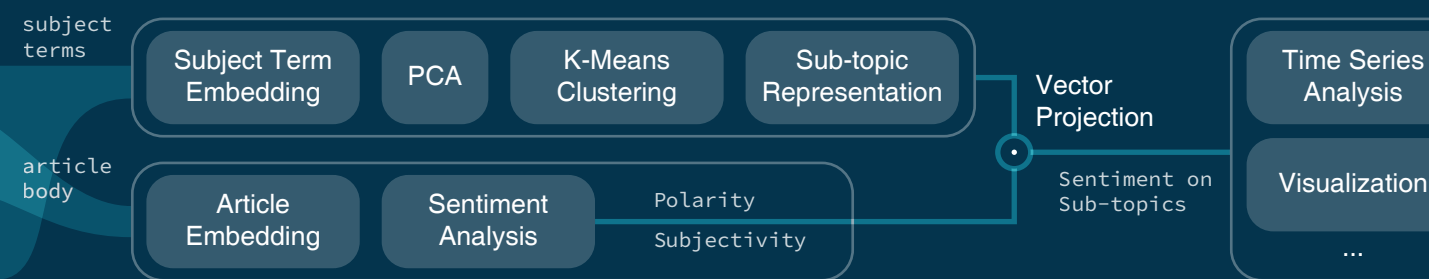
Research Question

What are the positions held by media on different dimensions of the political spectrum, each of which regarding different social issues?

Is there any temporal shift in such positions, representing how different social groups' attitudes towards different social issues has changed across time?

Data

Articles on topic "Climate Change" for each month from 2018 to 2022, from two media outlets *The Daily Mail and Mail on Sunday (London)* and *The New York Times*, which were considered correspondingly right-leaning and left-leaning on average. All articles were downloaded from Nexis Uni ®.

Framework

subject terms

Article

Article Corpus

article body

Subject Term Embedding → PCA → K-Means Clustering → Sub-topic Representation → Vector Projection

Article Embedding → Sentiment Analysis → Polarity / Subjectivity

Sentiment on Sub-topics

Time Series Analysis

Visualization

...

Methodology

Two pre-trained NLP frameworks, Sentence-Transformers (BERT based, state-of-the-art) and TextBlob (NLTK based), were adopted for this project.

For each single article, sentence embedding and sentiment analysis was performed first to obtain its vector representation in a 384-dimensional space, as well as its polarity (-1 ~ +1) and subjectivity (0 ~ 1) scores.

For the entire article corpus, a list of significant subject topic terms was obtained based on the statics of metadata accompanying the articles. This is followed by a series of unsupervised representation learning processes, consisting of sentence embedding, Principle Component Analysis and K-Means clustering, which enables the system to automatically decide and calculate the vector representations for different sub-topic clusters.

Afterwards, based on the ability of sentence embeddings to capture semantic differences through canonical distance metrics, by projecting the vectors of each article onto the corresponding feature vectors of different sub-topic clusters, we could finally obtain the relative sentiment representation components of each article in different sub-topic dimensions.

Such data serves well for various downstream tasks, including time-series analysis and visualizations.
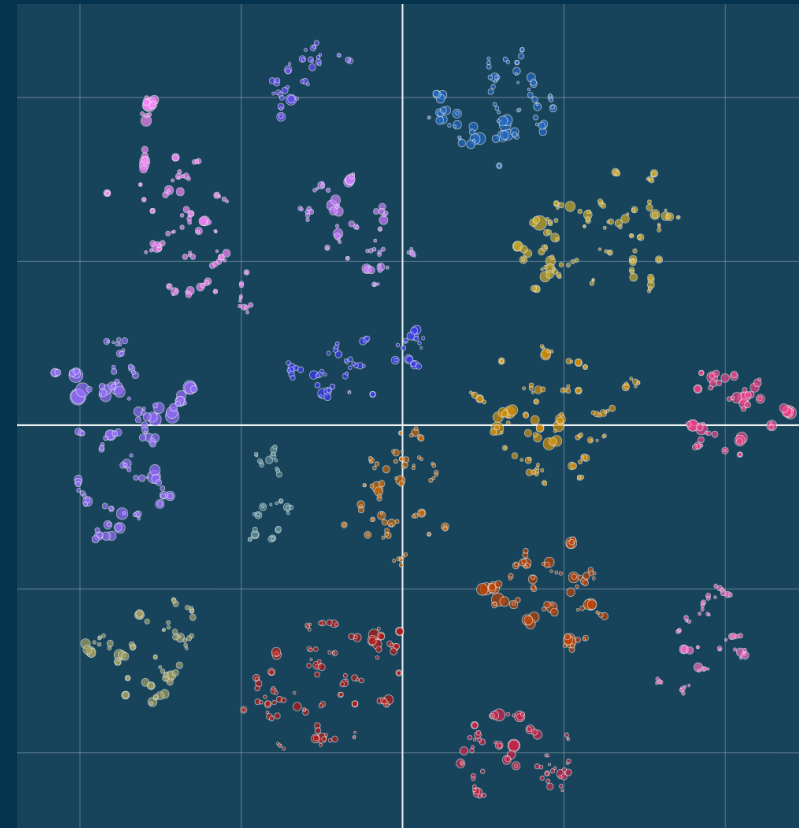
- Mountains, Rivers, Parks & Playgrounds
- Us Presidential Candidates 2020, Political Parties, Elections & Politics
- United Nations, Agreements, United Nations Institutions
- Economy & Economic Indicators, Economics, Taxes & Taxation
- Weather, Extreme Temperatures, Severe Weather
- Executives, Business News, Managers & Supervisors
- Science & Technology, Writers, Research Reports
- Climatology, Wildlife, Mammals
- Renewable Energy Industry, Energy & Utility Policy, Electric Vehicles
- Food Crises & Shortages, Meat Free Diets, Agriculture Departments
- Government & Public Administration, Legislative Bodies, Public Policy
- Society, Social Assistance & Lifestyle, Cities, Regional & Local Governments
- College & University Professors, Students & Student Life, Colleges & Univers
- Climate Change, Emissions, Global Warming
- Infectious Disease, Coronaviruses, Viruses
- Negative News, Negative Personal News, Negative Misc News

Results

As can be observed, it is true that the chosen two media outlets did convey different sentiments on multiple sub-topics, and it is also true that these positions they took have shifted over time.

The Daily Mail and Mail on Sunday (London)

year
- 2022
- 2021
- 2020
- 2019
- 2018

The New York Times

year
- 2022
- 2021
- 2020
- 2019
- 2018

Polarity Projection
−0.02   0   0.02   0.04   0.06