News Bias Detection

Zero Shot Zealots: Albert, Sandesh, Prathit

Previous Solutions

- [Stanford NLP Model News Bias Detection](#)
- [Multi-Bias Detection with LLMs](#)
- [Recent Article on Media Bias Detection with DL](#)

Datasets

HEADLINE ROUNDUP

US Inflation Cools to 2.8% in February

From the Center	From the Right	From the Left
Inflation rate hits 2.8% in February, less than expected	Inflation slowed slightly to 2.8% in February ahead of Federal Reserve meeting	US inflation cooled in February, but Trump's tariff plans and trade war loom
CNBC L L C R R	Fox Business L L C R R	CNN Business L L C R R

Qbias Dataset from IR Group

- GitHub Repo released 2023 with:
  - [Pre-collected data](#) from AllSides
  - [Scraper](#)
- All based on All-Sides labels from professionals
  - Each article labeled as left, right, or center
- Dataset creation paper: [Qbias - A Dataset on Media Bias in Search Queries and Query Suggestions](#)

Data Exploration

Allsides_balanced_news_headlines-texts.csv

- Indexed CSV file with columns for title, tags, heading, source, text, and bias rating
- 21,747 training examples
    - 10,273 articles tagged as left, 7,222 tagged as right, 4,252 tagged as center
    - We could scrape more if we needed using their scraping tool

Data Cleansing

- The given dataset is quite simple and neat
- Cleansing further may involve one-hot encoding for source and tokenizing the title, tags, heading, and text

Data Preparation

- Label Y is the prediction "Left, Right, or Center", and features are all other components like Title and Text.
- Limited data cleansing and Text Normalization by converting words to lowercase
- <u>Potential strategy:</u> Text processing techniques like stemming and grouping all the similar words together.

Challenges and Future Tasks

- Experimenting with different tokenization methods using different base models. BERT, Llama, DeepSeek...
- Figuring out parameters for fine-tuning our model