Name: Michael Tamburello
NetID: mjt11
Date: October 25, 2023

Project Proposal
Visualized Media Bias and Polarization Detection

Team Members

Name (NetID) - Role
Michael Tamburello (mjt11) - Captain

What is your free topic?

A platform leveraging cutting-edge NLP techniques to analyze and uncover pervasive bias in popular news outlets across diverse political landscapes. Providing end users with visualizations and comparisons that are easily digestible and enlightening.

What is the task?

Create a sophisticated platform that employs state-of-the-art NLP algorithms to uncover ideological bias and narrative divergence in major media channels, amplifying contrasting viewpoints on the same current events.

Why is it important or interesting?

In an age marked by extreme media polarization, this platform seeks to shed light on the undue mixing of opinions, assumptions, and hypotheticals in news reporting. The goal is to elevate the quality of public dialogue and reinforce the foundations of democratic society.

What is your planned approach?

- Incorporate spaCy's Named Entity Recognition to pinpoint key subjects and actors in news articles.
- Leverage spaCy's Part-of-Speech tagging and Dependency Parsing for intricate sentence analysis, illuminating potential bias.
- Utilize spaCy's advanced text classification features to train a machine learning model that quantifies media bias.
- Employ spaCy's similarity metrics to juxtapose articles from disparate outlets, thereby revealing journalistic dissonance.
- Design an embedding projection visualization to spatially represent word biases in a multi-dimensional vector space which highlights stark differences found between sources on identical events.

What tools, systems or datasets are involved?

Core NLP Framework: spaCy
Machine Learning Framework: PyTorch
Datasets: Curated collections from NewsAPI and the Media Bias/Fact Check database

What is the expected outcome?

A sophisticated platform capable of discerning and quantifying media bias and dissonance.
Dynamic visualizations that reveal the ideological landscape of media language.

How are you going to evaluate your work?

- Bias Detection
    - Metrics including Precision, Recall, F1-score, and ROC-AUC will be used to evaluate the model's ability to accurately identify bias.
- Dissonance Measurement
    - Custom metrics rooted in cosine similarity algorithms will be used to quantify narrative divergence between news sources.
- User Feedback
    - Subjective evaluations collected from users will serve as additional indicators of the platform's efficacy and impact.

Which programming language do you plan to use?

Programming Language: Python

Workload Justification

- Data Procurement and Cleansing: 5 hours
- Text Preprocessing Pipeline with spaCy: 5 hours
- Model Building/Fine-tuning and Hyperparameter Optimization with PyTorch: 5 hours
- Embedding Visualization Design and Development: 5 hours
- Frontend and User Experience Design: 5 hours
- Documentation and Evaluation: 5 hours

Overall Time Commitment: 30 Hours