Leveraging Language Models for Temporal Political Bias Analysis

Syntax Errors: *Clayton Carlson, Ryan Diaz, Charlie Rapheal, Sanjali Roy*

Introduction

Previous work has been done for detecting political bias in media sources on an individual level, but this work has not been applied to detect bias levels across time. Can language models be trained to capture patterns in how the political landscape shifts over time? More importantly, do these patterns line up with our intuition of political and economic events?

Literature Review

Political bias detection in text has been done in varying scales of granularity, such as sentence-level, article-level, and network-level bias. As discussed in Baly et al. (2020) [1], article-level bias classification runs the risk of the model overfitting to the media source rather than analyzing the article itself, and so they present several techniques to "de-bias" these models such as adversarial adaptation and triplet loss pretraining.

Proposed Ideas and Methods

- To analyze the trends of political leanings over time, we applied articles and transcripts from various news sources and authors to **politicalBiasBERT** [1], a model trained to classify articles based on political leaning (left, center, or right).
- Aligning this temporal data with real world 'benchmark events' allows for analysis of changing political leanings before and after these events.
- Benchmark events include:
  - 9/11 Attacks
  - 2008 Stock Market Crash
  - 2016 Election
  - Covid-19 Pandemic
- To explain the decisions of the classifier, we make use of Local Interpretable Model-Agnostic Explanations (LIME) [2].

Novel Contribution

We provide a starting point for performing temporal analysis of politically biased leanings of news networks and journalists using pretrained models. Additionally, we provide initial temporal analyses for several major news outlets such as CNN, Fox News, and the New York Times.
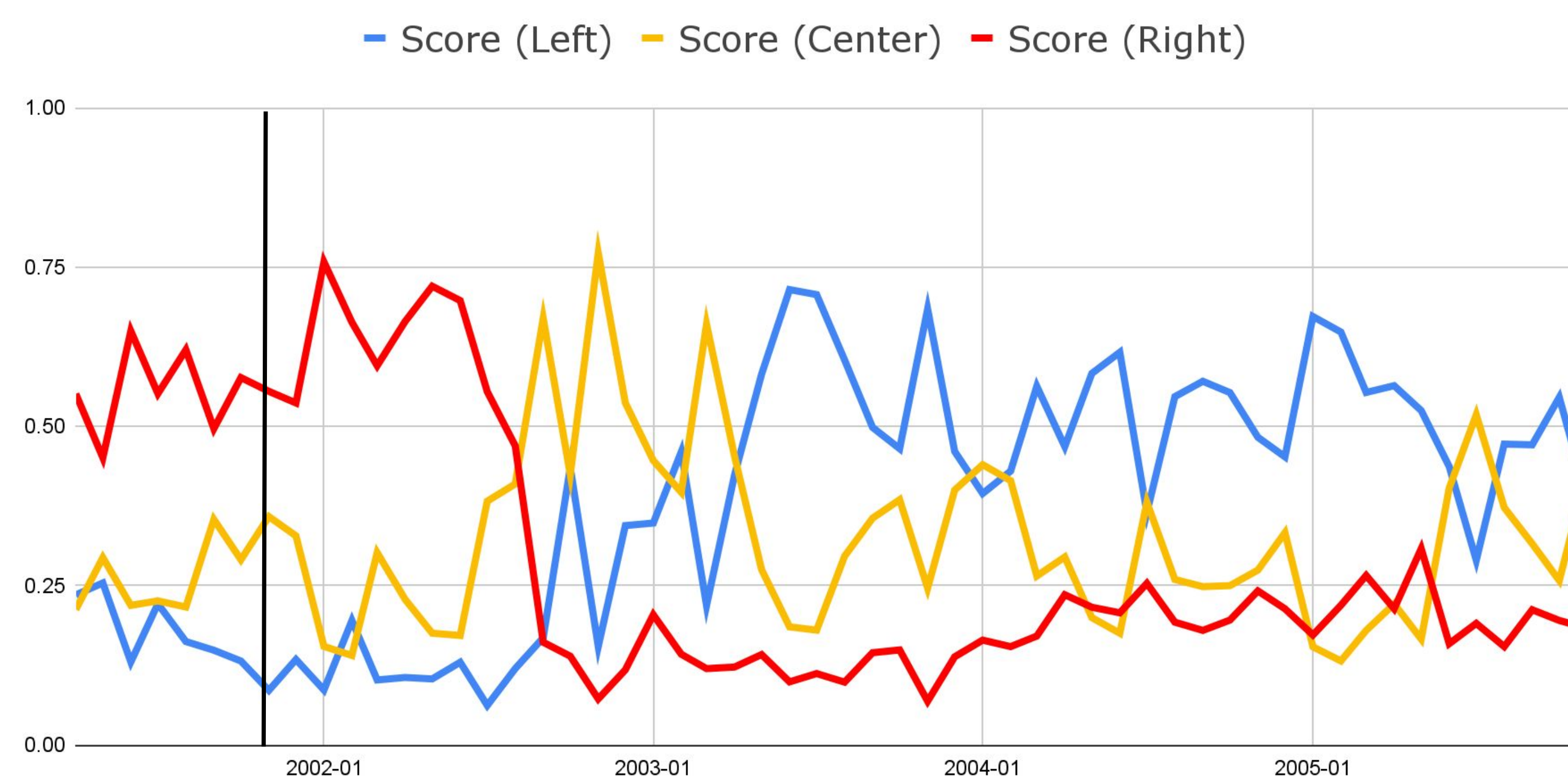
NYT Article Summary Score Ratings
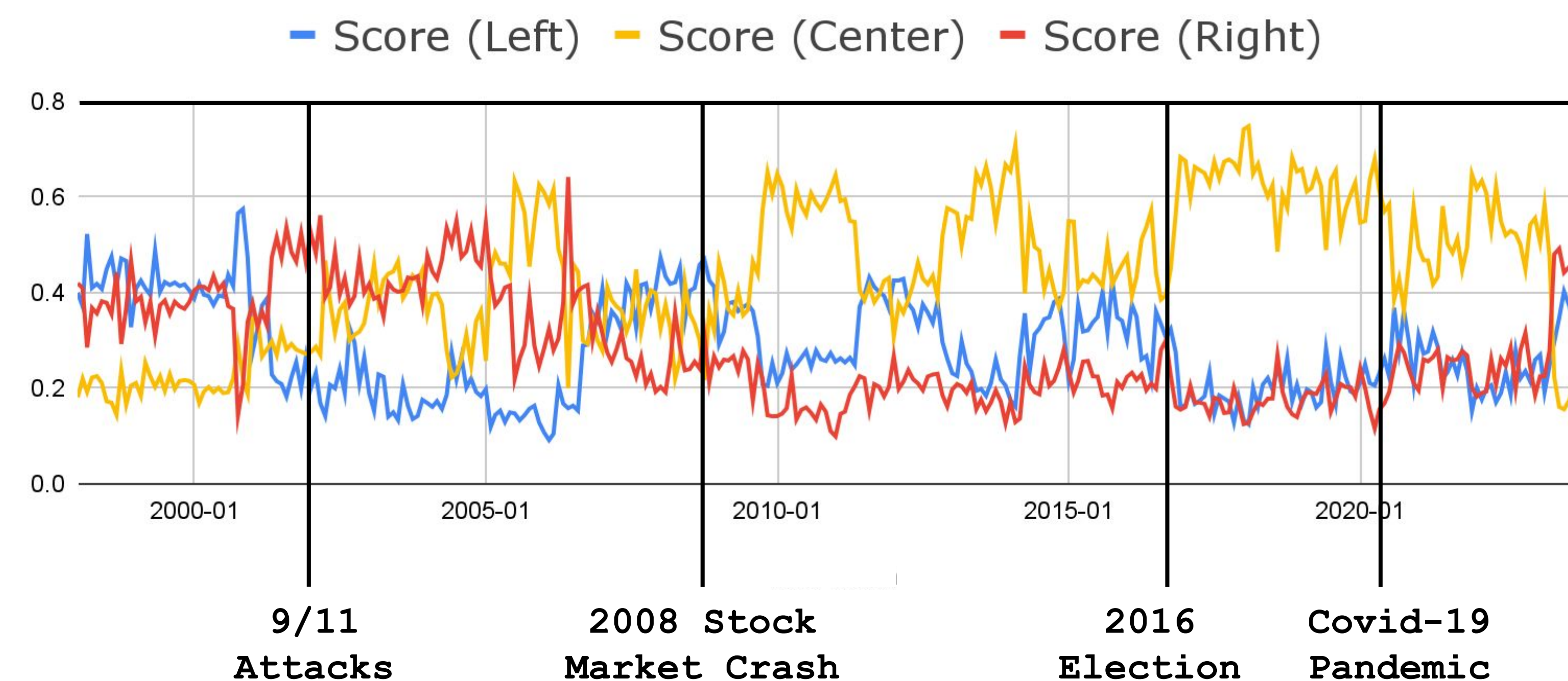
CNN Live at Dawn Score Ratings Unfiltered
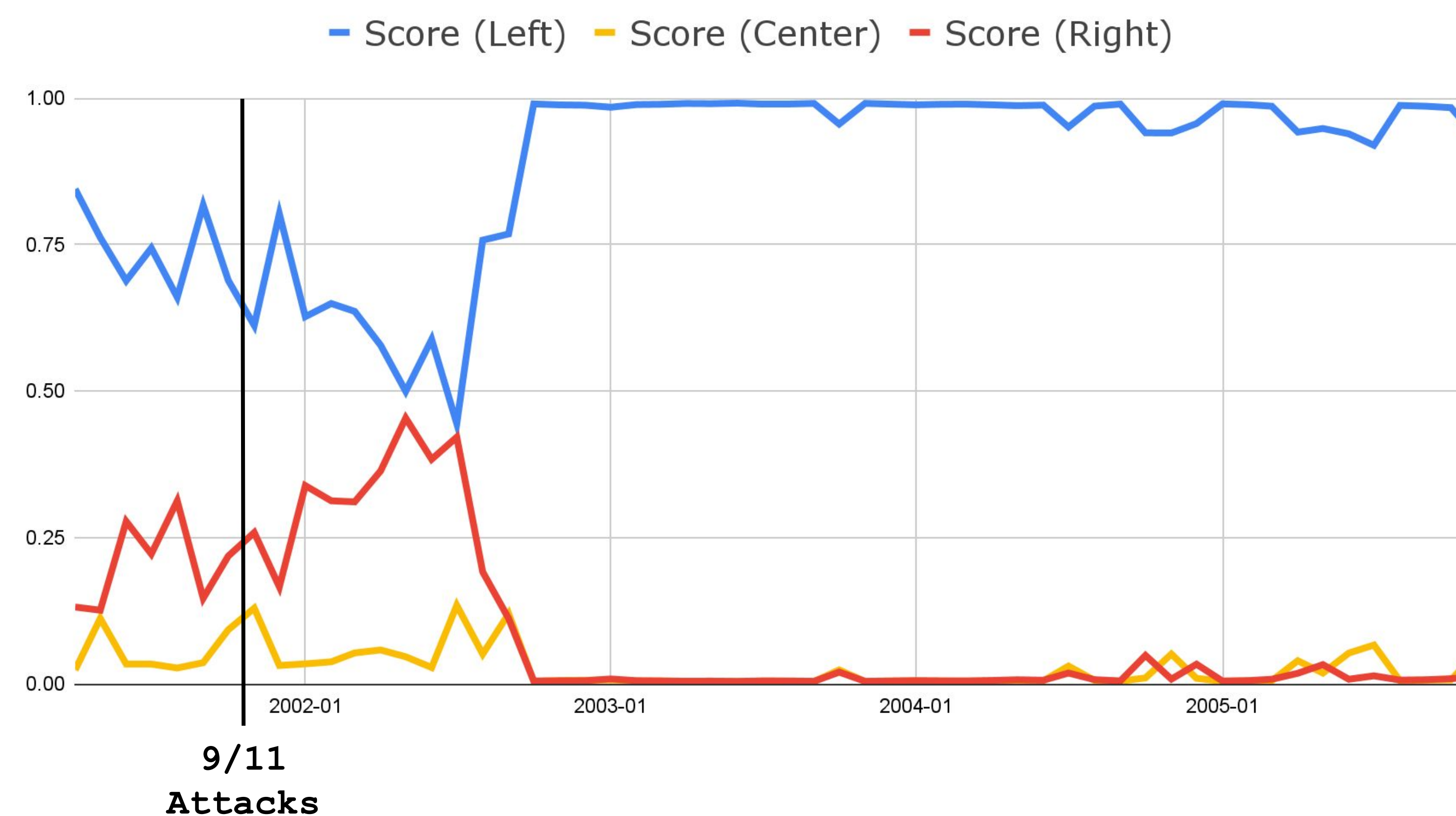
CNN Live at Dawn Score Ratings Filtered

Results and Findings

- Ran the collected articles and transcripts on **politicalBiasBERT**, saving the dates and softmax scores. For analysis, we averaged the softmax scores of the data per month to analyze trends over time.
- For **NYT article summaries**, benchmark events are correlated with an increase in center-leaning scores. Our hypothesis is that journalists report factually on significant events as they happen, which tends to result in less biased news.
- **CNN Live at Dawn** is rated as extremely left-leaning after September 2002, around when the U.S. publicly began planning to war on Iraq. LIME shows that the most heavily weighted word was "CNN". Replacing "CNN" with "News Network" causes trends to become less skewed, showing the model's sensitivity to words that are not relevant to the context.

Limitations

- Data quantity: CNN and FOX have transcripts, but the New York Times, Wall Street Journal and Associated Press prevent mass scraping of their content.
- LIME is computationally expensive and can take a while to run, limiting analysis of transcripts en masse.
- We focused on American news in written in English. This analysis could be done on news networks in different countries and in different languages covering American news for further insights into public policy.

Plan for Report

- Expand upon literature review and motivations in detail
- Discuss other models and why we chose PoliticalBiasBERT.
- Provide analysis of authors Don Lemon and Tucker Carlson and their political leaning.
- Conduct a more in-depth analysis of the NYT article summary trends, as well as the difference between using "CNN" and "News Network" in the CNN transcripts.

[1] Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We Can Detect Your Bias: Predicting the Political Ideology of News Articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991, Online. Association for Computational Linguistics.

[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "'why should I trust you?": Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, pp. 1135–1144, 2016.