**Detecting Media Bias in the Coverage of U.S. Presidential Elections Using a Hybrid BERT-BiLSTM Model**

**Abstract**

1. **Modeling the Experiments**  
   **2.1 Datasets**  
    **2.1.1 Allsides.com**  
    **2.1.2 Lean-Left and Lean-Right Newspapers Dataset**  
    **2.1.3 Left and Right Newspapers Dataset**  
    **2.1.4 Extended NewB**  
   **2.2 Models**  
    **2.2.1 Parameters for BERT**  
    **2.2.2 BERT Plus a Classification Output Layer**  
    **2.2.3 BERT Together with a LSTM**  
    **2.2.4 BERT Together with a CNN and a BiLSTM**  
   **2.3 Experiments**  
   **2.4 Evaluation Metrics**  
   **2.5 Comparisons with Other Approaches**  
    **2.5.1 The Custom Datasets**  
    **2.5.2 SemEval-2019 Task 4**  
   **2.6 Mathematical Model**
2. **Case Study**  
   **3.1 Creation of the Left and Right Newspapers Dataset**  
   **3.2 Implementation of the Models**  
    **3.2.1 Technologies Used**   
    **3.2.2 Implementation Details**  
   **3.3 Results**  
    **3.3.1 Metrics**  
    **3.3.2 Examples**  
    **3.3.3 Comparison with Dbias**
3. **Related Work**  
   **4.1 Survey of Other Approaches**  
   **4.2 Differences and Similarities with my Approach**  
   **4.3 Areas Expected to Improve Upon**  
   **4.4 Benchmark Datasets to Use**
4. **Source Code Repository**

**Abstract**

Recognizing political bias in the media one consumes is essential for being an informed citizen. One common situation where media outlets might lean towards a certain end of the political spectrum is during the elections. This paper aims to contribute to understanding political bias in written media, particularly in the context of pivotal events. We will present both a BERT+LSTM and a BERT+CNN+BiLSTM model for detecting such biases during U.S. presidential elections. These models classify news articles in 3 categories: favoring the Left, favoring the Right, and neutral. Their performance will be evaluated on 3 custom datasets consisting of articles extracted from U.S. newspapers, as well as on the “SemEval-2019 Task 4” Dataset (Kiesel et al., 2019). The performance on the latter dataset will determine how well our models generalize to other types of media biases.

**1 Modeling the Experiments**

**1.1 Datasets**  
Lazaridou et al., 2016 builds a dataset consisting of political articles about the world and the UK using Guardian and The Telegraph articles. Similarily, we will extract articles from American newspapers to build two datasets. Additionally, we will build and use an extended version of the NewB (introduced in Wei, 2020) dataset.

**4.1.1 Allsides.com**

We will use the chart on Allsides.com to determine the political orientation of a newspaper. It organizes newspapers into 5 categories: left, lean left, center, lean right, and right.

**4.1.2 Lean-left and lean-right newspapers dataset**

The first dataset will use articles from lean-left and lean-right newspapers. We will extract sentences about all aspects of the presidential race. We will consider the sentences to be slightly biased towards the newspaper’s orientation. We will use American newspapers, both Democrat- (e.g. New York Times) and Republican-leaning (e.g. Wall Street Journal). By training on this dataset, we increase the models’ ability to recognize subtler biases. During the test stage, we expect weaker results on this dataset, compared to the left and right dataset described below, due to the aforementioned subtleties.

**4.1.3 Left and right newspapers dataset**

We will build a similar dataset using articles from left and right newspapers. We can say with a higher confidence that the sentences extracted from these have a certain bias.

**4.1.4 Extended NewB**

NewB (Wei, 2020): collection of over 200,000 sentences about Republican (right-wing) candidate Donald Trump extracted from 9 mainstream newspapers. The sentences are labeled by source. We expect the sentences to lean in the same direction on the political spectrum as the newspapers they originated from. We will extend it by adding sentences about Clinton, Biden, or other Left candidate extracted from the same 9 newspapers.

**1.2 Models**

**1.2.1 Parameters for BERT**  
For all three models, we will subsequentially try the parameters that the authors of BERT recommend in Devlin et al., 2019:

1. *Batch size*: 16 and 32
2. *Learning rate*: 5e-5, 3e-5, 2e-5
3. *Number of epochs*: 2, 3, and 4

**1.2.2 BERT plus a classification output layer**

This will be the simplest model we will try. Devlin et al., 2019 claims that it is enough to add an additional output layer to BERT to create well-performing models. We will add a dense output layer that predicts the class choosing from 3 options: “left,” “neutral,” and “right.”

**1.2.3 BERT together with a LSTM**

BERT will take the raw text as input and generate contextual embeddings. BERT’s embeddings will be fed into a LSTM. LSTMs capture context over longer sequences, as they incorporate elements called “gates” specifically for this purpose (Siami-Namini et al., 2019). We will feed their output into a dense classification layer that will choose from the three output classes.

**1.2.4 BERT, CNN, and BiLSTM**

BiLSTM is an upgrade to LSTM. As opposed to LSTMs, BiLSTMs look both in the future and in the past, while LSTMs only exploit historical context (Liu et al., 2019). By combining BERT and a BiLSTM, we hope to obtain a model whose understanding of the input is better than those of the individual models.

The raw input text will go through BERT. Then will follow several CNN layers. Using this approach, Kaur et al., 2023 obtained a classification model that “performs better than the state-of-the-art baseline approach.” The output of the CNN layers will be taken as input by the BiLSTM. In the end, a dense classification layer will choose the output class.

To the author's knowledge, as of November 1, 2024, no studies have been done on building a neural network-based model specifically designed to detect bias in written-text media coverage of the 2024 U.S. presidential elections.

There have been a considerable number of studies on detecting either political bias in written media or political ideology of certain people, without focusing on the 2024 elections. Their approaches vary:

a) *Bayesian classifier* (Fang et al., 2015) or *Bayesian Hidden Markov Models* (Sim et al., 2013)

d) *statistical models such as support vector machines* (Lin et al., 2006)

c) *recursive neural networks* (Iyyer et al., 2014)

b) *fine-tuning Transformers models, such as BERT, on custom datasets* (Raza et al., 2022)

c) *LSTMs* (Baly et al., 2020)

**1.3 Experiments**

We will train each model on all three datasets. Based on the metrics obtained, we will adjust their hyperparameters (e.g.: number of epochs, learning rate) to improve each model’s performance. After optimizing the models, we will compare the best-performing one with state-of-the-art political bias detection models.

**1.4 Evaluation Metrics**

We will use the following metrics to compare the performance of the models: *accuracy, precision*, *recall*, and *macro F1* (suggested by Baly et al., 2020).

**1.5 Comparisons with Other Approaches**

As we mentioned above, we will compare our best-performing model with state-of-the-art political bias detection models. These state-of-the-art models include: the Transformers-based Dbias model (Raza et al., 2022) and the LSTM-based model proposed in Baly et al., 2020.

We will compare how they perform both on our custom datasets, and on datasets that are widely used, such as the “SemEval-2019 Task 4” Dataset (Kiesel et al., 2019).

Note that the performance on some datasets will determine how well our models generalize to political bias in general, as the most popular datasets for political bias do not specifically focus on U.S. elections.

In some cases, we might need to adjust our model to compare it with others. For example, the Dbias model proposed in Raza et al., 2022 has a binary output (biased or not). We can transform our 3-class output into a binary one by considering “left” and “right” to be biased, and “neutral” to be unbiased.

**1.6 Mathematical Model**

TODO

**2 Case Study**

**2.1 Creation of the Left and Right Newspapers Dataset**

We have so far only created and used the left and right dataset. We have used neither the lean-left and lean-right dataset, nor the Extended NewB dataset yet.

The end goal of the data gathering was to obtain a CSV file with three columns: *Text, Year* (of the elections)*,* and *Leaning* (“left,” “right,” or “center”).

We collected data from the 2012, the 2016, and the 2020 elections. We skipped the 2024 one because one of the parties changed candidates late in the race.

For each of the 3 leanings, we used *allsides.com* to choose a representative newspaper. We chose Jacobin (left), AP News (central) and Newsmax (right) due to them being free to access and having an extensive archive.

We chose between 2 and 6 search terms for each election. For example, for 2012 (Obama vs Romney) we chose “barack obama 2012 elections,” “barack obama rally 2012,” “mitt romney 2012 elections,” and “mitt romney presidential rally.” Then, we searched articles containing these phrases using the web scrapping Python library BeautifulSoup (TODO citation). To get articles from Newsmax, we used Google, as Newsmax’s search engine was not of satisfactory quality. Once we got an article, we searched for the names of the candidates inside the article and extracted fragments of 2-4 sentences around that occurrence of their names. Finally, the dataset will include an entry which has that 2-to-4-sentence-long fragment as *Text*, and the leaning of the newspaper (as per *allsides.com*) as *Leaning*.

In the end, we obtained a dataset with 948 entries: 316 for each leaning.

**2.2 Implementation of the Models**

**2.2.1 Technologies used**

To build, train and evaluate the three models, we used Google Colab notebooks with T4 GPUs. We used machine learning Python libraries such as Torch (TODO cite), SkLearn (TODO cite), and data manipulation libraries such as Pandas and Numpy. To load the pretrained BERT models we used the Transformers library made available by Hugging Face (TODO Cite)

**2.2.2 Implementation Details**  
We are first loading the data from the CSV file and splitting it into train and evaluation datasets. Then, we set hyperparameters such as learning rate, the number of epochs, and the batch size. For each of our 3 architectures, we tried many combinations of these hyperparameters. We were guided by the suggestions offered by Devlin et al. that we presented above. Additionally, in the more complex two architectures we used the Adam optimizer, the cross-entropy loss function (recommended for classification tasks – TODO citation), and we trained the model sequentially on chunks of data to avoid memory issues.

For the simple BERT model, we do not have much control over the architecture.

For the BERT+LSTM model, we had a BERT layer similar to the simple model followed by a LSTM layer with hidden size 128. Finally, we had a fully connected layer with 128 input and 3 output neurons.

For the BERT+CNN+BiLSTM model, we added three CNN layers between the BERT encoder and the LSTM layer. We also made the LSTM bidirectional. We set the hidden size of the BiLSTM layer to 256, and that of the CNN layers to 128. The CNN layers have different kernel sizes (2, 3, and 5), in order to capture varied textual patterns. To prevent overfitting, we added dropout with a 0.5 rate (TODO cite paper recommending dropout). Finally, a softmax layer is applied to make the 3 output values sum up to 1, as if they were probabilities. When predicting leanings, the model chooses the leaning with the highest probability.

**2.3 Results**

**3 Methodology**

**3.1 Parameters for BERT**  
For all three models, we will subsequentially try the parameters that the authors of BERT recommend in Devlin et al., 2019:

1. *Batch size*: 16 and 32
2. *Learning rate*: 5e-5, 3e-5, 2e-5
3. *Number of epochs*: 2, 3, and 4

**3.2 BERT plus a classification output layer**

This will be the simplest model we will try. Devlin et al., 2019 claims that it is enough to add an additional output layer to BERT to create well-performing models. We will add a dense output layer that predicts the class choosing from 3 options: “left” (favoring Harris), “neutral,” and “right” (favoring Trump).

**3.3 BERT together with a LSTM**

BERT will take the raw text as input and generate contextual embeddings. BERT’s embeddings will be fed into a LSTM. LSTMs capture context over longer sequences, as they incorporate elements called “gates” specifically for this purpose (Siami-Namini et al., 2019). We will feed their output into a dense classification layer that will choose from the three output classes.

**3.4 BERT, CNN, and BiLSTM**

BiLSTM is an upgrade to LSTM. As opposed to LSTMs, BiLSTMs look both in the future and in the past, while LSTMs only exploit historical context (Liu et al., 2019). By combining BERT and a BiLSTM, we hope to obtain a model whose understanding of the input is better than those of the individual models.

The raw input text will go through BERT. Then will follow several CNN layers. Using this approach, Kaur et al., 2023 obtained a classification model that “performs better than the state-of-the-art baseline approach.” The output of the CNN layers will be taken as input by the BiLSTM. In the end, a dense classification layer will choose the output class.

**4 Experiments**

**4.1 Datasets**  
Lazaridou et al., 2016 builds a dataset consisting of political articles about the world and the UK using Guardian and The Telegraph articles. Similarily, we will extract articles from American newspapers to build two datasets. Additionally, we will build and use an extended version of the NewB (introduced in Wei, 2020) dataset.

**4.1.1 Allsides.com**

We will use the chart on Allsides.com to determine the political orientation of a newspaper. It organizes newspapers into 5 categories: left, lean left, center, lean right, and right.

**4.1.2 Lean-left and lean-right newspapers dataset**

The first dataset will use articles from lean-left and lean-right newspapers. We will extract sentences about all aspects of the presidential race. We will consider the sentences to be slightly biased towards the newspaper’s orientation. We will use American newspapers, both Democrat- (e.g. New York Times) and Republican-leaning (e.g. Wall Street Journal). By training on this dataset, we increase the models’ ability to recognize subtler biases. During the test stage, we expect weaker results on this dataset, compared to the left and right dataset described below, due to the aforementioned subtleties.

**4.1.3 Left and right newspapers dataset**

We will build a similar dataset using articles from left and right newspapers. We can say with a higher confidence that the sentences extracted from these have a certain bias.

**4.1.4 Extended NewB**

NewB (Wei, 2020): collection of over 200,000 sentences about Republican (right-wing) candidate Donald Trump extracted from 9 mainstream newspapers. The sentences are labeled by source. We expect the sentences to lean in the same direction on the political spectrum as the newspapers they originated from. We will extend it by adding sentences about Harris extracted from the same 9 newspapers.

**4.2 Evaluation metrics**

We will use the following metrics to compare the performance of the models: *accuracy, precision*, *recall*, and *macro F1* (suggested by Baly et al., 2020).

**4.3 Results**

We will present the above metrics for all models, both on the custom datasets we used for training and on the “SemEval-2019 Task 4: Hyperpartisan News Detection” dataset (Kiesel et al., 2019).

The latter dataset will offer a glimpse into how well our model can generalize to other types of media political bias detection, as it is not related to the 2024 elections. It expects a binary output signifying whether the input text is radical or not, regardless of whether to the left or to the right. To test our 3-class classifiers on this binary classification dataset, we will add a layer that maps classes “left” and “right” to radical, and class “neutral” to non-radical. If our models are successful on this dataset, our process could be viewed as solving a transfer learning task.

**4.4 Comparisons**

We will use the metrics described above to compare how the models perform on the 3 custom datasets and on the SemEval dataset.

**5 Conclusion**

In the “Conclusion” section of the paper, we will decide whether the hypothesis has been proved or not, based on an overview of the metrics described previously.

**References**

1. Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein and Martin Potthast. 2019. SemEval-2019 Task 4: Hyperpartisan News Detection. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 829–839.
2. Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita. 2019. A Survey of the Usages of Deep Learning for Natural Language Processing. In IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. XX, NO. X, JULY 2019.
3. Alghamdi Jawaher, Lin Yuqing and Luo Suhuai. 2022. Modeling Fake News Detection Using BERT-CNN-BiLSTM Architecture. 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR).
4. Aixiang He and Mideth Abisado. 2024. Text Sentiment Analysis of Douban Film Short Comments Based on BERT-CNN-BiLSTM-Att Model. In IEEE Access 12 (2024), pages 45229-45237.
5. Bhart Gupta, P. Prakasam and T. Velmurugan. 2022. Integrated BERT embeddings, BiLSTM-BiGRU and 1-D CNN model for binary sentiment classification analysis of movie reviews. In Multimedia Tools and Applications, volume 81, issue 23, pages 33067 – 33086.
6. Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186.
7. Anjie Fang, Iadh Ounis, Philip Habel, Craig Macdonald and Nut Limsopatham. 2015. Topic-centric Classification of Twitter User’s Political Orientation. In SIGIR '15: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 791 – 794.
8. Yanchuan Sim, Brice D. L. Acree, Justin H. Gross and Noah A. Smith. 2013. Measuring Ideological Proportions in Political Speeches. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 91–101.
9. Wei-Hao Lin, Theresa Wilson, Janyce Wiebe and Alexander Hauptmann. 2006. Which Side are You on? Identifying Perspectives at the Document and Sentence Levels. In Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X), pages 109–116.
10. Mohit Iyyer, Peter Enns, Jordan Boyd-Graber and Philip Resnik. 2014. Political Ideology Detection Using Recursive Neural Networks. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1113–1122.
11. S. Raza, D.J. Reji and C. Ding. 2022. Dbias: detecting biases and ensuring fairness in news articles. In Int J Data Sci Anal 17, pages 39–59.
12. 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.
13. S. Siami-Namini, N. Tavakoli and A. S. Namin. 2019. The Performance of LSTM and BiLSTM in Forecasting Time Series. In 2019 IEEE International Conference on Big Data (Big Data), pages 3285-3292.
14. Gang Liu and Jiabao Guo. 2019. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. In Neurocomputing, volume 337, pages 325-338.
15. Kamaljit Kaur and Parminder Kaur. 2023. BERT-CNN: Improving BERT for Requirements Classification using CNN. In Procedia Computer Science, volume 218, pages 2604-2611.
16. Konstantina Lazaridou and Ralf Krestel. 2016. Identifying Political Bias in News Articles. In Bull. IEEE Tech. Comm. Digit. Libr., volume 12, n. pag.
17. Jerry Wei. 2020. NewB: 200,000+ Sentences for Political Bias Detection. In ArXiv abs/2006.03051, n. pag.