Fake News Detection using a Deep Learning Approach

Abstract:

In today's digital era, the proliferation of fake news has emerged as a pressing concern, especially during pivotal events like elections and pandemics. The rapid dissemination of false information across social media platforms has magnified the challenge of differentiating between reliable information and misinformation. This project aims to leverage the capabilities of artificial intelligence (AI) algorithms to address the issue of fake news propagation by enhancing its detection capabilities.

1. Introduction:

The pervasive nature of fake news has serious consequences, including the distortion of public opinion, manipulation of political discourse, and erosion of trust in media. The dissemination of fake news is exacerbated during critical events, making it a matter of utmost concern. In this context, AI-based solutions hold the potential to automate the identification and mitigation of fake news, ultimately aiding in the preservation of information integrity.

2. Related Work:

This project builds upon a body of prior research in the field of fake news detection. Various methodologies have been proposed, encompassing user-based, knowledge-based, social network-based, and style-based approaches. Additionally, BERT (Bidirectional Encoder Representations from Transformers), a prominent deep learning model, has shown promise in addressing this challenge.

3. Dataset and Data Preprocessing:

The ISOT Fake News dataset, compiled by the ISOT Research Lab at the University of Victoria in Canada, serves as the foundation for this project. It consists of both fake and truthful articles gathered from diverse sources. Data preprocessing techniques involve text cleaning, sentence cropping, and the selection of news titles as the primary data source. The dataset is partitioned into training, validation, and test sets for experimentation.

Figure1(a)

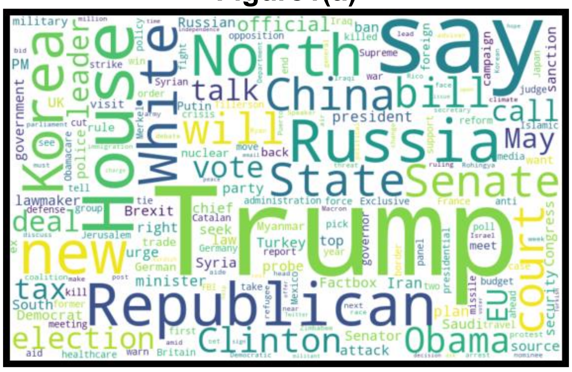
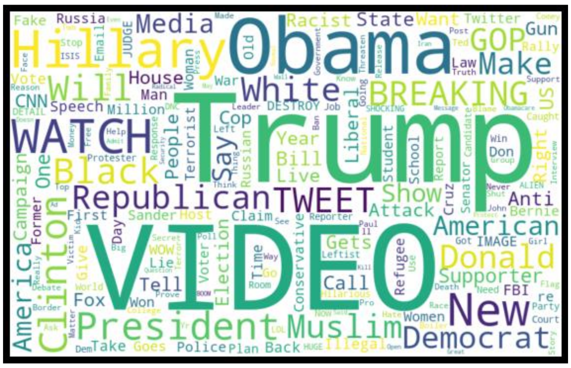


Figure 1(b)



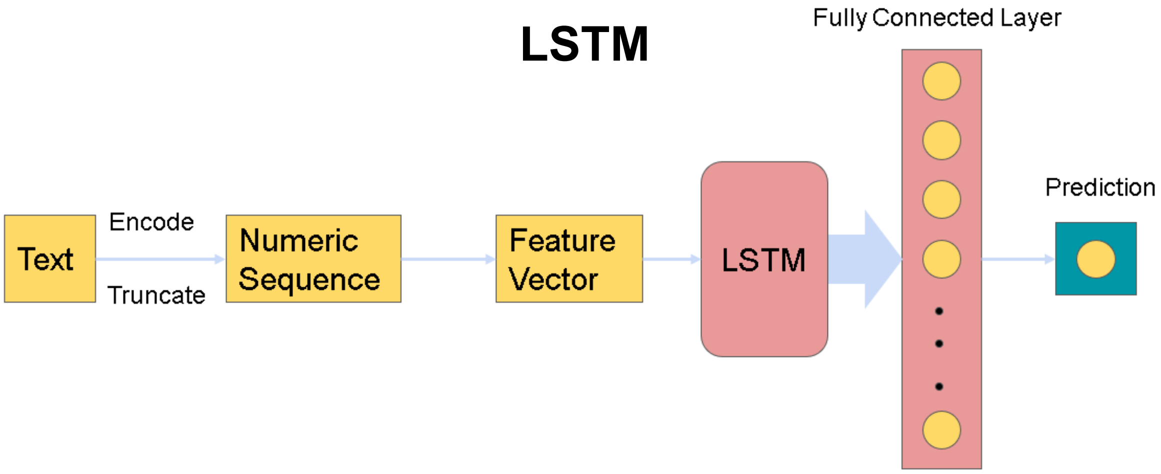
We can see that, in the real news word cloud, ‘Trump’, ‘say’, ‘Russia’, ‘House’, ‘North’, and Korea’ appeared frequently; while in fake news one, ‘VIDEO’, ‘Trump’, ‘Obama’, ‘WATCH’, and ‘Hillary’ appeared the most frequently. ‘Say’ appears frequently in real news but does not in fake news. ‘VIDEO’, and ‘WATCH’ appear frequently in fake news but do not in real news. From these two word clouds, we can get some important information to differentiate the two classes of data. The original form of the dataset is two CSV files containing fake and real news respectively. We combined the dataset and split it into training, validation, and test sets with shuffling at the ratio of 64%:16%:20%. The original combined dataset contains 44,898 pieces of data, and Table 1. shows the distribution of data in the training, validation, and test sets

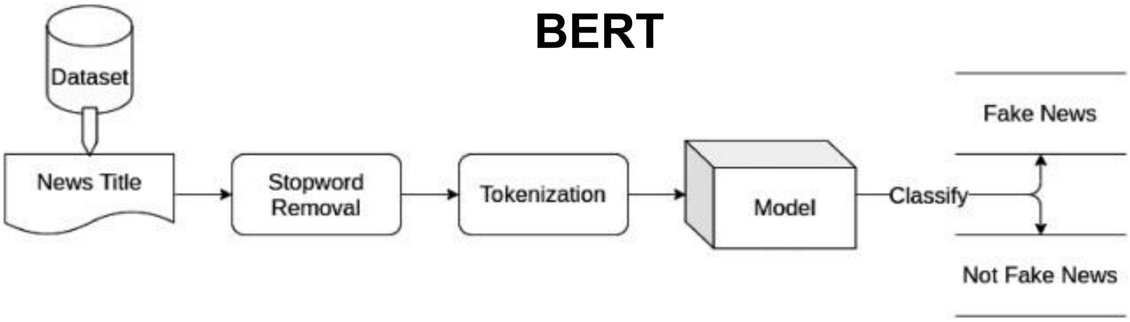
4. **Word Embedding:**

The preparation of textual data for machine learning models hinges on the use of word embedding techniques. Different models, including LSTM, BiLSTM, CNN-BiLSTM, and BERT, necessitate distinct embedding methodologies. Tokenization, sequence padding, and embeddings initialization are integral components of this process.

5. Models:

This project employs four distinct models for fake news detection: LSTM, BiLSTM, CNN-BiLSTM, and BERT. Each model offers unique advantages in the context of natural language processing (NLP) tasks.





6. Model Fine-Tuning:

Fine-tuning of the BERT-based model, specifically bert For Sequence Classification, is a crucial step in adapting the pre-trained model to the task of fake news detection. Components of the model, such as embeddings, transformer encoder, and linear classifier, play pivotal roles in this process.

7. Experiments and Results:

Experimental evaluations encompass both balanced and imbalanced datasets, reflecting real-world scenarios. Performance metrics include accuracy, precision, recall, and F1-score. The results underscore the superiority of transformer-based models, particularly BERT, and reveal that imbalanced data may yield slightly better results.

8. Analysis of Misclassifications:

In-depth analysis of misclassified data provides valuable insights into the models' limitations and areas for improvement. The distribution of misclassified data highlights the challenges faced in distinguishing fake news from real news articles.

9. Conclusions and Future Directions:

The findings underscore the critical role of fake news detection in contemporary society. BERT-based models exhibit superior performance in the context of fake news detection. Future research may explore avenues such as hyperparameter tuning and the integration of additional NLP techniques.

10. Experiment

Our dataset is balanced between real news and fake news. However, in the real world, real news and fake news are not as balanced as what the dataset shows. We assumed that the amount of real news is way larger than the amount of fake news to reflect the real-life situation. To build the imbalanced dataset, we shrunk the original fake dataset into one-tenth of the original size to imitate the real-world situation.

Below table shows the true/fake distribution of data in the original training set, imbalanced training set, original validation set, imbalanced validation set, original test set, and imbalanced test set.

Table 2. Distribution of Imbalanced Data
(Ori: Original, Imb: Imbalanced)

Data	Training Set		Validation Set		Test Set	
	Ori	Imb	Ori	Imb	Ori	Imb
True	13765	13765	3409	3409	4243	4243
Fake	14969	1497	3775	378	4737	474

11.Result Summary:

In the evaluation of our models for the fake news detection project, we employed common metrics widely used in the field of machine learning, particularly for classification problems. These metrics include:

- True Positive (TP): When the model correctly identifies fake news as fake news.
- True Negative (TN): When the model correctly identifies true news as true news.
- False Negative (FN): When the model mistakenly classifies true news as fake news.
- False Positive (FP): When the model mistakenly classifies fake news as true news.

We can define the following metrics based on the value of the above 4 situations.

$$\begin{aligned}
Precision &= \frac{|TP|}{|TP| + |FP|} \\
Recall &= \frac{|TP|}{|TP| + |FN|} \\
F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \\
Accuracy &= \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}
\end{aligned}$$

These metrics offer valuable insights into the performance of our classifier from various angles. Among these metrics, 'Accuracy' is often considered the most representative, as it provides a comprehensive view of classification performance.

The key findings from our evaluation are as follows:

- Training on imbalanced datasets yielded slightly superior results compared to training on balanced datasets for the fake news detection task.
- Transformer-based models outperformed other models significantly.
- Deep learning models with more attributes demonstrated better performance compared to models with fewer attributes.

ON BALANCED DATASET

Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.9697	0.97	0.97	0.97
BERT	0.9874	0.99	0.99	0.99

ON IMBALANCED DATASET

Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.9780	0.98	0.98	0.98
BERT	0.9903	0.99	0.99	0.99

In conclusion, our project's findings suggest that Transformer-based models and deep learning models with an increased number of attributes are the most effective for the task of fake news detection. The use of imbalanced datasets can also be beneficial for improving model performance. Our primary metric for evaluating classification performance was Accuracy.