Fake News Detection
CS3008: Deep Learning

Thribhuvan S (EC21B1093)
Chetan Vasista (EC21B1093)
Amar Rohith (EC21B1093)

April 20, 2025

Abstract

The rapid proliferation of fake news across digital platforms has emerged as a critical societal challenge, enabling the widespread dissemination of misinformation that significantly impacts public discourse, democratic processes, and social cohesion. While existing fact-checking mechanisms remain essential, their manual nature and delayed response times render them inadequate for combating the real-time velocity of modern information ecosystems. This underscores the urgent need for intelligent automated detection systems capable of synthesizing multimodal features—including linguistic patterns, visual content, and contextual relationships—to reliably identify deceptive information. However, developing such systems presents fundamental challenges, including the need to process heterogeneous data sources, detect sophisticated manipulation techniques, mitigate algorithmic biases, and maintain operational efficiency at scale. Addressing these limitations requires novel computational approaches that balance detection accuracy with interpretability while adapting to evolving misinformation tactics. Effective solutions in this domain could empower platforms, policymakers, and users to curb misinformation while preserving freedom of expression in digital spaces.

1 Introduction

Fake news detection is formulated as a binary text classification problem. Leveraging the Transformer architecture enables the model to capture long-range dependencies using self-attention, providing improved parallelization over recurrent approaches.

2 Methodology

2.1 Data Preparation

News articles labeled as `True` or `Fake` are combined by concatenating titles and bodies, then cleaned by lowercasing and removing non-alphanumeric characters. A 60–20–20 split is used for train, validation, and test sets.

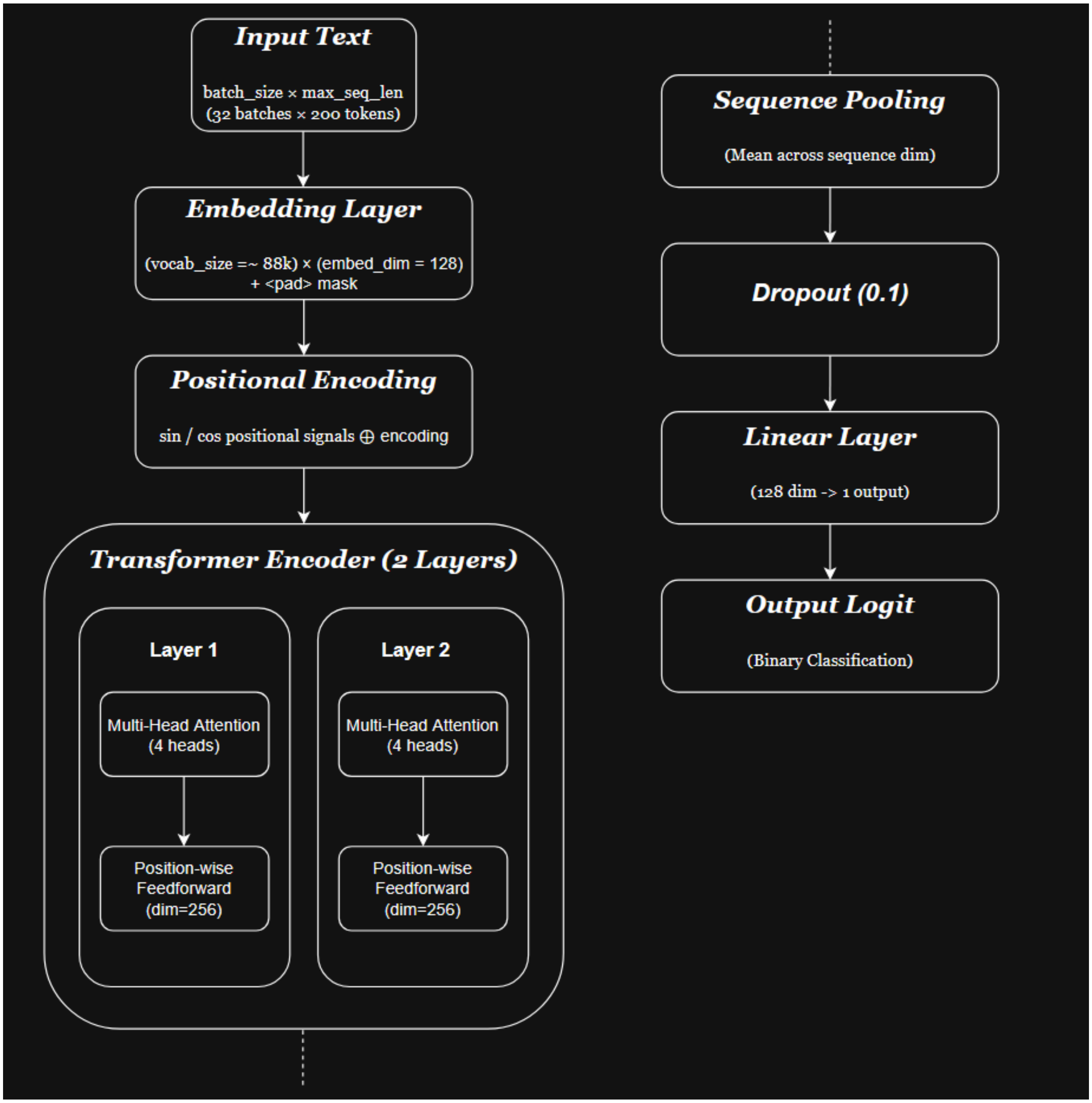
2.2 Vocabulary Construction

A vocabulary is built from the training set, filtering out tokens with frequency below a chosen threshold to balance coverage and model size.

2.3 Model Architecture

- **Embedding Layer:** Maps tokens to 128-dimensional vectors.
- **Positional Encoding:** Adds fixed sinusoidal signals to embeddings.
- **Transformer Encoder:** Two layers of multi-head self-attention (4 heads) and feedforward sublayers (dim=256).
- **Pooling & Classification:** Mean pooling across sequence positions, dropout (0.1), and linear layer reducing to a single logit.

3 Block Diagram



4 Training and Evaluation

The model is trained using BCEWithLogitsLoss and the Adam optimizer (lr=1e–3) for 5 epochs. Evaluation metrics include accuracy, AUC, precision, and equal error rate, computed on the held-out test set.

- Accuracy : 0.9979
- AUC : 0.9998
- Precision : 0.9998
- EER : 0.0024 at threshold 0.0131

5 Results and Visualization

Training and validation loss curves, ROC and precision–recall plots, and confusion matrices are generated to assess model performance and identify potential overfitting or class imbalance issues.

6 Conclusion

The Transformer-based approach demonstrates strong performance on the fake news detection task, with clear benefits from self-attention and positional encoding. Future work may explore multimodal extensions and domain adaptation for broader applicability.