How do Transformers compare to LSTMs with attention in performance and efficiency for text-based sentiment analysis?

Authors: Louis Vidal, Jack Woodhouse
Mentor: Matej Cief

Introduction

This study compares the performance and efficiency of Transformer networks and LSTMs with attention mechanisms on the IMDb reviews dataset, a large collection of movie reviews commonly used for sentiment analysis.

Motivation

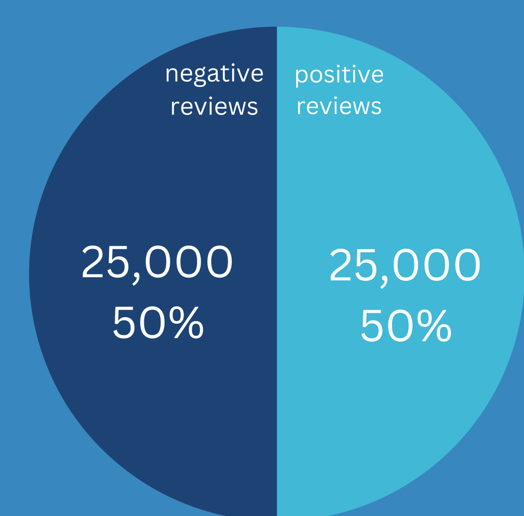
Sentiment analysis plays a critical role in analyzing consumer opinions and public sentiment. The IMDb dataset, consisting of 50,000 balanced positive and negative movie reviews, is ideal for testing binary sentiment classification models. Comparing LSTM and Transformer models helps identify the best approach for various real-world applications, considering both performance and computational constraints.

Objective

- Compare the performance of LSTMs with Attention and Transformers for sentiment analysis.
- Evaluate their efficiency regarding training time, inference speed, and resource usage.

IMDb Dataset:

The IMDb reviews dataset consists of 50,000 movie reviews, with an even split of positive and negative labels. So this is perfect for training and testing binary classification with sentiment analysis.



Introduction
Methodology
Results
Discussion

Abstract

This study evaluates the performance and efficiency of Transformers and LSTMs with attention mechanisms on the IMDb reviews dataset for sentiment analysis. We compare the models based on accuracy, training speed, and resource requirements, providing insights into their strengths and limitations

Overview

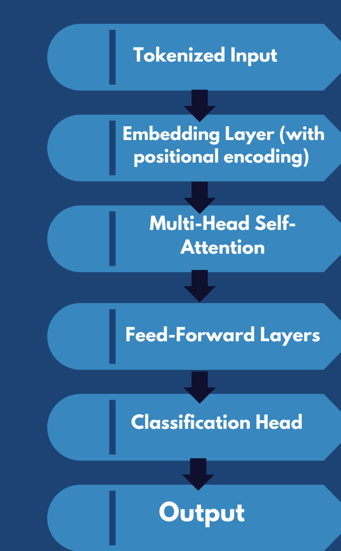
This study employs two distinct deep learning architectures, LSTMs with Attention and Transformers, to perform sentiment analysis on the IMDb reviews dataset. Both models were trained and evaluated on their ability to classify movie reviews as positive or negative. The methodology comprises data preprocessing, model training, and a comparative analysis of performance and efficiency. The model used are trained from scratch as opposed to using pretrained models such as Bert. This insures a fair test when comparing both models.

Transformer:

Transformers, use self-attention to process all words in a sequence simultaneously. This allows the model to capture relationships between words, regardless of their distance in the text. The parallelized processing results in faster training and inference times.
In this example, the multi-head self-attention layer captures different aspects of sentiment, such as phrases like 'highly recommend' or 'waste of time,' improving overall sentiment prediction.

Key Features:

- Processes text in parallel, enabling scalability.
- Multi-head self-attention extracts multiple relationships between words.
- Pre-trained on large datasets for enhanced generalization.



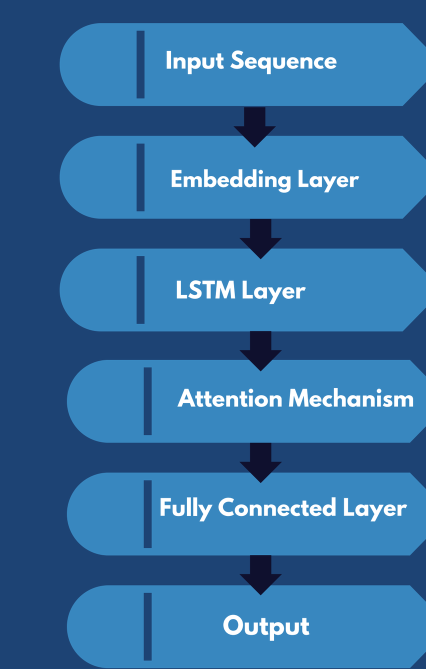
LSTM with Attention:

Long Short-Term Memory (LSTM) networks process sequences step-by-step, capturing dependencies over time. The attention mechanism further improves this by identifying the most important words in a sequence, allowing the model to 'focus' on sentiment-indicative parts of the text.

In this example, the attention mechanism assigns higher weights to emotionally significant words, such as 'amazing' or 'disappointing,' helping the LSTM make more accurate predictions for long reviews.

Key Features:

- Processes text sequentially, maintaining context across words.
- Attention prioritizes crucial parts of the sequence.
- Suitable for smaller datasets or resource-constrained environments.



Efficiency

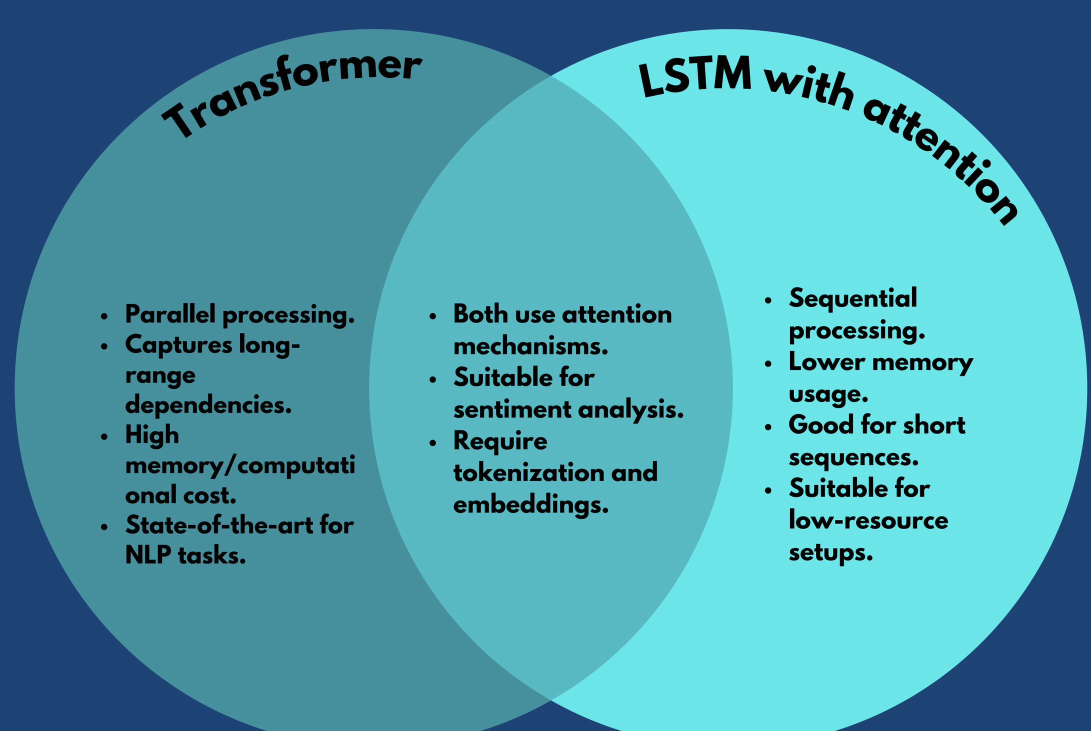
- Training Time: Transformers require more GPU resources and training time due to their larger number of parameters and complex attention layers.
- Inference Speed: Despite their computational intensity, Transformers process input faster than LSTMs during inference, thanks to parallelized architecture.
- Memory Usage: LSTMs with Attention use significantly less memory, making them more suitable for deployment in resource-constrained environments.

Performance

- Overall Performance: Transformers outperform LSTMs with Attention across key metrics.
- Self-Attention Mechanism: Allows Transformers to capture long-range dependencies and subtle text patterns more effectively.
- Accuracy: Higher accuracy, especially for longer reviews where LSTMs may lose context due to sequential processing.
- Precision & Recall: Superior precision and recall, particularly for detecting subtle sentiments in complex sentences.
- F1-Score: Higher F1-scores indicate a balanced and reliable performance in real-world sentiment analysis tasks.

Limitations

- LSTM with Attention: A viable choice for smaller datasets, shorter texts, or environments with limited computational power.
- Transformer: Ideal for large-scale datasets, long-text analysis, or applications requiring high accuracy and scalability.
- However, Transformers' computational demands may limit their applicability in real-time or low-power scenarios.



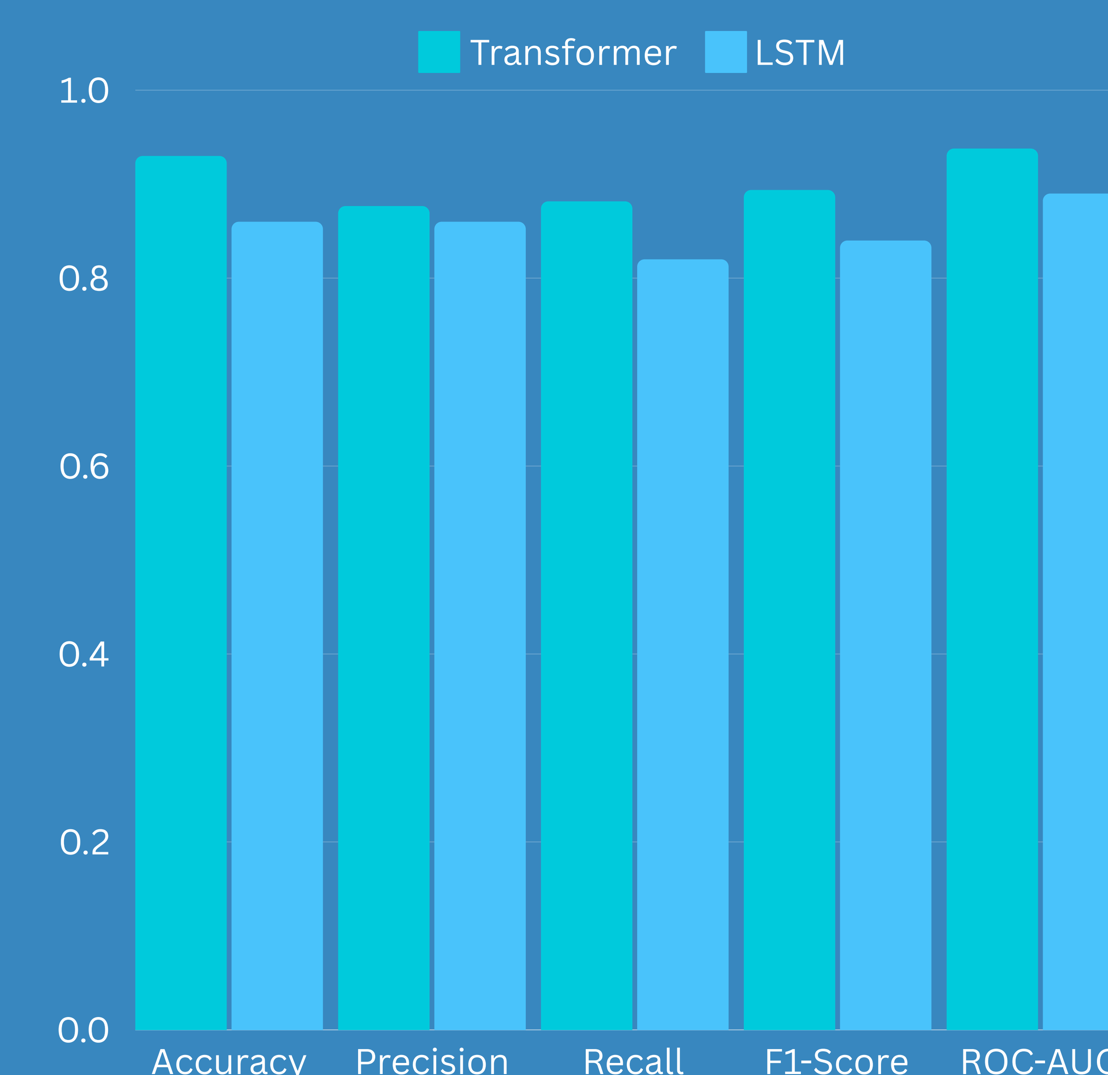
Conclusion

Transformers generally outperform LSTMs with Attention in sentiment analysis tasks, excelling in both performance and efficiency. This makes them ideal for larger datasets and faster processing. However, LSTMs with Attention remain a viable alternative for scenarios with limited computational resources or a need for smaller models. Future research could explore hybrid models that combine the strengths of both architectures or investigate the application of Transformers to smaller, less structured datasets.

Training and Efficiency Metrics

In terms of efficiency, the Transformer required approximately 6 minutes to train across 5 epochs. Its self-attention mechanism enables parallel processing, resulting in faster inference times compared to the sequential nature of LSTMs. However, the Transformer's larger architecture incurs higher computational costs.

Metric	Transformer	LSTM with Attention
Training Time (5 epochs)	~6 minutes	~5 minutes
Memory Usage	2.5 GB (training)	1.5 GB (training)
Inference Speed	12 ms/example	20 ms/example



Performance Metrics Comparison

The performance of both models on the IMDb dataset was evaluated using key metrics:
Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The Transformer outperformed the LSTM with Attention in most metrics, showcasing its superior ability to handle long-range dependencies and complex sentence structures.

LSTM with attention:

Transformer:

Vaswani, A., et al. (2017). Attention Is All You Need. (Reference: https://arxiv.org/abs/1706.03762)

Bahdanau, D., et al. (2015). Neural Machine Translation by Jointly (Reference: https://arxiv.org/abs/1409.0473)