Abstractive Text Summarisation Using Transformers and Bi-LSTMS: A Comparison

Presenters: Aman Chopra, Niharika Sinha, and Siddhartha Singh

Problem Statement:

To generate a summary based on understanding as opposed to extractive summarization, we have trained the system to understand the whole context. This project aims to understand and present the comparison of accuracy and training time for Bi-Directional LSTMs (with encoder-decoder and attention mechanism), and Transformer based approaches. We also aim to present a discussion on the need to shift to transformers and its advantages over Bi-LSTM.

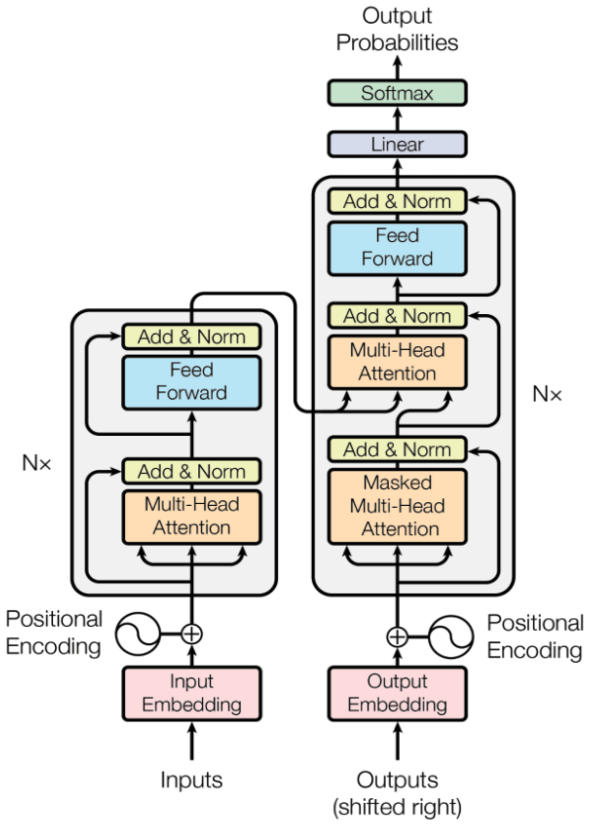

Figure: The architecture for Transformers

NYU

Results

Dataset: InShorts news dataset with 55,104 news - headline pairs was used (~100K unique words)

Pre-processing: Lower casing, abbreviation and punctuations handling, start and end tokens to target sequences, stop-words, vocabulary of words, text-to-sequence, glove word embedding, padding.

Why Transformers?

1. At each step we have direct access to all the other steps (self-attention)
2. Sentences are processed as a whole rather than word by word.
3. Transformers are faster than RNN-based models as all the input is ingested once

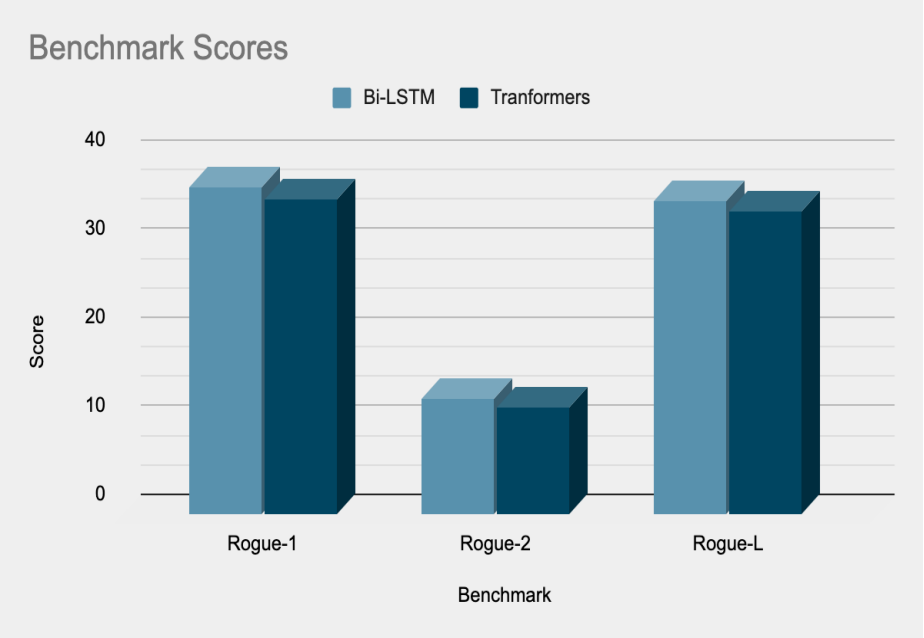

Figure: ROGUE results for Bi-LSTM and Transformers