Neural Text Compression: A Hybrid LSTM + Entropy Coding Framework

TEAM MEMBERS:

Vijayalakshmi G 22MIC0082

Mythili G 22MID0301

Hemasree R 22MID0303

Ramya M 22MID0357

Abstract

Data compression does an important role in reducing storage and improving the speed of transmission in this data driven world though the traditional methods like Huffman coding, Arithmetic coding, Lempel-Ziv-Welch (LZW) and DEFLATE (used in ZIP/Gzip) is being efficient and widely used their performance is so much limited by fixed symbol distribution or assumptions. The advancing in deep learning and machine learning have giving raise to neural network such as RNNs and LSTM with the use of these we can do compression more efficiently, this works by learning the patterns in a sentence and the trying to predict the next word. By doing this we will get higher compression ratio mainly in text domine.This paper extends the existing literature by not only reviewing traditional and neural text compression techniques but also by presenting a hybrid framework that combines the strengths of both approaches. The proposed hybrid model integrates statistical methods such as Huffman and Arithmetic coding with neural architectures like LSTM-based predictors to achieve a better balance between compression ratio, computational cost, and adaptability.Through this integration, the model leverages the speed and stability of traditional algorithms while incorporating the context-awareness and predictive power of neural networks. Such an approach demonstrates that hybrid models can achieve near optimal compression efficiency with significantly reduced resource consumption compared to purely neural methods.In summary, this survey highlights the evolution of text compression—from classical entropy and dictionary-based techniques to advanced deep learning-based systems—and positions hybrid frameworks as a promising future direction for achieving efficient, adaptive, and scalable compression across diverse text domains.