

Advanced NLP (CS7.501)

Monsoon 2022, IIIT Hyderabad
Assignment 2

Abhinav S Menon (2020114001)

Questions (Theory)

ELMo & CoVe

ELMo and CoVe differ in both architecture and pretraining details, although both provide contextual representations.

ELMo uses a six-layer bidirectional LSTM, while CoVe uses a two-layer bidirectional LSTM. Furthermore, ELMo sums up the hidden states of each LSTM in its stack, while CoVe takes the final outputs only.

Moreover, ELMo is pretrained on a bidirectional language modelling task (next-and previous-word prediction), while CoVe is trained on a machine translation task. Thus CoVe is paired with a two-layer decoder LSTM during its pretraining, while ELMo is simply used with a classification head to predict the next word.

Another point of difference lies in the incorporation of *global embeddings* (typically GloVe in both cases) into the contextual embeddings. ELMo simply adds them to the forward and backward embeddings, while CoVe concatenates its contextual representation with GloVe to create the final representation.

Character Convolutional Layer

A character convolutional layer is simply an application of CNN (convolutional neural networks) to character sequences instead of pixels. They give the model information extracted from the subword level, which is not available to it otherwise.

An alternative to this is any form of subword tokenisation, of which many methods have been identified. One popular method is BPE (byte-pair encoding), under which we start with characters as different tokens and merge them until we reach a certain vocabulary size. Morph analysis is another method that can be followed.

Analysis