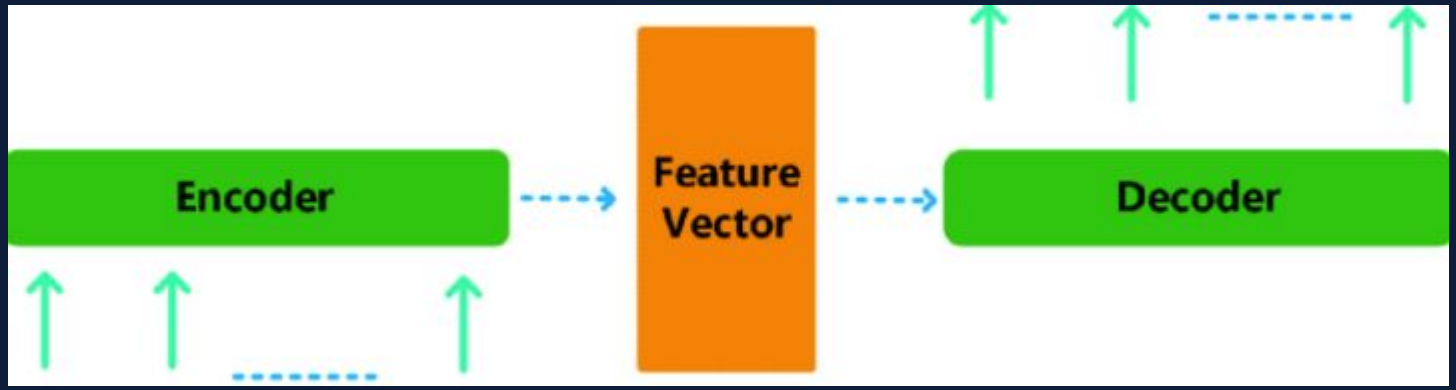
Neural Machine Translation with Attention Mechanism

By Aditi Singhania

What is NMT ?

NMT is basically translating one language into the another with the help of artificial neural network based models .These are known as sequence to sequence models and the architecture is called encoder-decoder.


Encoder
Feature Vector
Decoder

Why do we not convert words directly into another language ? why seq-seq model is used ?

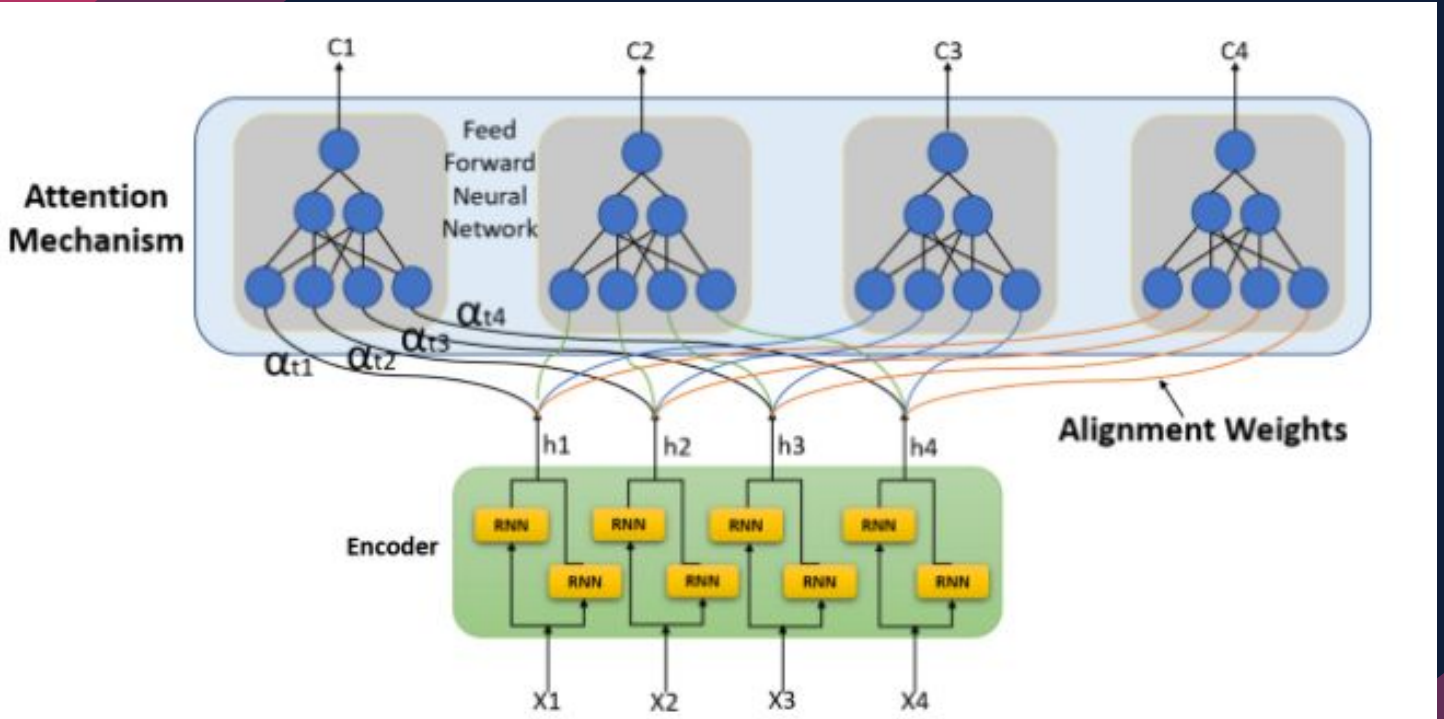
Why attention mechanism ?

The standard seq2seq model is generally unable to accurately process long input sequences, since only the last hidden state of the encoder RNN is used as the context vector for the decoder.

On the other hand, the Attention Mechanism directly addresses this issue as it retains and utilises all the hidden states of the input sequence during the decoding process.

It does this by creating a unique mapping between each time step of the decoder output to all the encoder hidden states.

This means that for each output that the decoder makes, it has access to the entire input sequence and can selectively pick out specific elements from that sequence to produce the output.



Feed Forward Neural network for attention mechanism

Step By Step to Apply Attention Mechanism

1. Producing the Encoder Hidden States - Encoder produces hidden states of each element in the input sequence

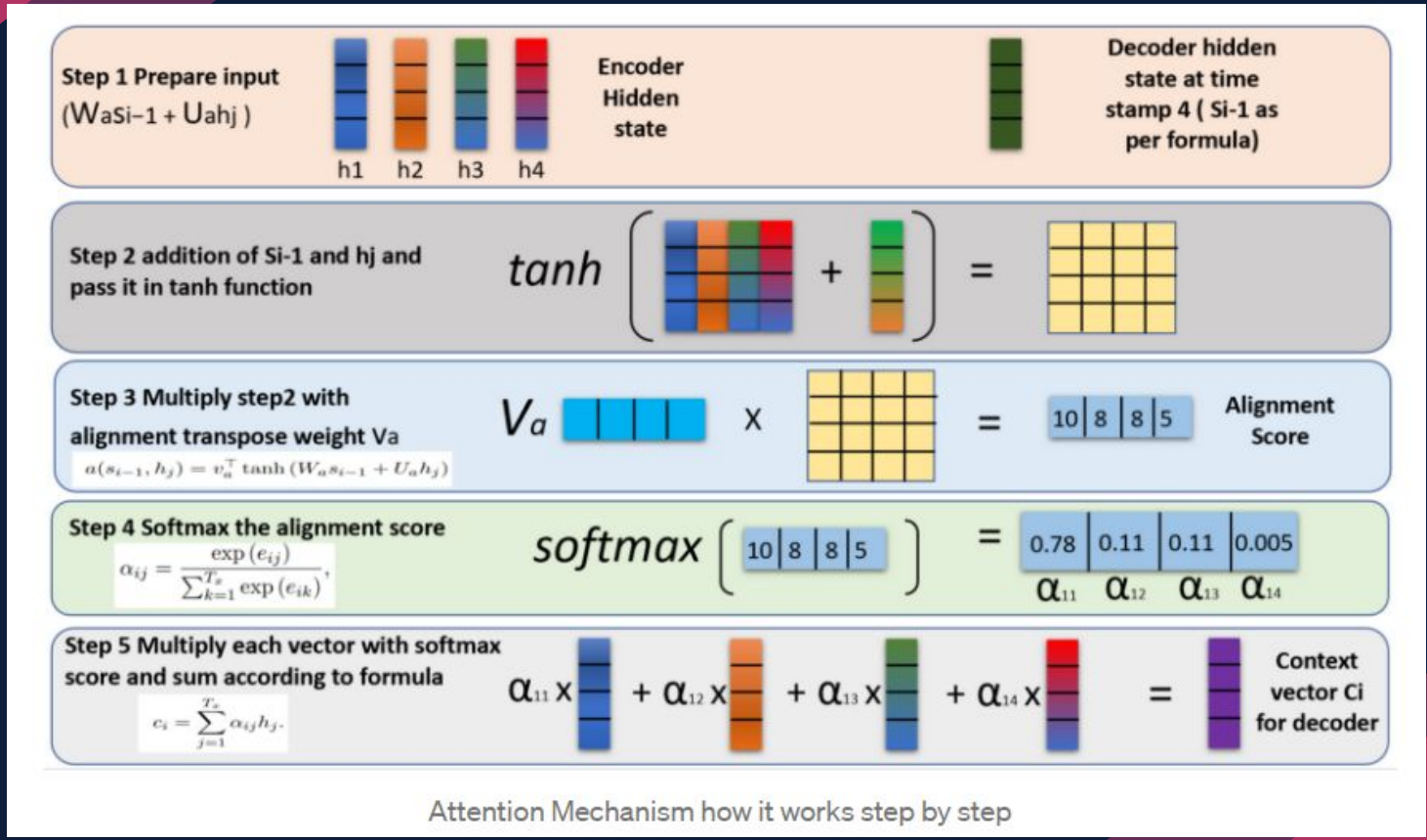
2. Calculating Alignment Scores between the previous decoder hidden state and each of the encoder's hidden states are calculated (Note: The last encoder hidden state can be used as the first hidden state in the decoder)

3. Softmaxing the Alignment Scores - the alignment scores for each encoder hidden state are combined and represented in a single vector and subsequently softmaxed

4. Calculating the Context Vector - the encoder hidden states and their respective alignment scores are multiplied to form the context vector

5. Decoding the Output - the context vector is concatenated with the previous decoder output and fed into the Decoder RNN for that time step along with the previous decoder hidden state to produce a new output

6. The process (steps 2-5) repeats itself for each time step of the decoder until an token is produced or output is past the specified maximum length



Attention Mechanism how it works step by step

Thank You