1. Can you design a learnable positional encoding method?

After the word embeddings of the input sentence, we can add a vector which represent the position of word to the embedding results. Then the position vector can be modified during the process of training.

1. What can be challenges to Transformers if input sequences are very long? Why?

Using a fixed-length vector to represent the input sequence to decode an entirely new sentence is difficult. Because the context vector cannot store all the information if the input sequence is large. Furthermore, it is challenging to differentiate sentences with similar words but with different meanings.