T5 transformer for text summarization

Transfer Learning:

Transfer learning is a method in Machine Learning wherein a model developed for one problem is used as the beginning point and applied into a related problem. It is extremely popular in the deep learning domain given the large number of resources required in the successful training of DL models.

Transformers:

Transformers are an architecture introduced in the paper ‘‘Attention is all you need” by Vaswani et. al. It utilises the concepts of positional encoding, attention and self-attention to transform one sequence into another using an encoder-decoder mechanism. Using the attention mechanism, the model can focus on specific parts of the input while predicting the output, and in the context of Natural language processing, this can help the model decide exactly which parts of the sentence to give more attention to, while predicting the next word.

Since the concept of position or relative position is crucial in the semantics of languages, positional encoding is utilised, where a d-dimensional vector that contains the information about a specific position in a sentence is used to equip each word with information about its position. This is crucial in transformers since each word in a sentence enters the encoder/decoder stack simultaneously (unlike RNNs, where it is sequential processing).

T5 Transformers:

T5 transformers were first suggested by Colin Raffel et. al in Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.

It uses transfer learning for Natural Language Processing (NLP) by introducing a framework that converts text-based NLP problems into a text-to-text format.

T5 stands for ‘Text-to-Text Transfer Transformer’. It is an encoder-decoder model pre-trained on a mixture of multiple tasks (both supervised and unsupervised). Every task, be it translation, question-answering or classification is cast as a text-input to text-output or target- text generator.

T5 is trained using ‘teacher enforcing’, which basically means that as we need an input sequence and a corresponding text output sequence for training.

Summarization:

In the submitted model, we have picked up a paragraph from a news article covering the sinking of the missile cruiser Moskva.

The input paragraph was:

‘‘Russia on Thursday said the flagship of its Black Sea fleet was seriously damaged and its crew evac uated following a fire that caused an explosion, as a Ukrainian official said the vessel had been hit by missiles. The incident on the Moskva missile cruiser occurred after ammunition on board blew up, In terfax news agency quoted the Russian defence ministry as saying. Maksym Marchenko, governor of the region around the Black Sea port of Odesa, said in an online post that the 12,500 tonne ship was hit by two missiles, without providing evidence.’’

The output summary:

the missile cruiser was hit by two missiles, an official says. ammunition on board blew up, the russian defence ministry says. the 12,500 tonne ship was hit by two missiles, an official says. the ship's crew were evacuated after a fire caused an explosion, the ministry says. a u.s. official says the ship was hit by missiles, without providing any evidence.

Jatin

Bansal

2020A4PS02

41P