# T5 Transformer

1) Introduction

T5 stands for Text-To-Text Transfer Transformer, and it is created by GOOGLE. Where GOOGLE used a C4-dataset which is of total 7 TB Storage. It uses a transfer learning mechanism which makes it one of the best models in NLG (natural language generation). let us take a dive inside it and learn how it works, its architecture, working application and limitations

2) Important terms

Before taking a deep dive inside how the T5 transformer works let us see a few terms which will make our understanding clear about things which are coming next in order:
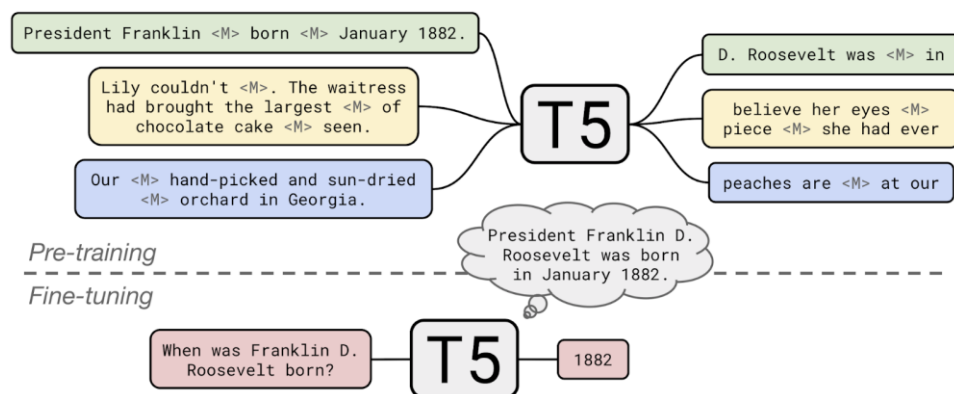
a) Pre-Training: When the model is trained on a large generic corpus, it is called 'pre-training'.
b) Fine- Tunning: When it is adapted to a particular task or dataset it is called as 'fine-tuning'.
c) Transfer Learning: Transfer learning is the process of creating new AI models by fine-tuning to previously trained (pre-training) neural networks.

3) What is T5?

a) T5 is a transformer based which deal with the encoding and decoding both to do text transformation Unlike its other competitors like GPT-3 and BERT it uses Both the encoding and decoding in process of text transformation.
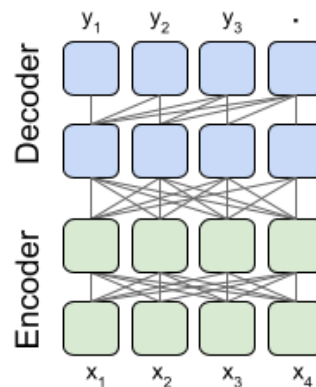
b) T5 used C4 To satisfy these requirements, we developed the Colossal Clean Crawled Corpus (C4), a cleaned version of Common Crawl that is two orders of magnitude larger than Wikipedia.

Dataset	# documents	# tokens	size
C4.EN.NOCLEAN	1.1 billion	1.4 trillion	2.3 TB
C4.EN.NOBLOCKLIST	395 million	198 billion	380 GB
C4.EN	365 million	156 billion	305 GB

c) C4 is used for cleaning process involved deduplication, discarding incomplete sentences, and removing offensive or noisy content.

d) T5 is Pretrained database which is of size 7TB which is too vast as GPT-3 trained on database of size only 40GB.

e) Also, T5 came with multiple number of variants which vary with its use case and its size.

- t5-small
- t5-base
- t5-large

- t5-3b
- t5-11b.

f) Google also published some follow-up works on T5

- T5v1.1is an improved version of T5 with some architectural tweaks and is pre-trained on C4 only without mixing in the supervised tasks.
- MT5 is a multilingual T5 model. It is pre-trained on the mC4 corpus, which includes 101 languages.
- byT5 is a T5 model pre-trained on byte sequences rather than Sentence Piece sub-word token sequences.

4) Architecture and Working

- T5 supports both transformer functionalities encoding and decoding and it will use them for two main processes called pretraining and fine tunning
- The first model structure we consider is an encoder-decoder Transformer, which consists of two-layer stacks.
- The encoder, which is fed an input sequence, and the decoder, which produces a new output sequence.
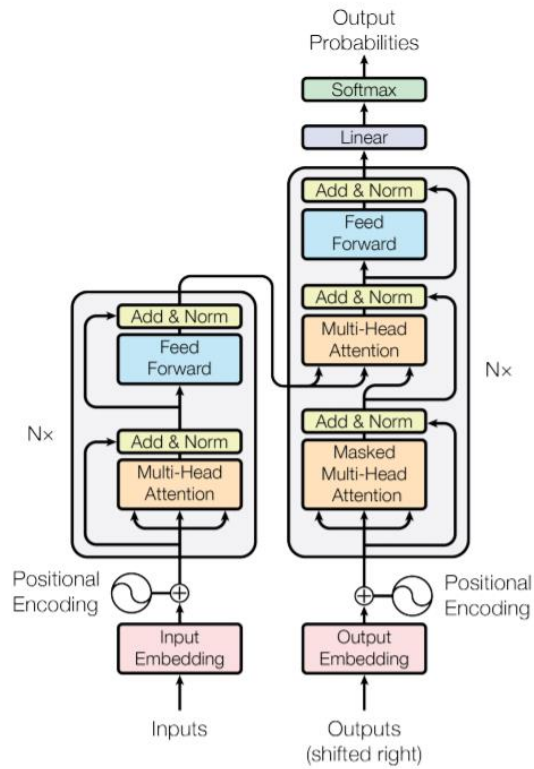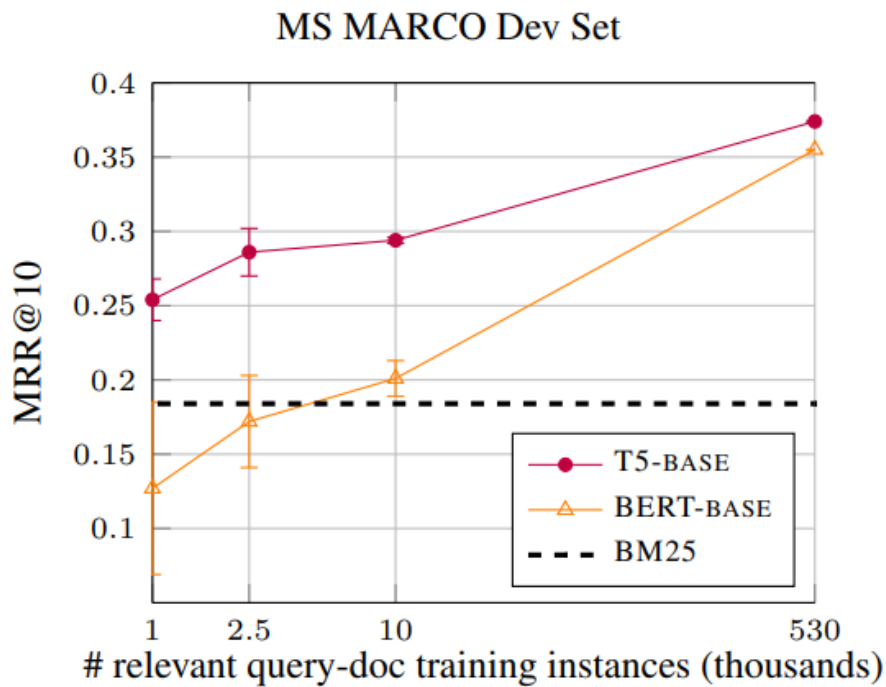
Fig: Transformer architecture. Left part is the encoder, right part is the decoder.

5) Performance

- Comparision between T5 BERT in context of Relevant query-doc Training

MS MARCO Dev Set

6) Applications

- Chat bot creation
- Humans like text creation
- Transfer learning has a great scope in future
- Text summarization

7) References

- htpps://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/1910.10683.pdf
- https://www.toptal.com/deep-learning/exploring-pre-trained-models
- https://paperswithcode.com/method/t5
  https://huggingface.co/docs/transformers/model_doc/t5