- TLDR: Text summarization from basic to advanced approaches

- Text summarization is the task of producing a concise and fluent summary of texts or documents while preserving key information. The goal of this project is to implement different abstractive summarization techniques on the same data set and compare them based on the same evaluation metrics.

-We use CNN and Daily Mail data set, which consists of online news articles paired with summaries and this data is shown in the table. For pre-processing step, we remove punctuation, number, and noisy words. To save computation time, we use only the first two sentence of stories together with 2 summaries.

- For our approaches,

First, our baseline model is a vanilla LSTM sequence-to-sequence model which is an encoder-decoder architecture. This model provides us a baseline of the training time and accuracy

For our second model, we finetune the BART-base model which is a pre-trained transformer model trained on XSum news data set. BART uses a standard sequence-to-sequence architecture with a bidirectional encoder similar to BERT and a left-to-right decoder similar to GPT.

For BART-large, we directly use the model without finetuning.

For T5, it is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks and is trained on C4 data set. The goal of this model is to reframe all NLP tasks into a unified text-to-text-format where the input and output are always text strings.

-To evaluate the generated summaries, we use F1 from the ROUGE metric. ROUGE-1 and ROUGE-2 measures unigram and bigram similarity of reference summary and model-generated one and ROUGE-L measures the longest common subsequence between the two summaries.

From the result table, we can see that LSTM model has a relatively disappointing performance compared with other models. The Transformer-based models like BART-large model and all three T5 models almost doubles the LSTM model. They have very similar performance to each other in all metrics regardless of size of models and underlying data sets.

On the other hand, our finetuned Transformer-based BART-base model outperforms LSTM.

Overall, the accuracy of the best model in our experiment is not high compared with the current state-of-the-art models. Nonetheless, in other experiments from different researchers, their best models hardly have an accuracy larger than 0.4.

-Here are the examples of our model outputs compared with a reference summary.

We can see that the summary output from LSTM model is incorrect as seen in the word like Manchester United repeating itself in both subject and object.

In contrast, the outputs of other models are correct and understandable in terms of meaning and grammar.

There are two key points in the reference summary.

The first is Manchester united have made Schweinsteiger as their top target. BART base model and all T5 models identified this point successfully.

The other point is that Van Gall eyes Scheweinsteiger reunion at Old Trafford. And for this point, only BART large model can capture this detail.

To summarize, almost all the models successfully summarize one key point from the text except the LSTM model. And the sentences they generate are readable and concise.

Although none of them can point out both of two key points.

-From our analysis, we conclude that the performance of text summarization models can be improved by using transformer model. Our current models can generate a correct and readable summary sentence in some cases.

Text summarization is still a challenging topic as the overall accuracy of our models is still relatively low; and in other experiments, researchers can hardly generate a model with the accuracy higher than 0.4.

For Future work, we can incorporate Named-entity recognition to our model since large portion of summaries involve named entity.