TEXT SUMMARIZATION

Daqian Dang
Rajkumar Conjeevaram Mohan
Andy Christian

Group 3
DATS 6312: NLP
Instructor: Dr.Amir Jafari

INTRODUCTION

- Motivation
  1. Critical to stay informed on latest trends for business leaders and technologists
  2. Higher-level concepts most important for understanding trends

- Objectives
  1. Proof of Concept for News and Article Summary App
  2. Prototype version would be able to ingest N articles of a given topic and summarize each
  3. Full version would be able to summarize the summaries, giving a one page report of the recent trend/news

- Methodology
  1. Load and preprocess text from Kaggle
  2. Develop transformer model to do summaries
  3. Future development of app that lets users input desired articles/reports and produces the summaries and meta summaries, along with highlights of key data

DESCRIPTION OF DATA SET

Source: Kaggle

Link: CNN-Daily Mail News Text Summarization

Columns: Full text as input, Summarized highlights as target

- **Full Text:** news articles from CNN or Daily Mail
- **Highlights:** summary of article written by author

Number of Observations:

- **Train:** 287113
- **Val:** 13368
- **Test:** 11590

Average Article Tokens: 781

Average Summary Tokens: 56

DESCRIPTION OF NLP MODEL AND TRAINING ALGORITHM

T-5 refers to Text-to-Text Transfer Transformer

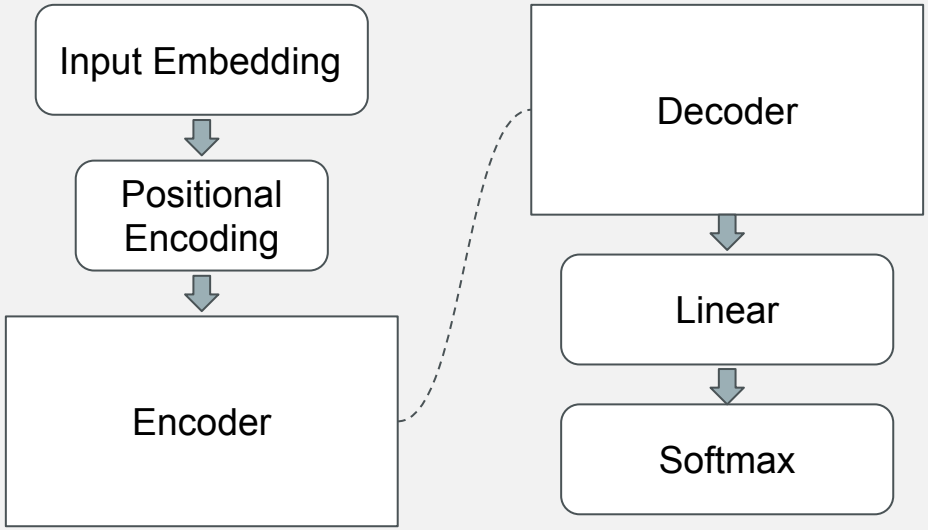
T-5 Small Model Key Characteristics

- T5-small model is a smaller version of the T5 architecture
- Faster and less computationally expensive
- Model size – approximately 60 million parameters
- Architecture – the encoder-decoder structure with 6 layers in both encoder and decoder
- Task-specific prefix – "summarize:" for summarization
- Pretraining objectives – using a denoising autoencoder setup, where the model learns to reconstruct the original text from a corrupted version with tokens masked
- Fine-tuning – after pretraining, T5 can be fine-tuned on specific task. Adjusting model weights to minimize the loss on a task-specific dataset

TRAINING ALGORITHM

- Tokenize – SentencePiece tokenizer
- Input embedding
- Positional encoding
- Encoder (multi-head self-attention, feed forward, layer norm, skip connections, dropout)
- Decoder (multi-head self-attention, feed forward, layer norm, skip connections, dropout)
- Output (linear layer, softmax)

T5 Architecture Visual Block Diagram



Cited from http://jalammar.github.io/illustrated-transformer/

Encoder:

- Multi-head self-attention layer
- Dropout
- Layer normalization (added after the residual skip connection)
- Feed forward network
- Dropout
- Layer normalization (added after the residual skip connection)

Decoder:

- Multi-head self-attention layer (decoder self-attention)
- Dropout
- Layer normalization (added after the residual skip connection)
- Multi-head cross-attention layer (encoder self-attention)
- Dropout
- Layer normalization (added after the residual skip connection)
- Feed forward network
- Dropout
- Layer normalization (added after the residual skip connection)

EXPERIMENTAL SETUP

How to use data to train and test the model?

1. The training data is tokenized, embedded and passed through the encoder and decoder
2. The produced result is then compared to the target summary using cross-entropy token-wise
3. The model continues training so long as the validation set's loss improves with subsequent epochs
4. The test set is used after training to compare produced summaries with the target version

How to implement the network and how judge the performance of the model?

1. The rouge metric is used to evaluate how well the model has summarized the text compared to a target
2. This is a variant of F1 score and can take a unigram, bigram, or max-shared sub-sequence approach

HYPERPARAMETER TUNING AND THE RESULTS

- Learning Rate - 1e-3
- Batch Size - 30
- Input Length - 1024
- Output Length - 128
- Gradient accumulation steps - 1
- Parallel calls - 10
- Number of attention heads - 8
- Linear projection size - 64
- Epochs - 30

Training loss with 256 input length:

0.9164646148681641

Rouge Score: 0.2413

Training loss with 1024 input length:

0.633148193359375

Rouge score: 0.2424

SAMPLE RESULT

Original Summary: Experts question if packed out planes are putting passengers at risk. U.S consumer advisory group says minimum space must be stipulated. Safety tests conducted on planes with more leg room than airlines offer.

Produced Summary: Tests conducted by the FAA use planes with a 31 inch pitch, a standard which on some airlines has decreased. Some airlines have 30 inches of space, while others offer as little as 28 inches. Some airlines offer as little as 28 inches.

SUMMARY AND CONCLUSION

- Working with transformers is very computationally expensive
- Need to have strategies for how to use the model well, beyond loss/metrics
- For building News Summary App
  1. Summarize each paragraph/page of an article/report
  2. Summary of the summaries of N article on a given topic
  3. Search through summaries to also provide chart of key data
  4. Interested to try zero-shot classification of the original text, human summary, and model summary

THANKS FOR WATCHING!
ANY QUESTIONS?

REFERENCES

Kaggle:

https://www.kaggle.com/datasets/gowrishankarp/newspaper-text-summarization-cnn-dailymail

Research Paper:

https://doi.org/10.48550/arXiv.1910.10683

http://jalammar.github.io/illustrated-transformer/