

Enhanced Attention Network for Question Answering System

Insop Song

Motivation

Answering Questions from given context paragraph

- Machine Comprehension, Automatic Question Answering
- End to End Deep Learning

Dataset

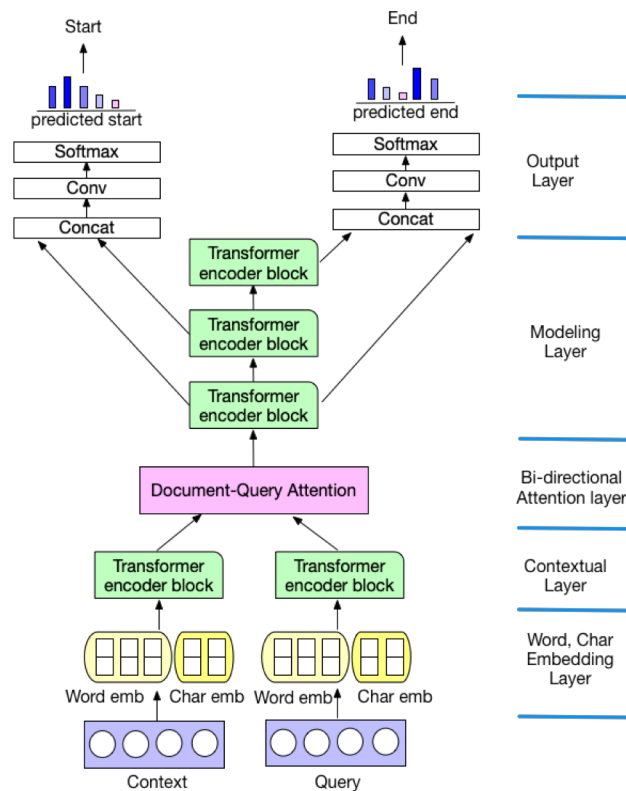
Data set:

- SQuAD (Stanford Question Answering Dataset)
- Data set from Wikipedia, 150k questions including no answer
- Example Context: "In 1973, Nixon named William E. Simon as the first Administrator"
- Question: "When was he elected by Nixon?"
- Ground truth answers: "In 1973", "1973"

Approach

- Embedding: Pretrained Word (Glove) and character embeddings
- Contextual: self-attention based encoder, Transformer instead RNN to parallelize for GPU
- Attention: find the correlation between context and question
- Modeling: capturing dependency of attention output
- Output: decoding layer to generate start and end pointer

Enhanced Attention Network



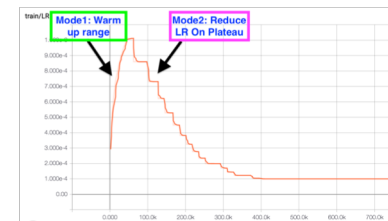
Observation

Training:

- Adaptive LR scheduling: slow warm up, dynamic LR reduction when validation plateaued
- Larger batch size training produce better score

Results:

- Dev. Set Results: EM/F1 56.7 and 59.4



Conclusion

- Added Character embedding helped the learning
- Training w/ larger batch size perform better
- Adaptive LR scheduling helps learning
- End to end deep learning method performs F1 score 56.7

References

1. Min Joon Seo, et al, Bidirectional attention flow for machine comprehension
2. Ashish Vaswani, et al., Attention is all you need
3. Adams Wei, Combining local convolution with global self-attention for reading comprehension