Question Answering using Transformers

Ichchhit Baranwal   Rohit Prasad

Abstract

Question answering refers to responding to queries asked in natural language based on a given passage. It is a task used for information retrieval in Natural Language Processing (NLP). In question answering, the background is given so that the system can relate to it and make inferences. For this case, We have decided to use transformer models which rely on an attention mechanism to relate the input and output as opposed to recurrence dependence in models like RNNs.

1. Technical Details

The transformer uses an attention mechanism that learns contextual relationships between words (or sub-words) in a text. The transformer has two different processes in its basic form: an encoder that reads the text input and a decoder that gives a prediction. For the task of Question Answering, we only require an encoder unit that captures the semantics of the text.

We have used Google BoolQ question answering Dataset (Clark et al., 2019). As a preprocessing step, we use pretrained Glove and BERT embeddings for our model. We call our Question Answering Transformer architectures as QAT-1 and QAT-2 respectively. The architectures of QAT-1 and QAT-2 is given in Figure 1 and 2.
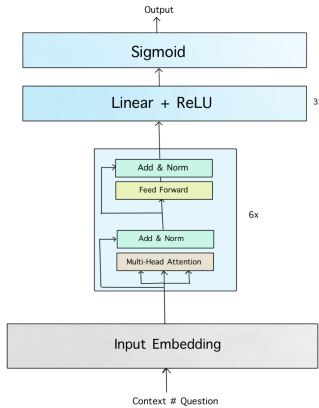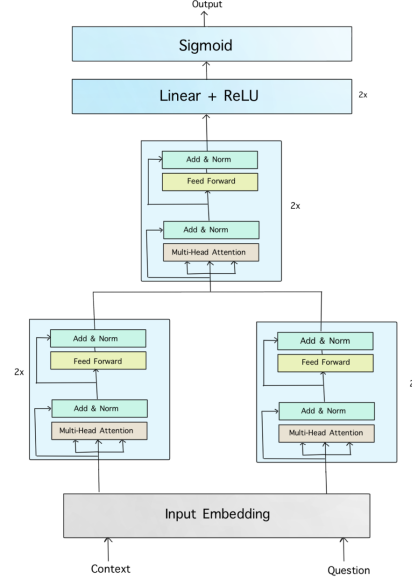

Figure 1. QAT-1

Figure 2. QAT-2

2. Results

We trained our models for 20 epochs at a learning rate of 10^{-5} with L_2 regularization value of 10^{-6} on Google Colab with 12 GB Nvidia K80 GPU which took about 1.5 hours. The result of training and subsequent evaluation is tabulated under Table 1.

MODEL	EMBEDDING	TRAIN ACC	TEST ACC	F1 SCORE
QAT-1	BERT	97.7	88.67	91.69
QAT-1	GLOVE	95.75	63.28	71.68
QAT-2	BERT	96.27	76.81	85.94
QAT-2	GLOVE	99.8	67.96	75.59

Table 1. Performance of models with different embeddings

3. Novel Contributions

We have designed two novel architectures for the task of Question Answering.

Designed and implemented architecture of QAT-1 and trained on BERT and Glove Embedding. (Rohit Prasad)

Designed and implemented architecture of QAT-2 and trained on BERT and Glove Embedding. (Ichchhit Baranwal)

4. Tools Used

- Google Colab

- pyTorch

- Numpy

- matplotlib

- spaCy

- simple-transformers

References

Clark, C., Lee, K., Chang, M.-W., Kwiatkowski, T., Collins, M., and Toutanova, K. Boolq: Exploring the surprising difficulty of natural yes/no questions. In *NAACL*, 2019.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Pennington, J., Socher, R., and Manning, C. D. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.