Question Answering Using Recurrent Neural Networks on SQUAD

Jacob James K **Kawin M**

Abstract

Question Answering is the task in which we build models capable of answering a question based on a paragraph, called context. To answer questions accurately based on the context, the model must learn the meanings and relationships of the words in the text. Thus, this is an essential problem in the field of Natural Language Processing. We tackle the problem of question answering using Recurrent Neural Networks (RNN) and attention mechanisms.

1. Technical Details

We have used the Network Architecture DrQA (1) as our baseline model. We first obtain encodings corresponding to the question and paragraph words by passing their respective embeddings through an RNN.

$$\{p_1, ...p_m\} = RNN(\{\widetilde{p_1}, \widetilde{p_2}, ...\widetilde{p_m}\})$$

$$\{q_1, ...q_l\} = RNN(\{\widetilde{q_1}, \widetilde{q_2}, ...\widetilde{q_l}\})$$

The paragraph encodings $(p_1, p_2, ...p_m)$ obtained from the RNN along with the encoding of the question q where

$$\mathbf{q} = \sum_j b_j q_j = \sum_j \frac{exp(w.q_j)}{\sum_{j'} exp(w.q_{j'})}$$

can be used to find the span of text in the paragraph corresponding to the answer as

$$P_{start} \propto exp(p_i W_s q)$$

$$P_{end} \propto exp(p_i W_e q)$$

The limitation of the above prediction step is that we treat the question and paragraph as two separate entities. In contrast, sometimes, a representation of the paragraph corresponding to a given question may help produce an exact span of answer from the paragraph. Therefore, combining the idea from the paper (4), We use a Co-Attention model (CoAtt) to get a representation of the paragraph that is aware of the question. We combine the question to paragraph representation (s_i) and the paragraph to question representation (a_i) where $i \leq i \leq m$ pass it through a BiLSTM.

$$u_1, u_2, ...u_m = BiLSTM([s_1 : a_1], ..., [s_m, a_m])$$

$u_1, u_2, ...u_m$ obtained is used in place of $p_1, p_2, ...p_m$ during the prediction step. We also explored the mechanism of using Self-Attention (SelfAtt) before applying Co-Attention and giving it as it's input. Doing so helps us in emphasize or deemphasize words in the paragraph and question before applying Co-Attention.

2. Result

We trained each of these models for 12 epochs and obtained the EM (Exact Match of Answers) and F1 (Overlap between Prediction and Ground Truth) score for the dev-set of the SQUAD 1.1 dataset after the end of each epoch (test set of SQUAD is not publicly available). We obtained the following results, where the DrQA model combined with Co-Attention gave an improvement of +0.7 EM and +0.56 F1 scores and the DrQA model combined with Co-Attention and Self-Attention showed a gain of +1.0 EM score and +1.20 F1 scores.

Table 1. Performance

	dev EM	dev F1
Base Model	67.59	77.02
DrQA + CoAttention	68.29	77.58
DrQA + CoAttention + Self Attention	68.58	78.22

3. Novel Contributions

- We studied the role of several attention mechanisms and their impact on the baseline model. The significant contribution of this project is to provide a mechanism to combine Co-Attention with Self-Attention and a mechanism to use these on the answer-span predictor from the DrQA model that gave an overall positive impact on the scores obtained. The model we proposed also converged faster than the baseline model.

- Jacob James K implemented the DrQA (baseline) model, paragraph and question encoding, and the mechanism to combine SelfAtt with CoAtt.

- Kawin M implemented CoAtt, Scaled Dot Product Attention (SDPA) and the mechanism to combine CoAtt with DrQA model and Co-Att with SDPA.

4. Tools Used

- PyTorch
- Spacy
- GloVe word embeddings
- numpy
- msgpack
- Tensorboard

References

[1] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. Reading wikipedia to answer opendomain questions. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, pp. 1870–1879, 2017.

[2] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100, 000+ questions for machine comprehension of text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1- 4, 2016, pp. 2383–2392, 2016.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000–6010.

[4] Caiming Xiong, Victor Zhong, and Richard Socher. Dynamic coattention networks for question answering. arXiv preprint arXiv:1611.01604, 2016b.

[5] https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1214/reports/final_reports/report264.pdf

[6] https://github.com/facebookresearch/DrQA

[7] https://github.com/hitvoice/DrQA