Question Answering Using RNN.

Kawin M

Jacob James K

MLSP Final Project 2021

Problem Definition

- Question Answering is the task in which we build models capable of answering a question based on a paragraph, called context.

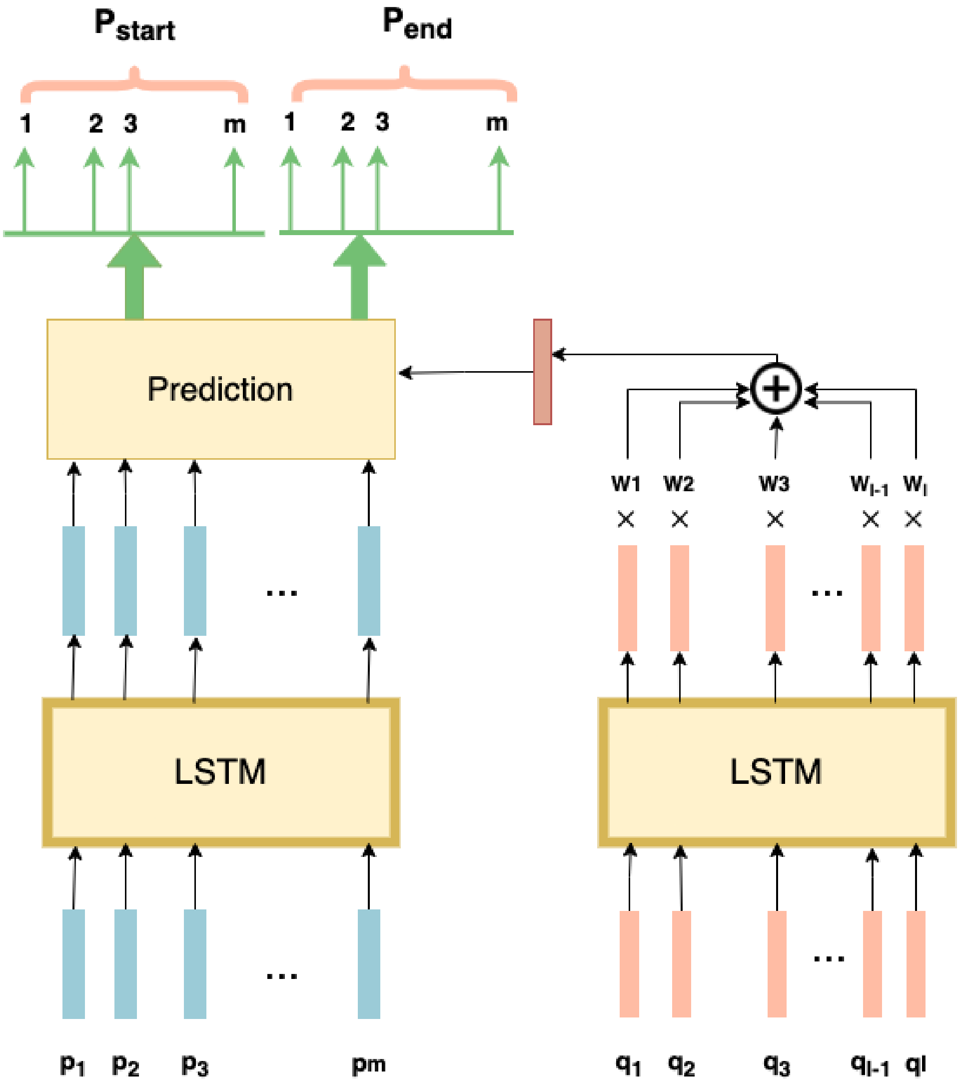
- Question Answering Systems have recently seen a significant leap mainly because of the increasing usage of chatbots and voice assistants.

- 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 and attention mechanisms.

- Our primary goal for the project is to improve upon the baseline DrQA model from the paper "**Reading Wikipedia to Answer Open-Domain Questions**" (https://arxiv.org/pdf/1704.00051.pdf) by experimenting with coattention and other attention mechanisms.

Technical Details



Network Architecture (DrQA)

- **Paragraph Encoding.**
  - $\{p_1, p_2, \ldots, p_M\} = LSTM(\{p_1, p_2, p_3 \ldots p_m\})$

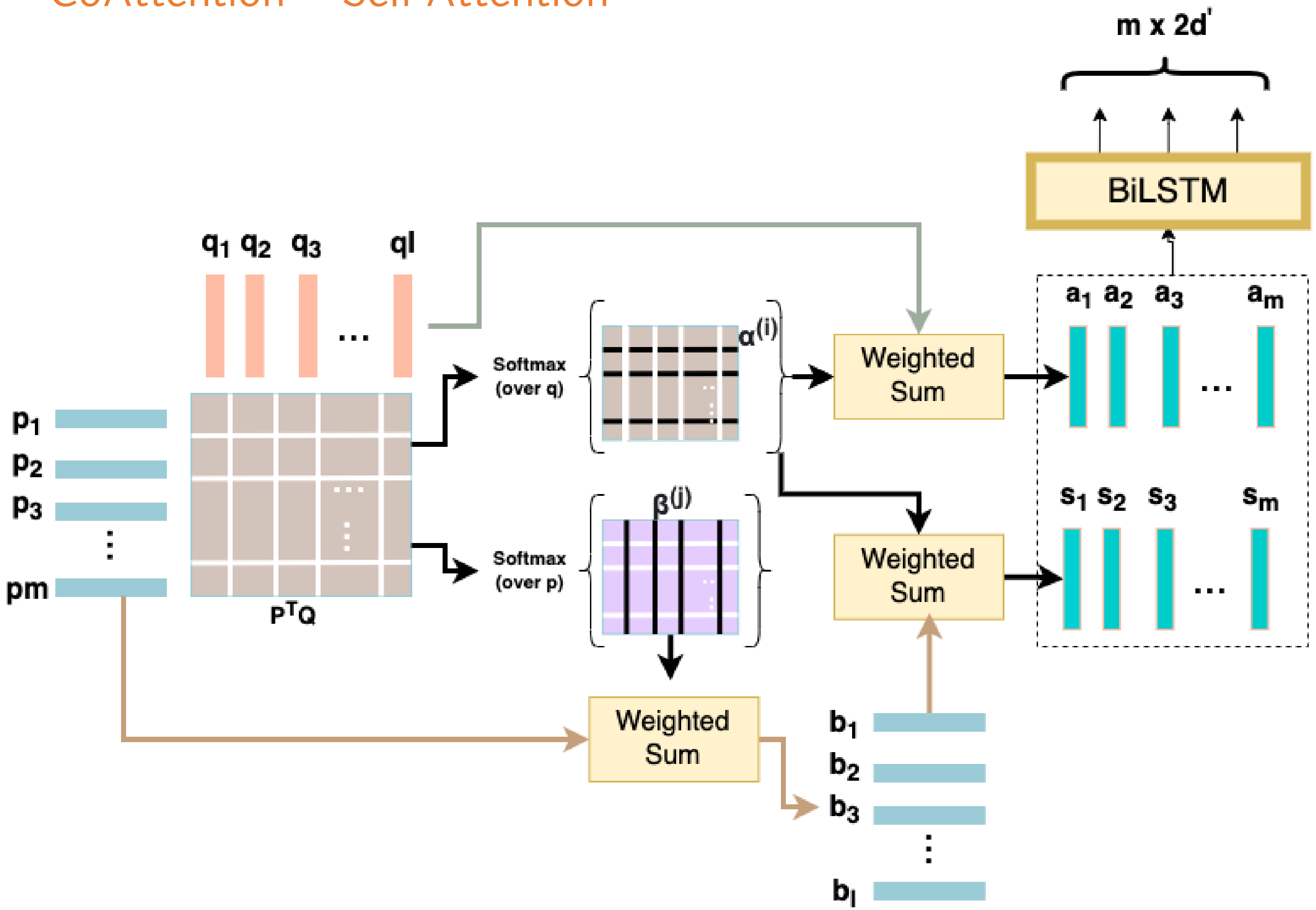
- **Question Encoding**
  1. $\{q_1, q_2, \ldots, q_l\} = LSTM(\{q_1, q_2, q_3 \ldots q_l\})$
  2. $q = \sum b_j q_j$

  Where $b_j = \frac{\exp(w.\boldsymbol{q_j})}{\sum_{j'} w.\boldsymbol{q_{j'}}}$; w is a weight vector

- **Prediction**
  - $P_{start} \propto \exp(\boldsymbol{p_i W_s q})$
  - $P_{end} \propto \exp(\boldsymbol{p_i W_e q})$

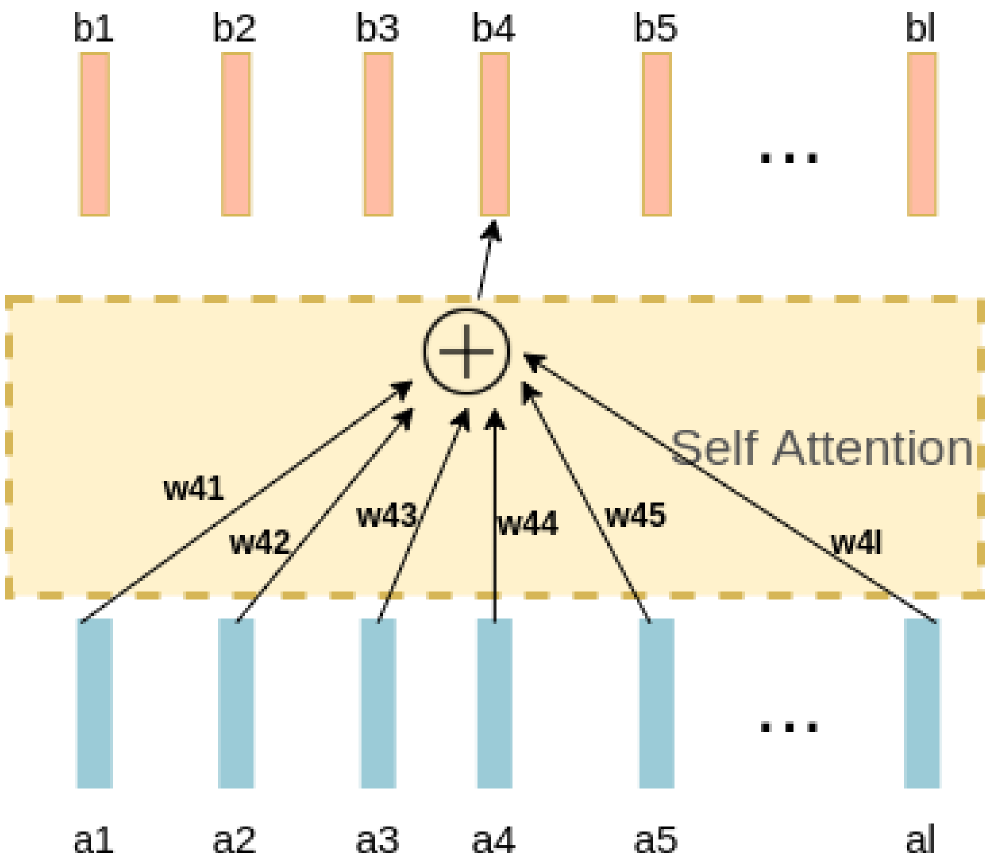
Technical Details

CoAttention + Self Attention



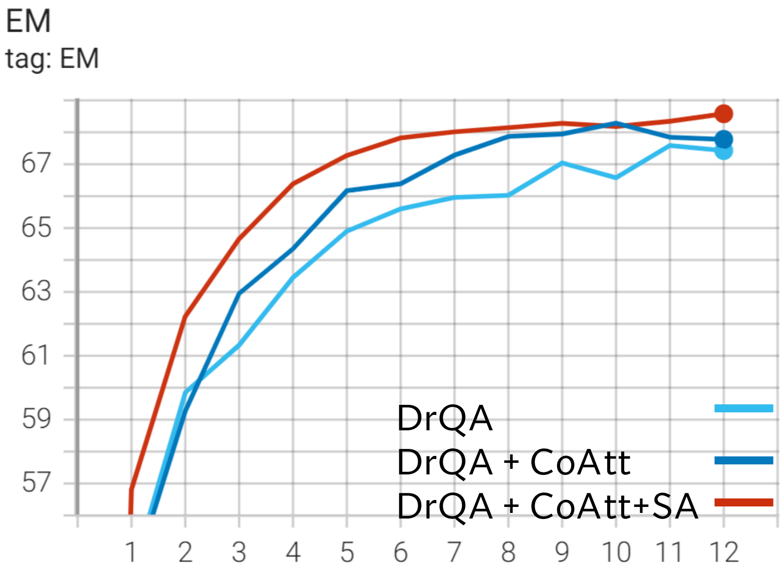
- **Affinity Matrix**
  - $L = P^T Q$
  - $P = \{p_1, p_2, p_3, .. p_m\}$ $\mathrm{Q} = \{\mathrm{q}_1, \mathrm{q}_2, .. q_l\}$
- **Context to Question Attention**
  - $\alpha^i = softmax(L_{i,:})$
  - $a_i = \sum_{k=1}^{l} \alpha_k^i \mathrm{q_k}$
- **Question to Context Attention**
  - $\beta^j = softmax(L_{:j,})$
  - $b_j = \sum_{k=1}^{m} \beta_k^j p_k$
  - $s_i = \sum_{k=1}^{l} \alpha_k^i b_k$
- $[a_i : s_i]$ is used as input to a Bi-LSTM to get question aware paragraph representation.

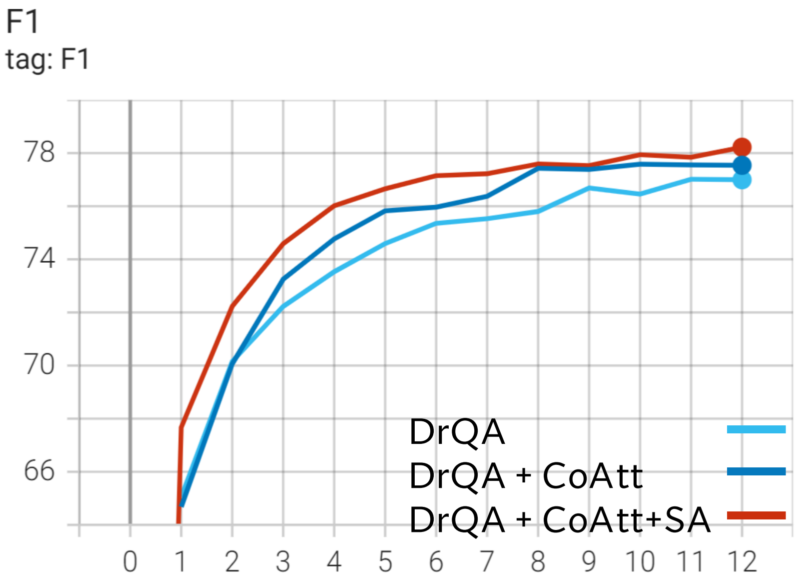
Self Attention



- Self Attention Mechanism can be used for emphasizing or deemphasizing words according to the context before co-attention is applied.
- $b_i = \sum w_{ij} a_j$

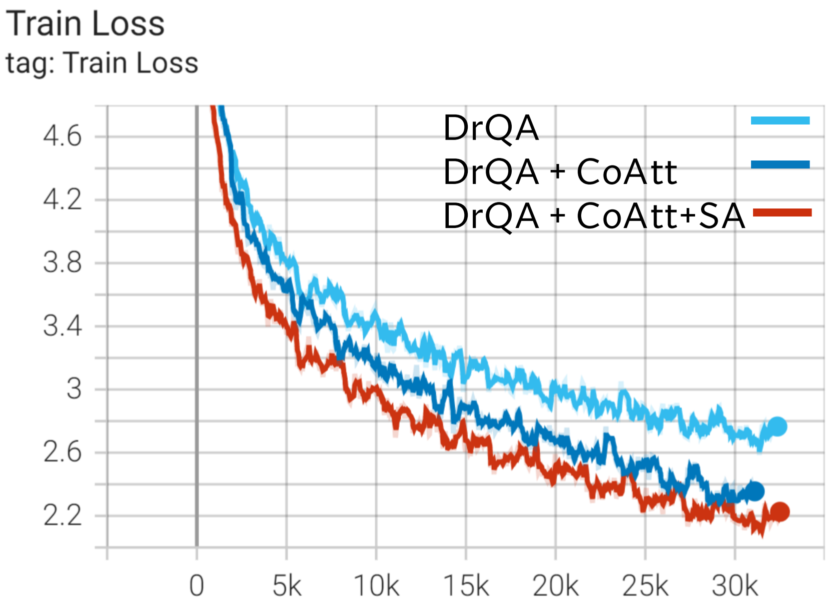
Results





Best Results

	dev EM	dev F1
DrQA (Base Model)	67.59	77.02
DrQA + Co-Attention	68.29	77.58
DrQA + Co-Attention + Self-Attention	68.58	78.22



Novel Contributions

- We have experimented with various attention mechanisms such as Simple Self-Attention, Co-Attention and Scaled Dot Product Attention and their role and impact on the baseline DrQA model.

- We used Co-Attention with the hidden states of the context and question encoding as the input, which gave an improvement of +0.7 EM and +0.56 F1 from our base model.

- To represent more relevant relationships between the context and question, we used Self-Attention combined with Co-Attention, where the input for the Co-Attention are the hidden states of Self-Attentions for the context and question encoding. We got an improvement of +1.0 EM score and +1.20 F1 score.

- Then, we combined the Triple Self Attention with Co-Attention, to make the representation even more complex and relevant, trying to capture all possible meaning.

- The intuition behind all these was to include as much relationship between the words in the context and in the question as an input to the answer span predictor to get a better model as a result.

- We then explored various choices of optimizers, dropout and more accurate language models and their behaviors on our model.