Attention based Neural Network for Machine Comprehension

CS224N Autumn 2018
Abhishek Bharani(abharani@stanford.edu), Balaji Kollu(vbkollu@stanford.edu)

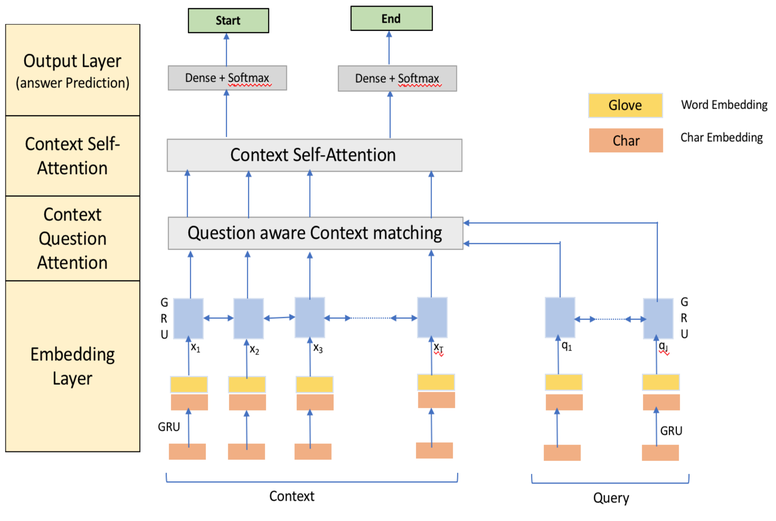
Overview

- Question answering is a special task of NLP , where computer system is presented with context paragraph and based on the question, the system predicts the answer.
- Here, we focus on reading comprehension style question answering system focused on SQuAD.
- Category prediction based on comments

Why Attention?

- Recurrent networks can only memorize limited context.
- For longer sequence, the possible answer will have no information for other parts of context and results in low accuracy.
- Attention allows modeling of dependencies without regard to their distance in the input or output sequences.

Model Architecture

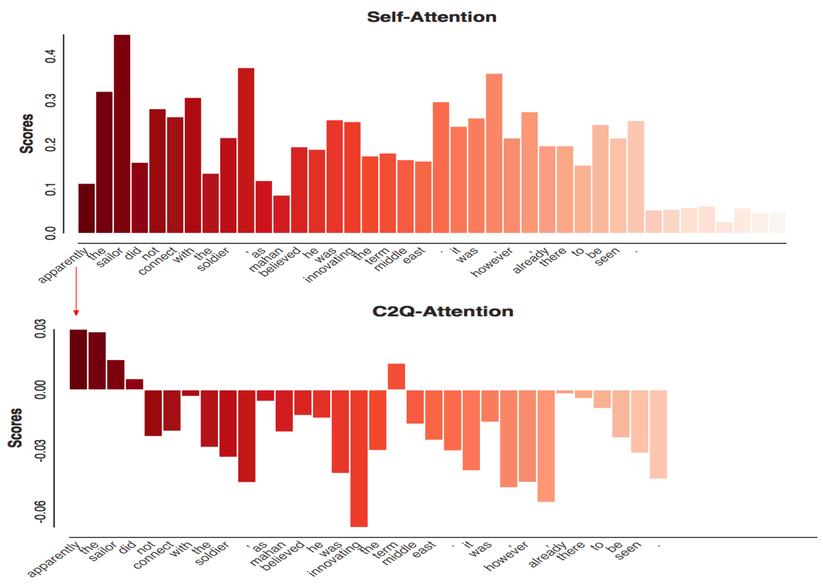


Attention Visualization

Context : apparently the sailor did not connect with the soldier , as mahan believed he was innovating the term middle east . it was , however , already there to be seen.

Question: who did not connect with the soldier ?

Answer: the sailor

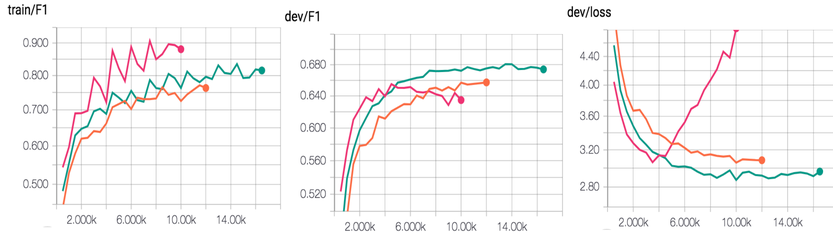


With self attention, the attention score for correct question aware word is high whereas in c2q attention we don't have the whole context representation.

Experiments

Various hyper parameters we experimented with

- lr of 0.005, 0.002, 0.001 and 0.0005; 0.001 gave best result
- dr – 0, 0.2 in all layers, and variable dr of 0.2, 0.4 & 0.5 in layer 1, 2 & 3. Model with 0.2 dr in all layers gave highest dev F1
- Changed context length to 350 which reduced run time by 10% per iteration



Performance with different drop out rates (dr)

0.2 dr Variable dr 0.0 dr

- Convex dev loss function with 0 dr
- Best dev F1 and smooth loss curve for 0.2 dr

Results

We achieved a final test score of F1 71.3 / EM 60.6 with our model.

Future Work

- Implement pointer network and char CNN.
- Try different word vector dimension and optimizations.

References

- R-NET paper from microsoft research team asia.