Question Answering with QANet

Xiluo He,[1] Zach Witzel[1]

[1]*CS 224N, Department of Computer Science, Stanford University*

Stanford
Computer Science

Problem

- Question answering systems aim to use a given passage and answer a question directly from the passage.
- However, understanding a passage of text well enough to answer questions about it is an area traditional machine learning approaches tend to struggle with as this task requires understanding over the entire passage and the systems needs to learn subtle nuances in language.
- In this work, we will explore the performance of two deep learning model architectures that worked well on SQuAD 1.1: Bidirectional Attention Flow (BiDAF) and QANet.

Models

BiDAF:
- Introduced bidirectional attention in RNNs
- Attention flows from question to the context and vice versa

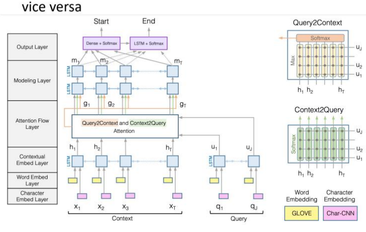

Figure 1: BiDAF Model Architecture

QANet:
- Only uses self-attention and CNNs
- Avoids the expensive training costs that came with RNNs

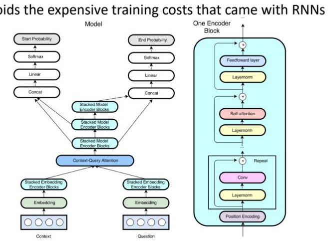

Figure 2: QANet Model Architecture

Analysis

- For all models, both the EM and F1 scores initially dropped before rising again and eventually plateauing. This is consistent with the theoretical results as the models will quickly get to 50% F1/EM score by just guessing N/A.
- We tried feeding the output of the character embedding layer into a 1D vs 2D convolutional layer. The 1D worked better than the 2D one. The QANet paper also used 1D convolutions so this is an expected result.
- Sometimes larger models would fail during training. We had to run the 7-Layer QANet twice before it worked. We believe this is due to memory limits of Azure as when we tried even larger models, they broke almost immediately.
- From qualitative inspection, the 7-Layer QANet performs especially well when the answer is N/A. It will also frequently give longer answers than the labels do, such as "$4.093 million available for disbursement" instead of just "$4.093 million".

Experiments

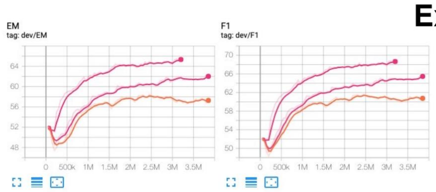

Figure 3: BiDAF (w/ Character Embedding) vs QANet Performance
(bottom = baseline BiDAF, middle = character embedding BiDAF, top = QANet)

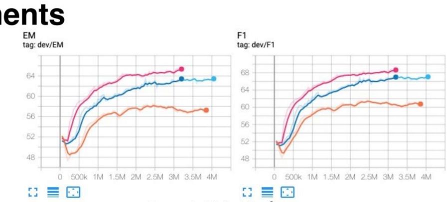

Figure 4: QANet Performance
(orange = baseline BiDAF, blue = 5-Layer QANet, red = 7-Layer QANet)

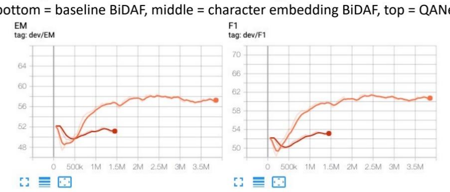

Figure 5: Co-Attention QANet vs Baseline
(red = Co-Attention QANet, orange = baseline BiDAF

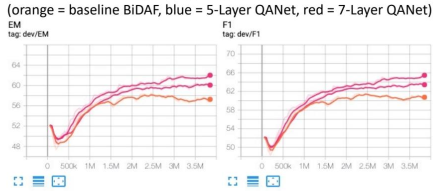

Figure 6: 1D vs 2D Convolutional Character Embedding vs Baseline
(bottom = baseline, middle = 2D Convolution, top = 1D Convolution

Conclusions

- We were able to successfully reproduce QANet from scratch and compare it to several architecture modifications. In doing so, our final result ended up with a 68.66 F1 versus the 60.75 baseline.
- We found the best success using the original QANet model architecture. All our personal modifications (10 QANet encoding block layers versus 7, co-attention versus self-attention, 2D convolutional character embeddings versus 1D) resulted in similar or worse performance compared to the original QANet design.
- We were generally limited by the number of runs we could do and the memory size available to us when training. We would frequently try a larger model architecture, but would run into a memory error partway through training.
- For future work, we would like to try testing parameterized positional encodings, letting the model learn the encoding parameters instead of requiring sinusoidal positional encodings.

References

[1] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V. Le. Qanet: Combining local convolution with global self-attention for reading comprehension. CoRR, abs/1804.09541, 2018.

[2] Min Joon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. CoRR, abs/1611.01603, 2016.

[3] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for SQuAD. In Association for Computational Linguistics (ACL), 2018.