Building a Robust QA System for IID Squad Using Coattention, Self-Attention, and Answer Pointer Network

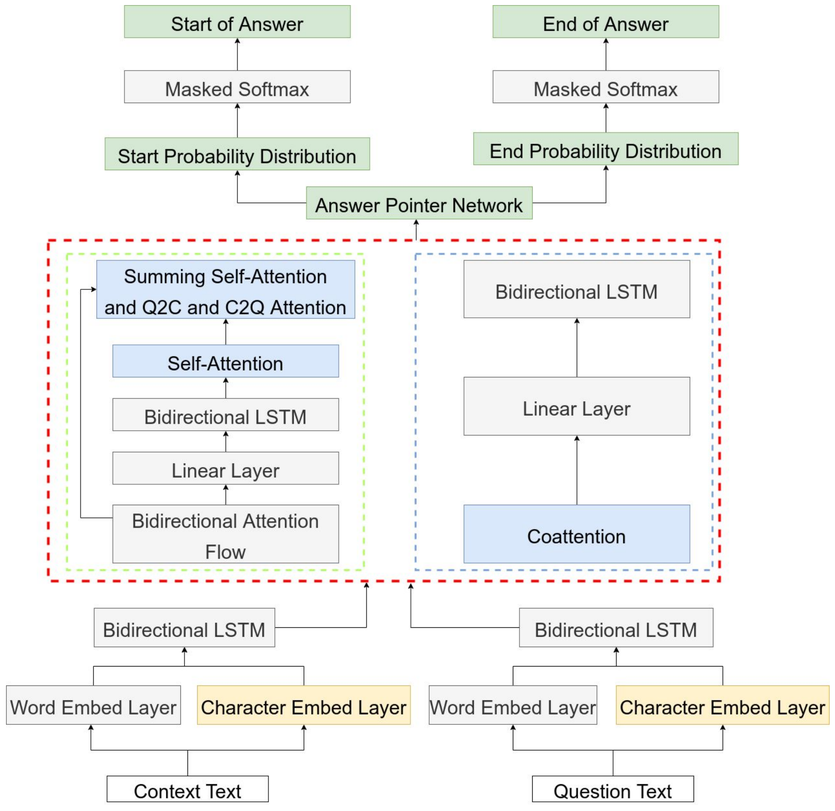
Mostafa Dewidar, George Wang, William Cai

Department of Computer Science; Department of Physics; Department of Mathematics, Stanford University

Problem and Background

- **Question answering**: important as many NLP tasks can be formulated as question answering
- **SQuAD**: > 129k data points from Wikipedia
- **Motivation**: explore effects of implementing and modifying techniques to improve traditional BiDAF model
- **Methods used**:
  - **Character-level embedding**: consider morphology
  - **Attention**: BiDAF → coattention, self-attention
  - **Answer pointer network**: condition end predictions on start predictions
- **Goal**: implement different combinations of various techniques to improve baseline

Methodology

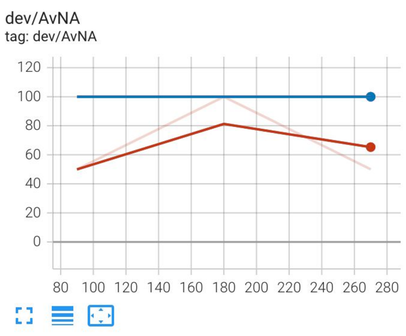


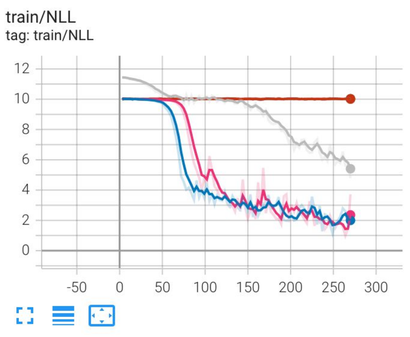
Abstract

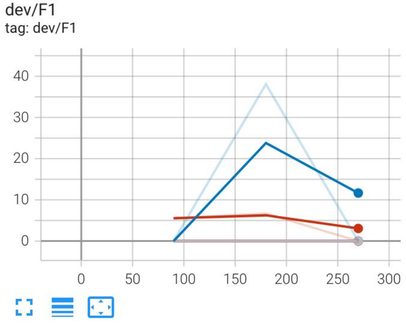
- Combinations of various techniques aiming to improve performance in question answering on SQuAD dataset
- Best model is the combination of coattention, character embeddings, and answer pointer network
  - EM = 52.20, F1 = 52.20 on training set
- No improvement yet, likely due to small training dataset used

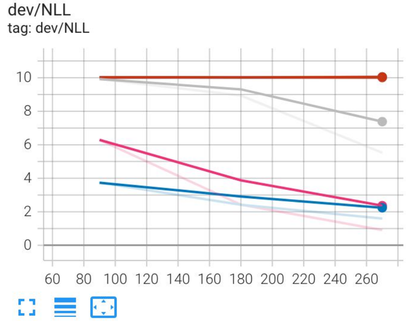
Experiments

1. Baseline (**baseline**)
2. Character-level embedding (**character**)
3. Self-attention (**selfattn**)
4. Answer pointer (**combined**)
5. Self-attention + answer pointer + character (**everything**)
6. Coattention (**coattention**)



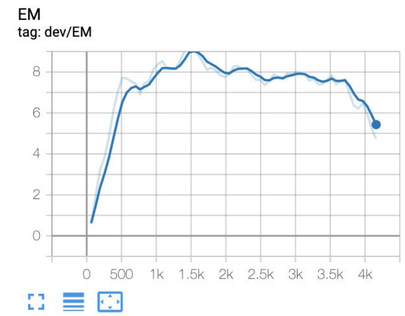




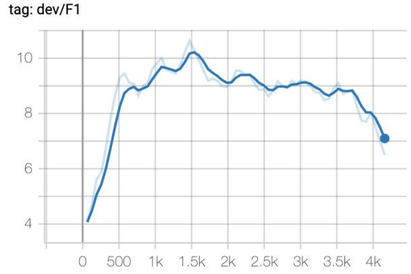


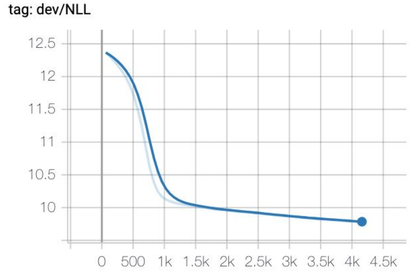
Results and Analysis

- Loss neared 0 at end of training 2e-4: overfit
- Reduced size of train dataset to ~1/10, increased # epochs (Colab's hardware limitations when Azure stopped working)
- May overfit due to training many epochs on small subset of train dataset



character embedding results (~5k)





Conclusion and Future Work

- Able to segment dataset and train different implementations of **coattention**, **self-attention**, **character embeddings**, and **conditioning**
- Identified the best combination of various techniques
- Learned how to read, understand, and implement NLP research papers; understanding large codebase; and using real-world datasets for interesting NLP applications
- With small training dataset, no performance improvement is found.
  - **Future work**: increase training size to better conclude performance

We appreciate Lucia Zheng for her advice, support, and encouragement.

Stanford University