Hit the QANet

By Chase Davis, Amanda Jacquez, and Vikas Munukutla

Problem Definition

Implement a deep learning model to more accurately answer questions in the given dataset, over the given baseline BiDAF model.

Data

We retrieve our features from the Stanford Question Answering Dataset, or, SQuAD.

Example:

[{"title": "Beyonc\u00e9", "paragraphs": [{"qas": [{"question": "When did Beyonce start becoming popular?", "id": "56be85543aeaaa14008c9063", "answers": [{"text": "in the late 1990s", "answer_start": 269}]

Future Work

Improve performance in the future by:

- Tuning hyperparameters
- Refine used features
- Polish QANet Implementation
- Implement a DCN

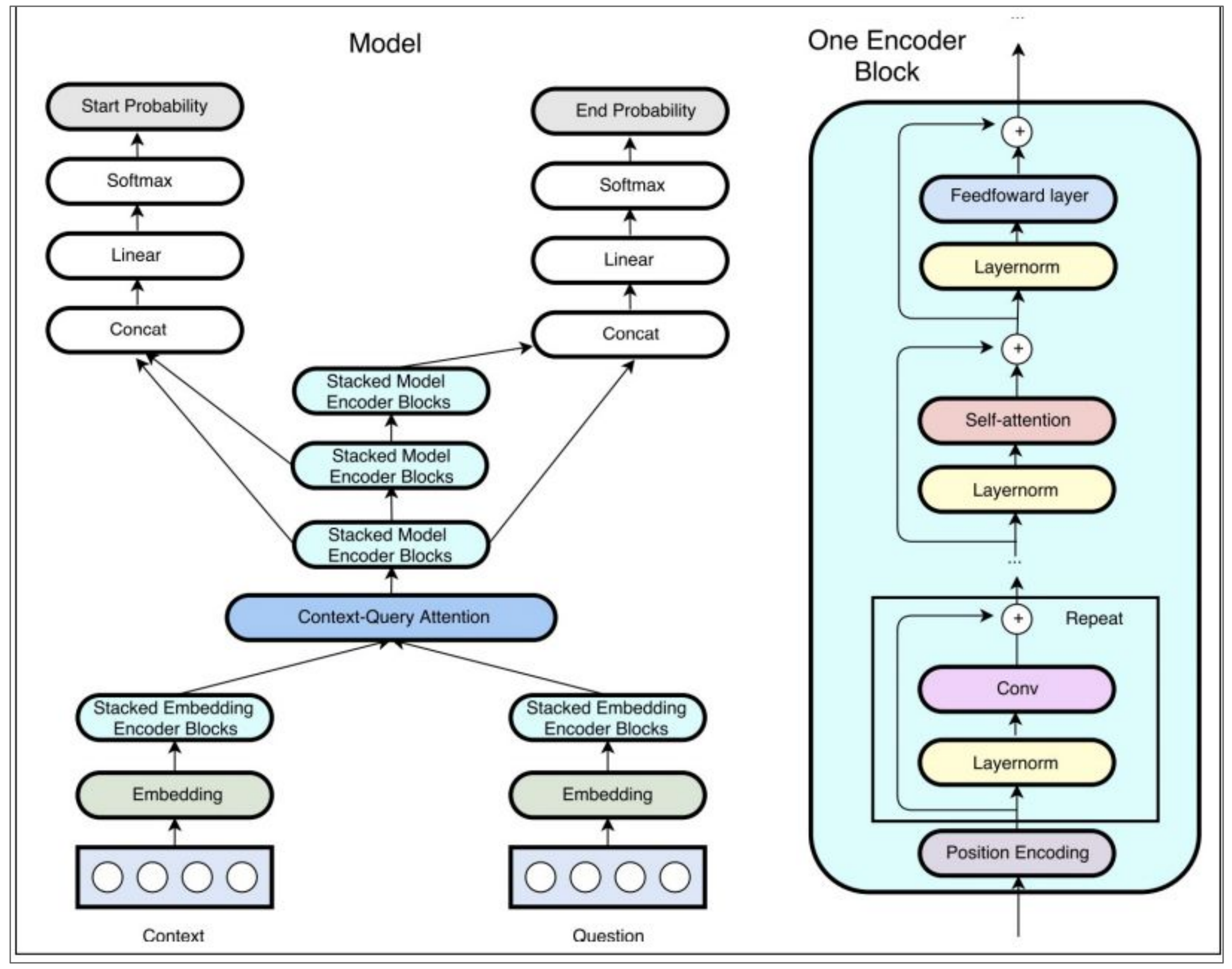
Challenges

- Limited available memory caused a number of bugs. Thus, we had to deviate from recommended parameters.
- Long training times made it difficult to tell when changes successfully improved training. For example when optimizer was changed from ADADELTA to ADAM the results were initially very promising

References

[1] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603, 2016.
[2] CS 224N Default Final Project: Question Answering on SQuAD 2.0. 28 Feb. 2019, web.stanford.edu/class/cs224n/project/default-final-project-handout.pdf
[3] Yu, Adams Wei, et al. QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension. 23 Apr. 2018, arxiv.org/pdf/1804.09541.pdf.

Approach



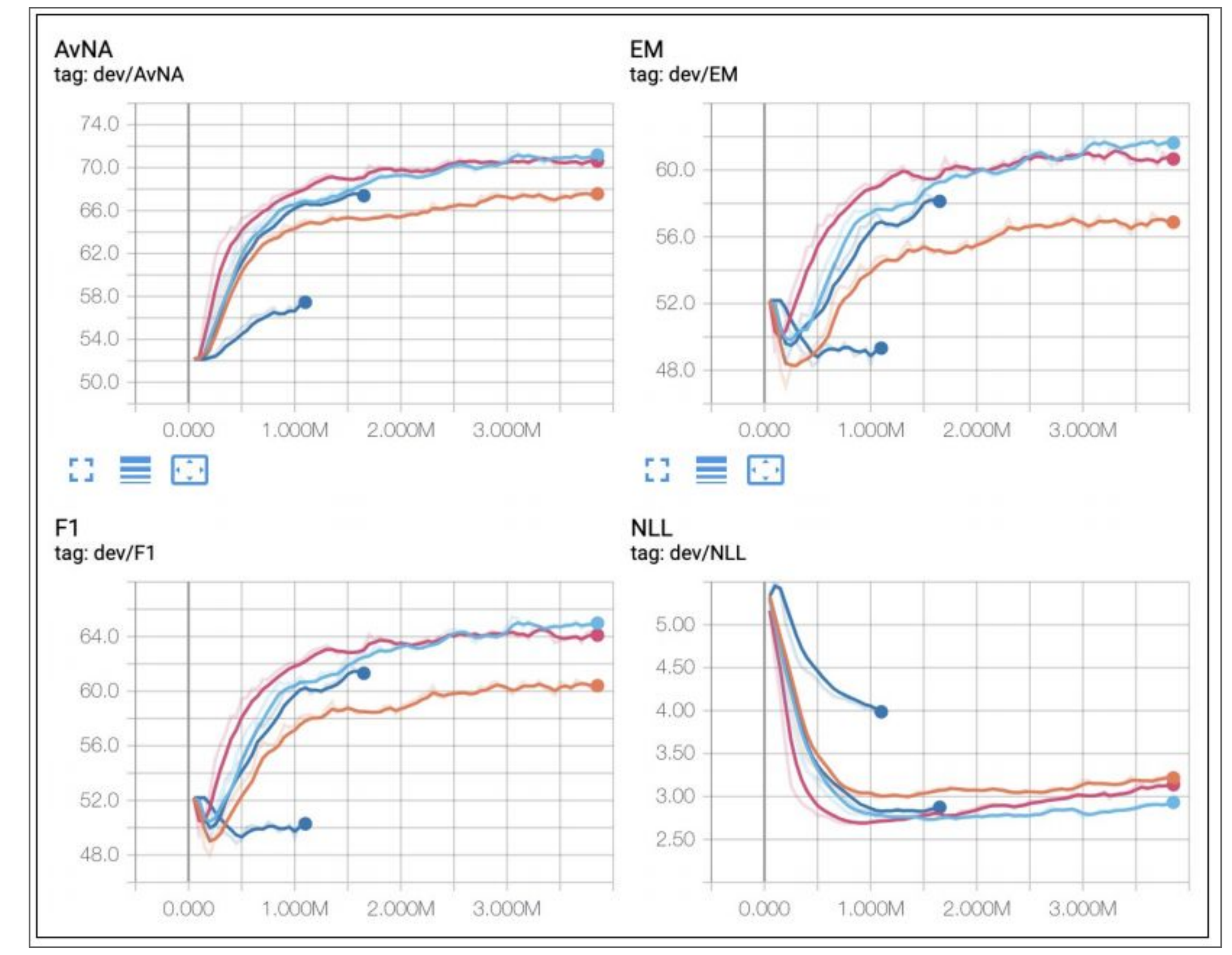
Step 1: Character Level Embeddings
-Produce a new embedding of dimension 300 by sending the character-level embedding through a CNN
-Concatenate this result with the original GLoVE 300-dimensional word vector to produce an overall embedding of dimension 600

Step 2: Encoder Blocks
-Initially adds positional encoding to the input passed in
-Repetition of similar structure, where the first is layer normalization on the input, then some function f, and then addition of the input to the output of f
-First perform convolutions, then self-attention, followed by a feed forward layer, all of which represent f in the formulation above

Step 3: Context-Query Attention
-This layer was given in the baseline implementation
-Computes similarity between the context and query

Step 4: QANet Output Layer
-Takes as input the results from each model encoder block, namely M1, M2, and M3.
-Concatenate M1 and M3, and M1 and M2 together, respectively
-Run each result through a linear layer and softmax function
-Output the probability distribution for start and end probabilities

Results



Scores: F1 : 60.101
EM : 63.716

Analysis

-Character embeddings using ADADELTA optimizer performed the best, with character embeddings using ADAM optimizer providing the second best results
-The ADAM optimizer most likely overfits data because of its added momentum
-The Character embedding models outperform the baseline in terms of subject-verb agreement

Example:

Question: *What German ruler incited Hugenot immigration?*
Answer: Frederick William

- Baseline: N/A Char Embedding: Frederick William
- Context: Frederick William, Elector of Bradenburg, invited Huguenots to settle in his realms, and a number of their descendants rose to positions of prominence in Prussia

Acknowledgments

We would like to thank the CS224n teaching staff for their guidance and support throughout the quarter!