Team Xuber: Question Answering via Modified R-NET Construction

Eric Xu, Damon Zuber

Department of Computer Science

Introduction

Following the release of the Stanford Question Answering Dataset (SQuAD), rapid innovations have been made in the realm of question answering.

> **Question**: Why was Tesla returned to Gospic?
> **Context paragraph**: On 24 March 1879, Tesla was returned to Gospic under police guard for not having a residence permit. On 17 April 1879, Milutin Tesla died at the age of 60 after contracting an unspecified illness (although some sources say that he died of a stroke). During that year, Tesla taught a large class of students in his old school, Higher Real Gymnasium, in Gospic.
> **Answer**: not having a residence permit

The dataset provides a passage and an associated question. The purpose of the dataset is to train deep learning models that can comprehend the provided passage and answer the associated question solely using information found within the passage.

Approach

To achieve better results than the baseline model, we will augmented the algorithm in three main ways:

- **Character-Level Embeddings:** to help handle out-of-vocabulary and infrequently used words.
- **Switching to a Self-Attention Layer:** to effectively aggregate evidence from the whole passage to infer the answer.
- **Using a gated recurrent unit (GRU) instead of LSTM:** to help solve vanishing gradient problems.

Related Work

Character Embeddings

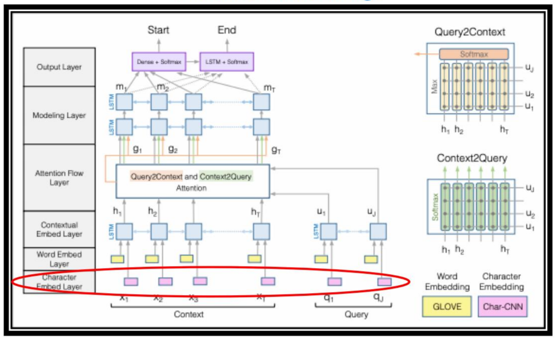

Figure 1. BiDirectional Attention Flow Model from Seo's University of Washington paper [3]. This inspired our character-level embedding implementation (circled in red).

R-NET (Self-Attention Layer)

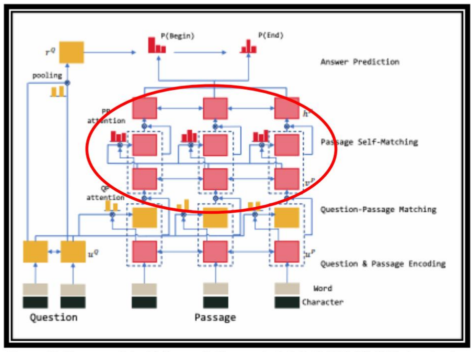

Figure 2. Flow model of Microsoft Research Asia's R-NET [2]. This inspired our self-attention implementation (circled in red).

Results

Score	EM	F1
Baseline	57.08	60.42
Milestone (Character Embeddings only)	59.02	62.62
Best dev Score	59.14	62.51
Final test Score	59.17	62.48

Conclusions

After continuously testing our model, we found that:

- Adding character-level embeddings and lowering batch size improved upon a BIDAF model.
- Replacing an LSTM cell type with GRU also improved performance.
- Self-matching attention to significantly slow down the model, and self-attention did not improve results over the BIDAF attention layer.

Our implementation ultimately **beat the baseline score**, but **fell short of improving** upon the performance of **Microsoft's R-NET model**.

Acknowledgements

Thank you to the entire CS 224N teaching team! Was an amazing class and quarter!

References

[1] CS 224N Default Final Project: Building a QA System (Iid ... Stanford University Department of Computer Science, \textit{https://web.stanford.edu/class/cs224n/project/default-final-project-handout-squad-track.pdf.}

[2] Microsoft Research Asia, Natural Language Computing Group. "R-NET: MACHINE READING COMPREHENSION WITH SELF-MATCHING NETWORKS." Natural Language Computing Group.

[3] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603, 2016.

[4] Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In \textit{Conference on Empirical Methods in Natural Language Processing} (EMNLP 2014)

[5] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In \textit{NIPS 2014 Workshop on Deep Learning}, December 2014

[6] Kaplan, J., McClandish, S., Henighan, T. "Scaling Laws for Neural Language Models." Computer Science. January 2020