Bidirectional Attention Flow Using Answer Pointer and QANet

Yuxiao Chen
yuxiaoc@stanford.edu

Qingxi Meng
qingxi@stanford.edu

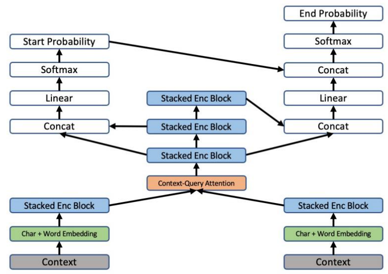
Introduction

- Machine comprehension is very important for natural language processing (NLP) research [1]
- We propose to replace the regular output layers in the baseline BiDAF with an Answer Pointer [2]
- We also explore and implement QANet [3]
  - Explore both light and complex models
  - Explore both with and without Answer Pointer

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

QANet

QANet is very similar to Transformer but its main component Encoder Block also has stacked convolutional sublayers.



Discussion

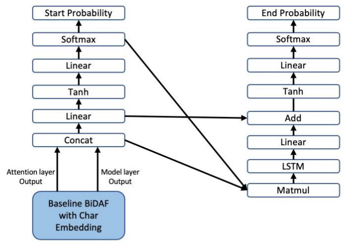
Model	Dev F1	Dev EM
BiDAF Baseline	58.26	55.05
BiDAF Baseline + Char Embedding	61.52	58.21
BiDAF Baseline + Char Embedding + Answer Pointer	62.18	59.13
QANet v1 + Char Embedding	61.69	58.01
QANet v1 + Char Embedding + Answer Pointer	64.35	60.93
QANet v2 + Char Embedding + Answer Pointer	68.29	64.98

Challenges / Limitations

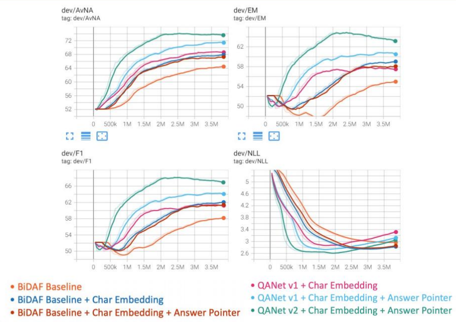
- Longer answer performs worse than shorter answer
- "Why" question type performs worse than other question types

Answer Pointer

The Answer Pointer conditions the probability distribution for the end location on the start location



Experiment



Summary

Summary

- Integrating Answer Pointer to BiDAF and QANet improves F1 and EM performance.
- Larger QANet performance better than lighter QANet.

Future Work

- Explore Transformer-XL for longer-term dependencies.
- Explore models with low memory use like Reformer.

Reference
[1] Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." arXiv preprint arXiv:1611.01603
[2] Shuohang Wang and Jing Jiang. Machine comprehension using match-lstm and answer pointer. international Conference on Learning Representations (ICLR), 2017.
[3] 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. arXiv preprint arXiv:1804.09541, 2018.