Extending QANet for SQuAD 2.0

Eun Jee Sung, Ofure Ebhomielen
Computer Science, Stanford University

Stanford
Computer Science

Problem

SQuAD2.0
Stanford University

Given a **query** (question) and **context** pair, our models have to…

- Detect if the question is answerable
- If it is, find the answer span within the context paragraph

Query

"What president eliminated the Christian position in the curriculum?"

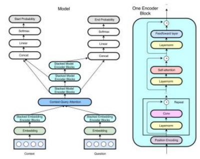
Context

"Charles W. Eliot, president 1869–1909, eliminated the favored position of Christianity from the curriculum while opening it to student self-direction. While Eliot was the most crucial figure in the secularization of American higher education, he was motivated not by a desire to secularize education, but by Transcendentalist Unitarian convictions. Derived from William Ellery Channing and Ralph Waldo Emerson, these convictions were focused on the dignity and worth of human nature, the right and ability of each person to perceive truth, and the indwelling God in each person."

Target Answer

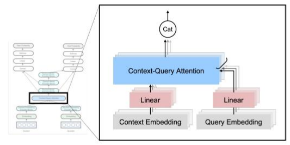
"Charles W. Eliot"

Base Model: QANet



We first implemented the **QANet** [1]. It uses the Transformer architecture and convolutions to replace RNN layers in QA models.

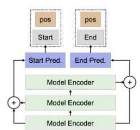
Approach 1: Multiheaded Context-Query Attention



We changed the basic context-query attention layer into a **multihead context-query attention.** We hoped to capture different types of context-query attention with multiple heads.

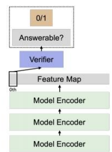
Approach 2: Output Layer Modules

Single FC Layer



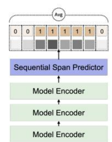
Two single fully-connected layers are each used for starting and ending point prediction.

(V) Answerability Verifier



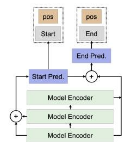
Using the feature map from the 0th index, this module predicts if a question is answerable or not.

(SEQ) Sequential Span Predictor



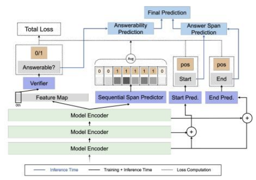
For each token in each a sequence, this module predicts if the token is in the answer span or not.

(COND) Conditioning Ending Point Pointer



The ending point predictor gets starting point prediction as one of its inputs.

Final Best Model Output Layer (Base + 1 + 2)



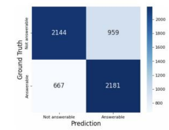
We use the auxiliary loss from (a) verifier and (b) sequential span predictor together with the basic boundary predictor loss.

[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. arXiv preprint arXiv:1804.09541, 2018.

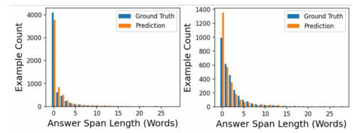
Result

Compared to the original QANet, our model achieves **a much better result for the answerability and answer span prediction**. Our verifier and sequential prediction modules also help outperform the basic QANet model. We achieved an ensemble F1 score of **69.57** and EM score of **66.54.**

Model	Embedding	AvNA	F1	EM
BiDAF (Baseline)	W	68.43	61.46	58.12
QANet (a)	W + C	68.66	61.57	57.97
QANet + V	W + C	71.16	63.96	60.28
QANet + V + SEQ (b)	W + C	**71.84**	**65.01**	**61.12**
Multihead QANet + V (c)	W + C	72.06	65.74	62.16
Multihead QANet + SEQ	W + C	**72.71**	66.4	62.85
Multihead QANet + COND	W + C	62.44	50.91	47.16
Multihead QANet + V + SEQ (d)	W + C	72.68	**66.88**	**63.38**
Multihead QANet + V + SEQ + COND	W + C	70.70	63.57	59.84
Ensemble (a+b+c+d)	W + C	**74.48**	**69.57**	**66.54**



We see high true negative and positive rate on the dataset, with the ensemble AvNA of 74.48%.



Length distribution of ground truth and predicted answer spans.

- Left: Answer span lengths of **all questions**
- Right: Answer span lengths of **answerable questions**