QAN-et al.: Exploring Extensions on QANet

Timothy Dai, Michelle Qin, Jessica Yu

Problem

Question answering is a significant challenge in the NLP space because it is one of the most effective ways to evaluate a model's understanding of language. In this project, we address a model's ability to produce the answer span of text for a question from the SQuAD 2.0 dataset. We aim to build and improve upon existing end-to-end models for question-answering tasks.

Background

In our project, we investigate the following concepts:

BiDAF Model, *Seo et al.*
- Introduces the concept of context-query attention
- Recurrent nature → slow & expensive to train

QANet Model, *Yu et al.*
- Uses convolutional layers to capture the local structure of the text and self-attention to capture the longer term interaction between words, borrowing from Vaswani et al.'s seminal work on transformers

Relative Positional Encodings, *Shaw et al., Dai et al.*
- Shaw et al. introduces method to encode relative positional information in the self-attention layer
- Dai et al. further develops relative positional encodings in Transformer-XL

Answerability, *Aubet et al., Levy et al.*
- EQuANt model by Aubet et al. adds an AvNA module to exclusively predict answerability
- Levy et al. predicts no-answer when predicting the prepended out-of-vocabulary (OOV) token

Methods

1. **Baseline**: Seo et al.'s BiDAF with slight modifications
2. **Our Implementation**: QANet described in Yu et al.
3. **Improvements on Our Vanilla QANet Model**:
   a. QANet with learnable positional encodings
   b. QANet with an AvNA module*
   c. QANet with conditioned end predictions*
   d. QANet with relative positional encodings**
4. **Assemble an Ensemble**: Select high performing models from our repertoire of BiDAF and QANet variants to achieve our highest performing result

* = includes original contributions. Ask us!
** = our best performing model (outside the ensemble)

Experiments

Model	Dev F1	Dev EM	Dev AvNA
BiDAF (baseline)	61.29	57.86	67.72
BiDAF with character-level embeddings	**63.46**	**60.14**	**69.82**
BiDAF with coattention	56.15	52.60	61.24
QANet	68.95	65.15	75.40
QANet with learnable positional encodings	69.83	66.21	76.07
QANet with relative positional encodings	**69.98**	**66.26**	**76.39**
QANet with AvNA module	68.57	64.95	74.76
QANet with conditioned end predictions	69.08	65.55	75.40
QANet + BiDAF ensemble	**72.35**	**69.43**	**77.31**

Model	Test F1	Test EM
QANet + BiDAF ensemble	70.23	67.29

Analysis

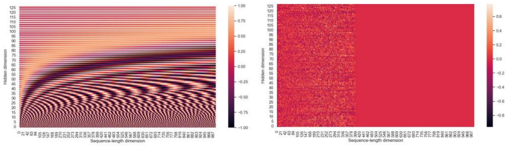
Relative Positional Encodings
- Outperforms vanilla QANet on almost all question types, especially "How" questions

Model	Dev F1 by question type						
	Who	What	When	Where	How	Why	Other
QANet	**71.70**	68.76	73.44	66.02	64.90	62.79	64.76
QANet w/ relative positional encs.	71.56	**69.71**	**73.89**	**66.26**	**68.69**	**64.64**	**66.26**

AvNA Module
- QANet is overly cautious in predicting no-answer

Discretization Method	Dev F1	Dev EM	Dev AvNA
AvNA only	66.14	62.09	73.45
AvNA && joint start-end	65.81	61.75	73.13
AvNA || joint start-end	**68.57**	**64.95**	**74.76**
Joint start-end only	68.22	64.59	74.42

Learning Positional Encodings
- No sequence-length invariance → loss of information



Conclusion

We implement a QANet model and explore extensions, from which we learn that changes in architecture are less important to improving model performance, except when addressing a bottleneck. We create an ensemble of our highest performing models, which achieves 70.23 F1 / 67.29 EM on the test leaderboard.