Word-Dropout and PointNet

Novel Methods For Question-Answering

John Peruzzi, Mason Swofford

Stanford University

Introduction

Question-Answering is an important task in modern machine learning. Given a **question** and a **context** paragraph, the model predicts the **answer**. We evaluate on the Squad 2.0 dataset, which has questions with no answers.

We experiment with 4 alterations to BiDAF: **character embeddings, word dropout, PointNet attention, and self attention**, of which word-dropout and PointNet attention are novel methods for the task.

Char-Level Embeddings

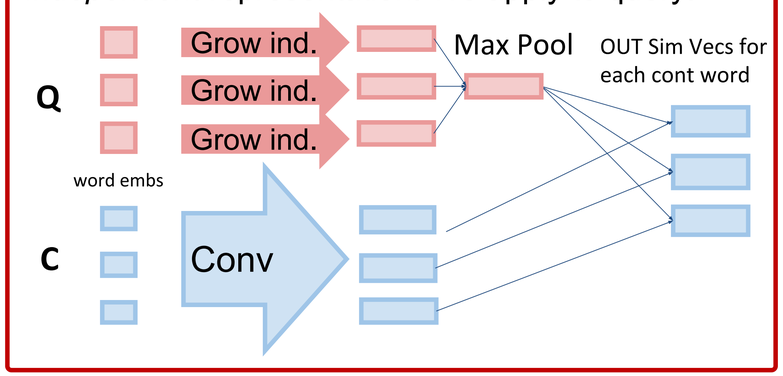
Problem: Unknown words in context and question
Solution: Character level embeddings to represent unknown words consistently.

known emb = | word emb | char emb |

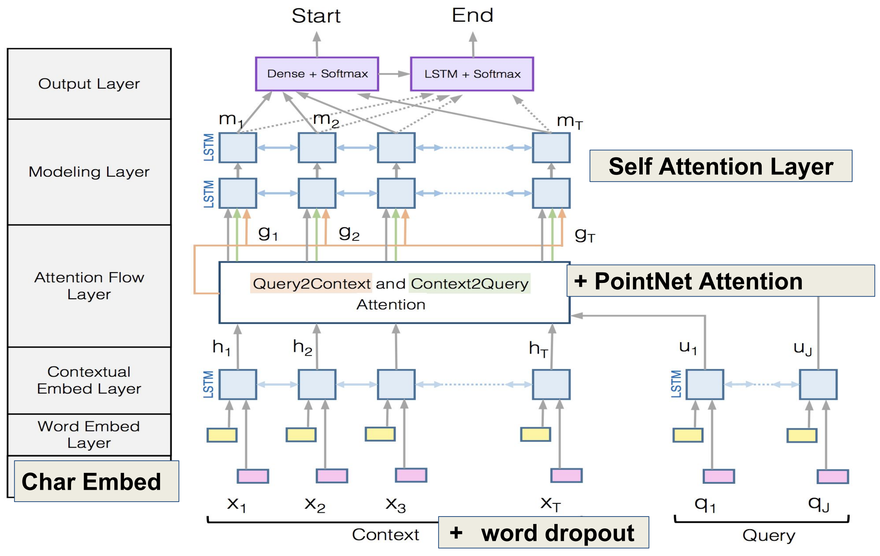
unknown emb = | <unk> | char emb |

PointNet Attention

PointNet: used in computer vision to learn *order-independent* representation. We apply to query.



BiDAF Architecture + Modifications



Self-Attention

Problem: Model is not understanding the context paragraph because context words do not communicate.

Solution: While the original BiDAF implementation contains an Attention Flow layer which exchanges information between the context and query, it relies on the modeling layer to exchange information within the context layer.

Self-Attention directly aggregates information within the context, without relying on an RNN which may have trouble capturing long term dependencies.

Error Analysis

Question: Who designed the garden for the University Library?
Context: The building was designed by architects **Marek Budzyński and Zbigniew Badowski** and opened on 15 December 1999. It is surrounded by green. The University Library garden, designed by **Irena Bajerska**, …
Answer: Irena Bajerska
Prediction: Marek Budzyński and Zbigniew Badowski

Word-Embedding Dropout

Question: Is the network simply looking for a few specific words and not understanding the context?
Experiment: Dropout words from context. Always drop out the same word everywhere.

Example: The dog chased the cat.
Dropout: < > dog chased < > cat.

Quantitative Results

- Char Embeddings: EM **+ 1.37** F1 **+ 1.28**
- Word Dropout: EM **- 2.72** F1 **- 2.13**
- Self Attention: EM **- 1.03** F1 **- 2.35**
- PointNet Attention: EM **- 0.58** F1 **- 1.23**

Discussion

- Char-level embeddings provide information about unknown words, particularly when query and context words match
- Word-Level Dropout drops important terms and prevents the model from learning
- Self-Attention and PointNet Attention overcomplicate the model and do not provide the hoped for context

References

[1] Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." *arXiv preprint arXiv:1611.01603* (2016).

[2] Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

[3] Vaswani, Ashish, et al. "Attention is all you need." *Advances in Neural Information Processing Systems*. 2017.