Introduction

Question-answering (QA) is a highly relevant area of focus in the Natural Language Processing (NLP) community. QA systems are both useful tools in themselves, and provide us with a better understanding of how well machines encode human language. Massive improvements have been made to QA models in recent years with the introduction of neural attention mechanisms. Focus on new such models, Bidirectional Attention Flow (BiDAF) [7]. While we do not propose any significant novel components, we explore the addition of character-level embeddings to BiDAF. Further, we examine the impacts of CoAttention, a secondary attention model, on BiDAF performance in order to better understand the functions of both BiDirectional and CoAttention and the interaction between the two. The models are evaluated on the Stanford Question Answering 2.0 (SQuAD 2.0) dataset [3].

Related Work

This exploration interprets the following previous work in the QA domain:

- Character embeddings use a single convolutional layer and max pooling to build character representations [1].
- Previously, QA models typically attended to small portions of the context with uni-directional attention
- BiDirectional attention represents direct query-to-context and context-to-query attention, but only performs a single pass [4]
- CoAttention attends over first-level bidirectional attention, outputting a second-level attention representation [5]

BiDAF Baseline

The baseline model is based on the BiDirectional Attention Flow model (BiDAF). BiDAF consists of a word Embedding layer, an Encoder layer, a BiDirectional Attention layer, a Modeling layer, and an Output layer. The BiDirectional Attention layer represents query-to-context and context-to-query attention. The original BiDAF model also includes character embeddings that are not included in the baseline, highlighted in green in Figure 1.

Approaches

1. Re-integrating Character embeddings to emulate original BiDAF models[5][4] Character embeddings are used in further explorations.
2. Replacing BiDirectional attention layer with a CoAttention layer described below[5]
3. Creating the Dual Attention layer by concatenating BiDirectional and CoAttention outputs.

CoAttention Layer

CoAttention includes a secondary attention computation that attends over the first-level attention output. For question hidden states $q_1, \ldots, q_n$ and context hidden states $c_1, \ldots, c_m$, a softmax over $\sum_{i=1}^{n} q_i c_i$.

Results

- Character embeddings and Dual Attention outperform BiDAF baseline
- CoAttention outperforms BiDAF baseline but reduces performance of character embedding model with BiDirectional attention
- Dual Attention outperforms all other models explored

Next Steps

This exploration highlights important attributes of the BiDAF model. Primarily, character embedding allow the model to generalize outside of the training set to unseen words by capturing sub-word meaning, greatly improving BiDAF performance. Secondly, BiDirectional attention is an integral component of BiDAF—this is the same reason it provides future layers with important first-level context to question and question-to-context information. On its own, CoAttention is a negligible contribution to BiDAF performance but likely because second-level attention does not carry enough information on its own to sustain the model. This is supported by the high performance of Dual Attention, suggesting the second level of information from CoAttention positively complements the more powerful BiDirectional attention. While I have improved upon the BiDAF baseline with character embeddings and Dual Attention, there is still great room for exploration and improvement within BiDAF and the SQuAD 2.0 question-answering task:

- exploring the effects of larger embeddings, e.g. 8-gram embeddings
- explore other attention models such as Self Attention by substituting or concatenating similar to Dual Attention
- with more resources, hyperparameter tuning in order to maximize model performance

References


CS 2282 Oodball Project Winter 2023