Exploring the effects of semantic tag augmentations for SQuAD

Jonathan Tseng 1  Narvin Phouksouvarath 2
1Computer Science, Stanford

Problem
Question-answering is a broadly researched NLP problem which has seen many contributions. Many publications that approach the QA task leverage an attention mechanism of some kind. In our project, we explore the effectiveness of augmenting three different models which use attention with character-level and semantic embeddings, and examine the qualitative effects on their attention behavior.

Dataset
We conducted our experiments on the SQuAD 1.1 Dataset.

![Image 1: Example of SQuAD Question Answer triple]

- Context paragraphs sourced from Wikipedia.
- Questions and answers created using Amazon Mechanical Turk.
- Answers are a span from the context paragraph.
- Roughly half the questions cannot be answered. Model should return N/A.

Experiments
BEAF, R-NET, and QANet are three models which use attention at different complexity levels for SQuAD. We experiment with adding three different semantic augmentations.

1. **Character-Level Embeddings**: added to the baseline BEAF model, are implemented using GRU layers applied over the character embeddings in each word token. Character embeddings can provide richer characterizations of words, especially those which are out of the model’s known vocabulary.

2. **Part of Speech**: POS tags were added to all model inputs. Adding POS tags to the model’s input may have an effect on its understanding of a word’s semantic meaning and its interactions with other tokens by giving it additional expert knowledge, as opposed to hoping it learns those relationships on its own.

3. **Named Entity Recognition**: NER tags were added to the BEAF model. Named entities include anything that can be described with a proper noun, such as a person, place, dates, etc. We hoped that NER tags would assist the model with identifying key tokens and spans in the context that would be relevant to the question, especially “who,” “when,” and “what” questions.

Analysis
- Character embeddings performance increase is attributed to improved understanding of out-of-vocabulary words.
- POS tags improved performance modestly. QANet virtually unaffected due to its complex question context representation in multi-layered R-NET.
- NER tags might have acted as a regularizer for smaller models.

Conclusion
- Character level embeddings provide significant benefits to models that initially only contain word embeddings.
- NER tags can act as an effective regularizer for smaller models, but diminishes in return for complex models.

References