QUESTION ANSWERING

EXAMINING THE TOP TRENDS AND METHODS FROM THE SQUAD TASK

INTRODUCTION

- Question answering is a research area that is gaining momentum.
- The Stanford Question Answering Dataset (SQuAD) is a well-known and large-scale dataset.
- We aim to combine these two to build a high-performance model for the SQuAD task.

PROBLEM

- Given a context paragraph and a question, the goal is to find an answer.
- There are two metrics reported for this task:
  - Exact Match (EM):
  - F1 score:

DATASET

- SQuAD 1.1 consists of 107,785 question-answer pairs from Wikipedia, where the answer to every question is a span from a single passage.

RESULTS

- Observed variations in performance across different models.

CHAR CNN

- A novel model architecture that uses character-level CNNs for improved performance.

ABSTRACTION

- The model layer provides a significant performance boost.
- We need to fix our ELMo integration since it is negatively affecting performance. It may be better to pass directly to the model layer.
- Self-attention adds only marginal benefit. Most likely due to the fact that the context only attends to itself.
- Character embeddings were more beneficial than we expected.
- BERT was more beneficial for this task.

FUTURE WORK

- Output layer that conditions the end prediction on the start prediction.
- High-performance models that are used in the BERT-A model.
- Experimenting with how to integrate ELMo and RNN layering methods such as ESET.
- Adding more features (POS tags, NER tags, EM tags, TIER features, etc.).
- Testing different optimization methods.
- Using non-RLN approach (CNN, Transformer).

REFERENCES