Neural Question Generation from SQuAD

Bilguunzaya Battogtokh, Daniel Do
{battogtokh,dktd}@cs.stanford.edu

Introduction

- Neural question generation has many practical applications such as generating educational content and non-trivial datasets for question answering.
- Our goal is given a context paragraph of text and an answer phrase, generate a relevant question.
- The baseline model uses a sequence to sequence model where the input to the encoder is the context and the decoder outputs the question.
- We used a recurrent BERT model and a seq2seq model replacing GloVe embeddings with BERT embeddings.

Dataset

- We used the SQUAD 1.1 dataset, which contains 100k+ question-answer pairs given a context text.
- The context texts are from Wikipedia, the questions are generated by crowdworkers, and the answer to each question is a segment of text inside the context text.
- This dataset contains a wide range of topics, ranging from sports to rainforests to historical plagues, which is ideal because we want our model to generalize to any corpus.

Challenges

- Existing techniques tend to perform syntactic manipulation on corpus of text rather than parse semantic meaning.
- Current approaches lack long-term dependencies between the input context and the question.
- There is no consensus on which quantitative metrics to use to evaluate model performance.
- Questions may need to take into account information from entire context paragraph instead of just from one sentence.

Approaches

- We utilized two main approaches: Replacing the encoder in Seq2seq with BERT and replacing GloVe embeddings with BERT.
- The baseline model was a simple seq2seq model using 512-dimensional GloVe embeddings.
- One of our approaches replaces the GloVe vectors with DistilBERT embedding vectors.
- GloVe vectors only depend on the input word, but DistilBERT embedding vectors also take into account the bidirectional context of the input word.
- DistilBERT has the same transformer structure as BERT, but requires only 66M parameters compared to BERT's 340M parameters.

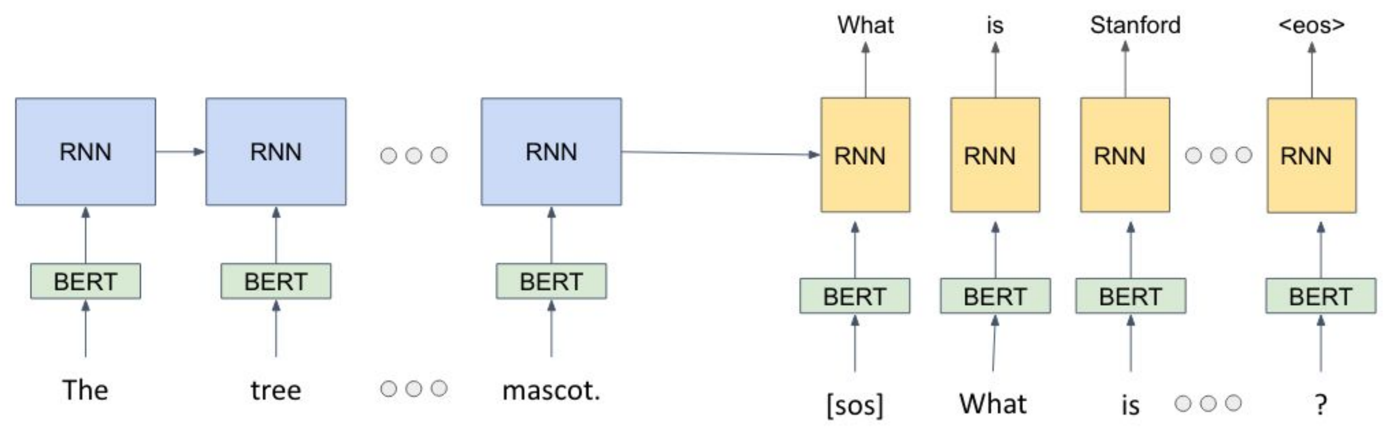

Figure 1: Approach #1

- Our second approach was to replace the encoder completely with BERT.
- For this approach, we used a version of BERT that produced sentence embeddings instead of contextual word embeddings.

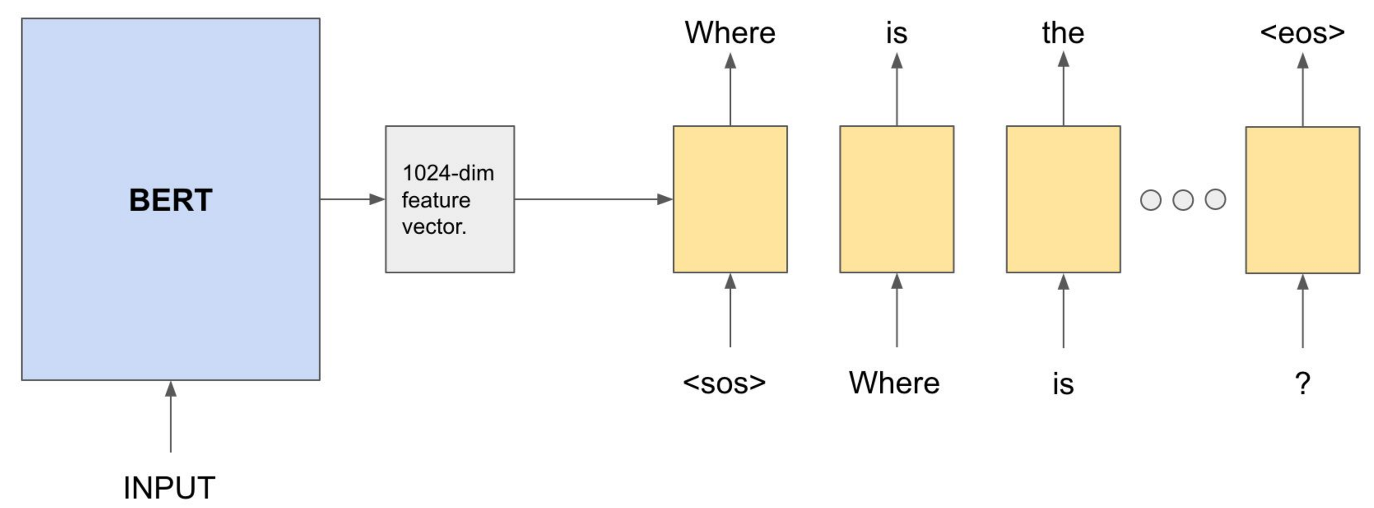

Figure 2: Approach #2

Results

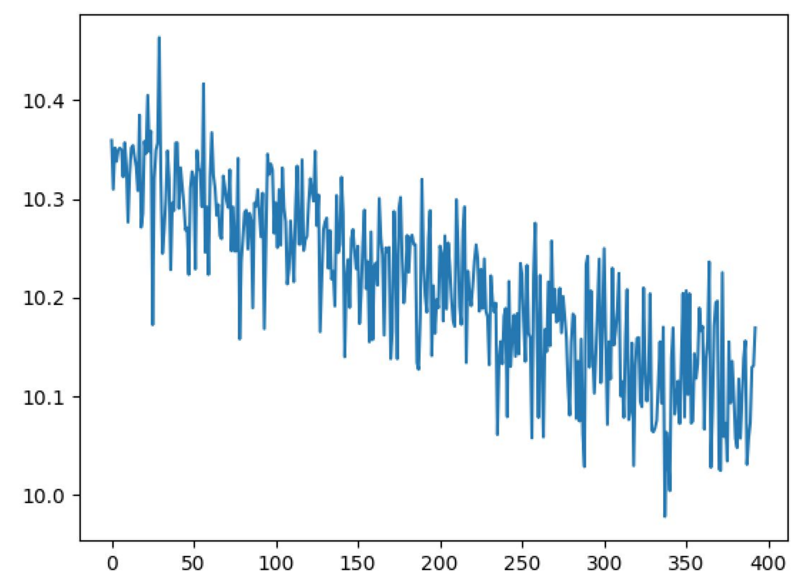

Figure 3: Loss curve for approach #1

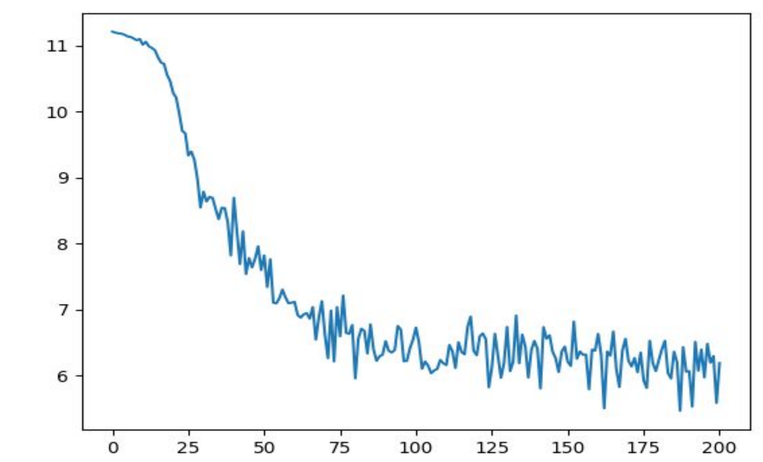

Figure 4: Loss curve for approach #2

Figures 3-4, show the results of training. We will compute the BLEU-4, ROUGE, and METEOR scores for the final report.

Analysis and Remaining Work

- We are in the progress of fine-tuning BERT for the specific question generation task
- Because of computational constraints, we did not examine context-level contexts which should provide richer questions
- We could look into using other Transformer models such as the newly released GPT-2 and XLNET that provide pretrained contextual embeddings.
- We could explore other methods of language generation such as the copy method for dialogue generation.