Motivations

Machine reading comprehension (MRC) is a challenging task, where the goal is to have systems read a text passage and then answer any question about the passage. Its importance is demonstrated by its wide applications. This task is an useful benchmark to demonstrate natural language understanding, and is applied across industries, e.g. conversational agents and customer service support.

Recently, MRC has largely benefited from the availability of large-scale benchmark datasets and it is possible to train large end-to-end neural network models.

Dataset

We trained, validated, and compared our model on the SQuAD 2.0 dataset, consisting of over 150k question-answer pairs on 500+ Wikipedia articles where the answer to each question is a span taken from the article. The major difference as compared with SQuAD 1.1 is that the updated dataset includes over 50k unanswerable questions. This forces the model to comprehend both questions and passages more thoroughly.

Evaluation

In the SQuAD 2.0 test set, we achieve 66.6 F1 and 63.7 EM score, which clearly demonstrates the challenge of MRC.

MRC Model Architecture

Encoding Layer: employs Bi-LSTM to refine the coarse word-embeddings and obtain contextual representations of both query and passage.

Co-Attention Layer: captures relationship between query and passage by using the hierarchical fusion kernel that combines representation from multiple granularities to better the model understanding and training efficiency.

Self-Attention Layer: employs a bilinear self-attention function to address the long-distance dependency within different contexts and allow contextual information to flow between passages and queries.

Matching and Output Layer: inherits the idea of stochastic dropout and bilinear matching function into the model span detector.

AvNA Classifier: a binary classifier to predict if the a given query is answerable. We apply the idea of multi-tasking to train the MRC model and classifier simultaneously.

Cross Entropy Loss Function:

Reference