Reading Comprehension with Neural Networks

Andrew Weitz
Department of Bioengineering
Stanford University
Stanford, CA 94303
aweitz@stanford.edu

Abstract

Developing algorithms that can successfully answer questions when given a paragraph of text – a task known as reading comprehension – is a challenging problem at the interface of machine learning and natural language processing. Like many other areas of machine learning, reading comprehension has become a data-driven field dominated by neural networks and deep learning. In particular, recent approaches for reading comprehension have relied on neural attention to capture the complex interactions that can exist between question and context. Here, we reimplement the high-performing Bidirectional Attention Flow model, combining it with several other previously suggested approaches, including token-specific learnable word embeddings, aligned question embeddings, and context featurization. The model achieves promising results on the Stanford Question Answering Dataset (SQuAD) with F1 and Exact Match scores of 72.3 and 61.9, respectively.

1 Introduction

Reading comprehension is a critical component in the development of modern artificial intelligence systems. It can be found in chatbots, document summarization systems, and digital assistants like Siri and Alexa. Unlike other natural language problems such as text classification, reading comprehension often requires logical inference and external world knowledge that can be difficult to explicitly define. For example, in the sentence *"The ball fell through the table because it was made of paper,"* a human reader relies on knowledge of the physical world to infer the identity of *it*.

In 2013, Burges [3] formulated the problem of reading comprehension as follows:

> A machine comprehends a passage of text if, for any question regarding that text that can be answered correctly by a majority of native speakers, that machine can provide a string which those speakers would agree both answers that question, and does not contain information irrelevant to that question.

Question answering has therefore become essential to assessing progress in reading comprehension. Here, we develop and validate a proposed reading comprehension system on the Stanford Question Answering Dataset (SQuAD). SQuAD was released in 2016 to provide researchers with a standardized benchmark test for reading comprehension performance [13]. It includes approximately 100K pairs of context paragraph, question string, and answer tokens (represented by start and stop indices within the context). The problem of question answering can therefore be formulated as predicting the start and stop indices of the answer for a given context-question pair.