Project Proposal
Answering System for Comprehension Using Match-LSTM

Natural Language Processing (CS 6120) - Spring 2018

Team : Pratik Jaipuria, Rohit Raul and Sumit Bhanwala

Abstract:

One of the core challenge of the natural language processing is to make a machine read a paragraph and then answer question based on the contextual paragraph provided.We will be representing the answer in a continuous span of index from the context paragraph where start index and end index will represent the starting pointer and ending pointer of answer from the contextual paragraph.Thus, the task at hand can be formulated as predicting the start and end index for a given pair of question and context paragraph.

Background:

Most traditional approaches to the task at hand rely heavily on the sequence of steps that involves linguistic analyses, syntax parsing, named entity recognition, question classification etc. Recently with the new approaches of combining neural networks with NLP has resulted into building end-to-end neural architectures for NLP tasks at hand. Recently, with the advances of applying neural network models in NLP, there has been much interest in building end-to-end neural architectures for various NLP tasks, including several pieces of work on machine comprehension (Hermann et al., 2015; Hill et al., 2016; Yin et al., 2016; Kadlec et al., 2016; Cui et al., 2016). However, given the properties of previous machine comprehension datasets, existing end-to-end neural architectures for the task either rely on the candidate answers (Hill et al., 2016; Yin et al., 2016) or assume that the answer is a single token (Hermann et al., 2015; Kadlec et al., 2016; Cui et al., 2016), which make these methods unsuitable for the SQuAD dataset. we propose a new end-to-end neural architecture to address the machine comprehension problem as defined in the SQuAD dataset.

Generally, a QA system includes two components: question analysis and answer retrieval. The question analysis component produces key terms, question class/category and question structure from an input question. Meanwhile, the answer retrieval component uses the information produced by the question analysis component to find an answer from a target knowledge base.

Dataset : link

For the dataset we plan on using SQuAD comprehension dataset. Stanford Question Answering Dataset (SQuAD) is a new reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage. With 100,000+

question-answer pairs on 500+ articles, SQuAD is significantly larger than previous reading comprehension datasets.

Algorithm :

Match-LSTM

Match-LSTM attempts to solve this task by layering a basic attention mechanism on top of a simple sequential feature representation for both the context and question. The system then uses another sequential model to "point" to the boundaries of the answer within the context. We will explain how the model combines the question and answer, and how its specifics could lead to deficiencies

Description

Inspired by Wang & Jiang (2016b), we seek out R-NET, an end-to-end neural network model for reading comprehension and question answering. But as basic RNN suffers from vanishing gradient problem we will be using LSTM (Long Short-Term Memory) RNNs to tackle this problem.

Our task is to predict an answer span tuple i.e. a sequence of two indices indicating the start position and end position of the answer in the paragraph, given a question and a context (paragraph). Our approach to this task is to use Match LSTM with Answer-Pointer model described in Wang et al.,2016 and Jiang et al.,2016.[3] The pointer net (Answer-Pointer) model was adopted into Match LSTM to allow us to generate multiple tokens from the original text rather than a large fixed vocabulary

Evaluation Plan :

To evaluate our model, we will be using the official evaluation script made available with SQuAD dataset, along with a sample prediction file that the script will take as input. Two metrics are utilized to evaluate model performance of SQuAD: Exact Match (EM) and F1 score. EM measures the percentage of the prediction that matches one of the ground truth answers exactly. F1 measures the overlap between the prediction and ground truth answers which takes the maximum F1 over all of the ground truth answers.

References:

[1]Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In Proceedings of the Conference on Advances in Neural Information Processing Systems, pp. 1693–1701, 2015.

[2]Felix Hill, Antoine Bordes, Sumit Chopra, and Jason Weston. The Goldilocks principle: Reading children's books with explicit memory representations. In Proceedings of the International Conference on Learning Representations, 2016

[3]Wenpeng Yin, Sebastian Ebert, and Hinrich Schutze. Attention-based convolutional neural network ¨ for machine comprehension. arXiv preprint arXiv:1602.04341, 2016

[4]Rudolf Kadlec, Martin Schmid, Ondrej Bajgar, and Jan Kleindienst. Text understanding with the attention sum reader network. In Proceedings of the Conference on Association for Computational Linguistics, 2016.

[5]Graczyk, Michael. "Implementation and Analysis of Match-LSTM for SQuAD." (2017)