Project: Machine Reading Comprehension with Deep Learning

Prasanna Kumar Challa, Prathwish Shekar Shetty, Sriram Jayaraman

Introduction

Machine reading comprehension is the task of answering a natural language question about a passage. The main objective is the rival the human performance which is about 86% on solving the task of reading comprehension. The task is to enable the computer to understand the documents and questions syntactically and semantically and also understand the types of reasoning required to answer the questions. The answers can be readily available from the corresponding passage, may not be readily available, a plausible answer could exist or the question is unanswerable based on the passage.

The aim is to provide an in-depth and thoughtful analysis of this dataset and what level of natural language understanding is needed to do well on it. This proposal will try to tweak some pre-trained models in transfer learning and deep learning to see how well they do and compare results.

Related Work

Attention based models are generally used for machine comprehension tasks to achieve high accuracies. This project aims to use improved technique called Bi-Directional Attention Flow, a multi-stage hierarchical process that represents the context at different levels of granularity and uses bidirectional attention flow mechanism to obtain a query-aware context representation without early summarization. Additionally we will also try to implement Universal Language Model Fine-tuning (ULMFiT), an effective transfer learning method that can be applied to any task in NLP, and introduce techniques that are key for fine-tuning a language model. The pre-trained models of ULMFiT has been previously applied for tasks such as text classification and topic modelling which produced great results, and this proposal hopes to extend this idea of transfer learning to Question Answering and Machine Reading comprehension tasks.

Dataset

The dataset used in this project is Stanford Question Answering Dataset (SQuAD) Version 2.0 which is a famous reading comprehension dataset, which consists questions posed by volunteers on a set of Wikipedia articles, wherein the answer to every question is a segments of text, or span, from the respective passage or at times the question is unanswerable.

The SQuAD 2.0 dataset has nearly 150k questions along with corresponding passage. The training dataset is split into two datasets, namely training and validation. And the number of questions and passages will be subject to the computational capabilities of the machines. The developed models will be tested on the Dev Set dataset provided by SQuAD.

Methodology

The project intends to do a comparative study of the methods that have been used in the papers referenced above and then use transfer learning to develop another model and compare the accuracy of the various models.

The first model that will used will be a Bi-Directional Attention Flow (BiDAF) model, which has proven to give high accuracy when used on tasks like Machine Reading Comprehension.

For the next stage of the model development, the team intends to take inspiration from the field of computer vision and use transfer learning to develop the model. We intend to experiment with some of the recent works in pre-trained models like ULMFit, ELMo, GLoMo or OpenAI transformer. The advantage of using a pre-trained model is that it allows a model to capture and learn a variety of linguistic phenomena, such as long-term dependencies and negation, from a large-scale corpus.

Evaluation

The evaluation plan is to compare the models using like metrics accuracy, F1 score and EM which are widely used in most of the machine comprehension tasks.

References

[1] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*, 2016.

[2] Jeremy Howard, Sebastian Ruder. Universal Language Model Fine-tuning for Text Classification. *arXiv preprint arXiv:1801.06146, 2018*.

[3] Stanford Question Answering Dataset https://rajpurkar.github.io/SQuAD-explorer/

[4] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:*1802.05365*, 2018*.

Extra
https://web.stanford.edu/class/cs224n/reports/2762106.pdf

https://arxiv.org/abs/1705.03551

https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2762006.pdf

https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2762006.pdf

https://github.com/allenai/document-qa/tree/master/docqa/elmo

https://thegradient.pub/nlp-imagenet/

https://medium.com/huggingface/universal-word-sentence-embeddings-ce48ddc8fc3a

https://github.com/danqi/rc-cnn-dailymail

https://rajpurkar.github.io/SQuAD-explorer/
https://arxiv.org/abs/1606.05250
ULMFit - https://arxiv.org/abs/1801.06146
{*}https://medium.com/dair-ai/a-light-introduction-to-transfer-learning-for-nlp-3e2cb56b48c8
ELMo - https://allennlp.org/elmo