EECS 595: Project Proposal
2-Stage Question Answering System Exploiting Dependency Parsing

Cheng Qian
University of Michigan
chengqia@umich.edu

Yin Yuan
University of Michigan
aoyin@umich.edu

1 Problem

This project works on the machine reading comprehension (MRC) problem, which aims to train the computer to answer certain questions based on the comprehension on related reading materials(Liu et al., 2019). It is a promising task which can be broadly used in the field of human-machine interaction, such as question answering and dialog system. Early-stage MRC systems are generally built with the hypothesis that it is guaranteed to be able to raise up an answer of given question from the related context, which leads to a lack of robustness and generalization for these systems(Zhang et al., 2020). Modern state-of-art MRC systems work on solving problems which contains questions that may not be answerable with given context, which raises up higher requirement of comprehending the contexts and questions.

Nowadays, MRC task can be classified into two parts: reading comprehension and answerability verification. Most models can perform well on finding the answer of answerable questions, while current approaches suffers from the accuracy of verifier, which is caused by the fact that these model fails to fully understand the dependency parsing information of the context. For example, given a sentence "Paris is located in France" and the question "Where is France located in?", the answer should be unknown. However, modern models are highly possible to answer "Paris" because of incomplete comprehension of the context. Inspired by the parsing technique, we want to design a MRC model which takes the parsed syntax and pos tags of context and question into account and hopefully achieve a deeper understanding about the structure within a sentence.

2 Approach

There are 2 approaches we try to use to solve the problem.

2.1 2-Stage Reading with Smaller Range

We would like to do reading comprehension 2 times. Just like people reading, the first time is to read extensively to get rough idea, and the second time is after reading the question, determine a range and read it carefully. We will input the context and question into the fine-tuned BERT model to get the word embeddings, then use [CLS] representation to do classify whether the question is answerable. The second time of reading is to use BM25 or another BERT to determine a sentence that is most related to the question. Then take this sentence and another one sentence before and after it as the accurate range of the answer. This won't loose any important information since all the questions are simple questions whose answer can be found with 1 or 2 sentences without inference. Do exactly what the first reading does to get a span of answer and the question's answerability. Weight the answerability of 2 satges to decide whether the question is answerable.

2.2 Exploit Part-of-Speech and Dependency Parsing

The existing model used to do QA cannot explore the exact relation between nominals well. To learn the dependency and relation between words in the context better, we want to utilize pos tag and parsing information. We are going to predict the pos tag of context and question, encode the tags and the pos tag encoding onto the embedding given by BERT. Dependency parsing information should be used in a similar way. The prediction of pos tag and parsing would be done by exsiting libraries such as Spacy and NLTK. This approach can be combined with the first approach.

3 Dataset

We will build our MRC model on SQuAD 2.0(Rajpurkar et al., 2016) dataset. Stanford Question

Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles. Answer to every question, if any, is a segment of text, or *span*, from the corresponding reading passage. When SQuAD updates to 2.0 version, it extends the problem definition by allowing for the possibility that no short answer exists in the provided paragraph, making the problem more realistic.

4 Previous Works

Nowadays, pre-trained language models have achieved success in MRC tasks, such as Bidirectional Encoder Representations from Transformers (BERT). BERT is is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. Pre-trained BERT model can be applied in various specific NLP problems without further modifications. What's more, thanks to the generalization of BERT, there are also models using BERT and implement their own schematics, such as Retrospective Reader(Zhang et al., 2020), U-Net(Sun et al., 2018) and SG-Net(Zhang et al., 2019).

From the results on leaderboard, we find existing models is not good at determining whether a question is answerable. Retrospective Reader(Zhang et al., 2020) introduce a 2 stage method to verify the answerability and do question-aware matching to find more accurate range of answer.

Failure to understand sentence structure is a reason for wrong answer. There has been work that try to exploit the dependency parsing information to the BERT model, called SG-Net(Zhang et al., 2019). However its practical approach is limited and does not reaches much improvement. In practice, SG-Net directly encodes all parsed pairs into arrays of ones, whose length equals the size of each pair. This approach only determine whether there is a relation but do not tell the difference of relations, which is believed to be the key to deeper reading comprehension.

Information retrieval based question answering is a broadly applied technique for finding a sentence segment as the answer for a question(Hirschman et al., 1999). We would use BM25 to find the most related sentence in order to shrink the range of passages with IR before using BERT to do comprehension work.

5 Plans And Milestones

We set three milestones for our project:

1. Milestone 1 at 11/12/2021: Finish the QA model using single BERT and use its results as the baseline, and finish designing flow chart and pseudo code of our own approaches.

2. Milestone 2 at 11/26/2021: Complete the code of two approaches with based on fine tuned BERT or its variants, and evaluate the effects of them.

3. Milestone 3 at 12/03/2021: Try to combine 2 approaches and see its effect. Finish data collection and analysis. Try potential improvement if time permits. Start preparing for project report and presentation.

6 Teamwork Division

1. Cheng Qian: Cheng is responsible for preprocessing the data in SQuAD and implementing the 2 stage approach.

2. Yin Yuan: Yuan is going to do the baseline using single BERT and try to include pos tag and dependency parsing based on BERT.

Other works such as analyzing data and writing the report will be completed by us together.

References

Lynette Hirschman, Marc Light, Eric Breck, and John D. Burger. 1999. Deep read: A reading comprehension system. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 325–332, College Park, Maryland, USA. Association for Computational Linguistics.

Shanshan Liu, Xin Zhang, Sheng Zhang, Hui Wang, and Weiming Zhang. 2019. Neural machine reading comprehension: Methods and trends. *Applied Sciences*, 9(18).

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text.

Fu Sun, Linyang Li, Xipeng Qiu, and Yang Liu. 2018. U-net: Machine reading comprehension with unanswerable questions.

Zhuosheng Zhang, Yuwei Wu, Junru Zhou, Sufeng Duan, Hai Zhao, and Rui Wang. 2019. Sg-net: Syntax-guided machine reading comprehension.

Zhuosheng Zhang, Junjie Yang, and Hai Zhao. 2020. Retrospective reader for machine reading comprehension.