

Question Answering System Based on BERT

Jiaming Li, student ID 300233503
CSI 5180 - Topics in AI: Virtual Assistant
Master of Computer Science, University of Ottawa
Final Project Definition

1 Introduction

Question answering (QA) is a significant Natural language processing (NLP) problem as well as a long-standing AI milestone. A user can ask a question in natural language and receive a rapid and concise response using QA systems. The capacity to read a piece of literature and then answer questions about it is known as reading comprehension. Reading comprehension is challenging for robots because it necessitates a combination of natural language comprehension and world knowledge. This research uses the BERT to learn a piece of text, and artificially asks some related questions based on the content of the article. The model provides answers to these relevant questions based on the content of the learned articles.

2 Goals and Limitations

- **What will you achieve (learn + produce) and what is your prior knowledge of this task?**

Hopefully the model that I am going to build can display: I give the model my own text and my own questions and see how well the answer the model does. I can ask Bert a question and it is going to give me an answer back which is indistinguishable from a human answer. That would be like the first true AI, but I am not there yet. So the actual task that is implied hereby question answering is much more constrained. But it is still really impressive what BERT is able to do on this task. And I think it's interesting to see kind of the details of how you apply it

- **Why do you want to do this project?**

Since English is not my first language, before I came to University of Ottawa, I had to take TOEFL, which is an English testing exam. There is a part called reading comprehension, which requires students to read a very long article and answer the questions based on each paragraphs. I am bad at this part, and that is why I am wondering if machine can do better than me.

- **What will be the final deliverable and by whom could it be used?**

The final project can be used for people who wants to quickly read an article and need to know some answers regrading to their questions. But they do not or they are not able to read such long article. Then this model can help them out.

- **What are the project boundaries that you are setting to be able to achieve your project within 30 hours? In other words, what is included and what is excluded of the project.**

The application of BERT to the Stanford Question Answering Dataset is what "Question Answering" refers to (SQuAD). The SQuAD benchmark has a little different task than you may expect. BERT must highlight the "span" of text matching to the correct response when given a question and a piece of text providing the answer. The SQuAD portal includes a great tool for studying the questions and reference text for this dataset, as well as showing the predictions generated by the best models.

3 Project Settings

- **Software Platform**

PyCharm, macOS, Github

- **Programming Environment**

Python 3.9.0

PyTorch 1.10.2

pytorch-transformers 1.2.0

- **Datasets - SQuAD**

The [Stanford Question Answering Dataset \(SQuAD\)](#) is a reading comprehension dataset comprised of questions posed by crowd workers on a collection of Wikipedia articles, with each question's response consisting of a text segment, or span, from the relevant reading passage, or the question being unanswerable. SQuAD includes reading comprehension questions with each passage. These questions are based on the substance of the passage and can be answered by rereading it. Finally, each question has one or more answers.

4 Activity Table

Activity	Why	Time planned	Deliverable
Research and study current articles and achievements, build background knowledge	Gather knowledge about QA system models	Have already done before writing the project definition, spent about 5 hours	None (Online resources and free video tutorial)
Explore and compare different datasets	Decide which dataset is the most suitable to the project	Have already decided to use SQuAD as the project dataset, spent 2 hours	None
Install huggingface transformers library and configure the environment	QA system requires a proper environment to run the program	Have already set the project configuration, spent nearly 2 hours	The projects can worked well on my laptop
Load Fine-Tuned BERT-large and the tokenizer	This class supports fine-tuning, and the project keeps things simpler and loads a BERT model that has already been fine-tuned for the SQuAD benchmark.	Expect to finish in 1 hour	The model was trained on version 1 of SQuAD
Build and run the model, then visualize Scores	Turn the QA process into a function so that the examples can be easily tried	Expect to train and show the plots in 8 hours	Some result plots should be generated
Test some examples of the model and compare the result both technically and artificially	See how accurate the model could be	Plan to be done in 2 to 3 hours	Answers of questions that I give to the model
Fix and improve the model	Get better results	2 hours	Output could be more accurate
Write report and record the presentation for final project	Summarize what I do for this project	7 hours	Video and paper

References

- [1] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [2] Wang, Z., Ng, P., Ma, X., Nallapati, R., Xiang, B. (2019). Multi-passage bert: A globally normalized bert model for open-domain question answering. arXiv preprint arXiv:1908.08167.
- [3] Qu, C., Yang, L., Qiu, M., Croft, W. B., Zhang, Y., Iyyer, M. (2019, July). BERT with history answer embedding for conversational question answering. In Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval (pp. 1133-1136).
- [4] Rajpurkar, P., Jia, R., Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. arXiv preprint arXiv:1806.03822.
- [5] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... Rush, A. M. (2019). Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771.
- [6] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations (pp. 38-45).
- [7] Su, D., Xu, Y., Winata, G. I., Xu, P., Kim, H., Liu, Z., Fung, P. (2019, November). Generalizing question answering system with pre-trained language model fine-tuning. In Proceedings of the 2nd Workshop on Machine Reading for Question Answering (pp. 203-211).