

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 Proposal

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 Related Work

2.1 BERT

[Bidirectional Encoder Representations from Transformers \(BERT\)](#) is one of the most extensively used and popular NLP models. BERT [1] models can look at the words that come before and after a word to determine its whole context, which is particularly valuable for determining the intent behind the question. It has a deeper sense of linguistic context and flow due to its bidirectionality, and is hence used in many NLP jobs nowadays. The performance of BERT is good and the architecture of it is worthwhile more thinking.

Pre-training with language models has been demonstrated to improve several natural language processing tasks. Among the several models, the Bidirectional Encoder Representations from Transformers (BERT) [1], which was recently disclosed by Google, is a conceptually simple but empirically powerful one. It excels in a variety of activities, including the SQUAD 2.0 competition, which we ran. The pre-trained BERT representations can be fine-tuned with additional architectures to succeed in certain tasks, according to the paper [1]. As a result, we'll design our own output network on top of the BERT pre-trained model.

2.2 Start and End Token Classifiers

BERT needs to highlight a "span" of text containing the answer—this is represented as simply predicting which token marks the start of the answer, and which token marks the end.

We feed the start token classifier the final embedding of each token in the text. The start token classifier only has one set of weights that it applies to every word (shown by the blue "start" rectangle in the above example).

We use the softmax activation to generate a probability distribution over all of the words after calculating the dot product between the output embeddings and the 'start' weights. We choose the term that has the best chance of being the start token. We repeat the process with the final token, which

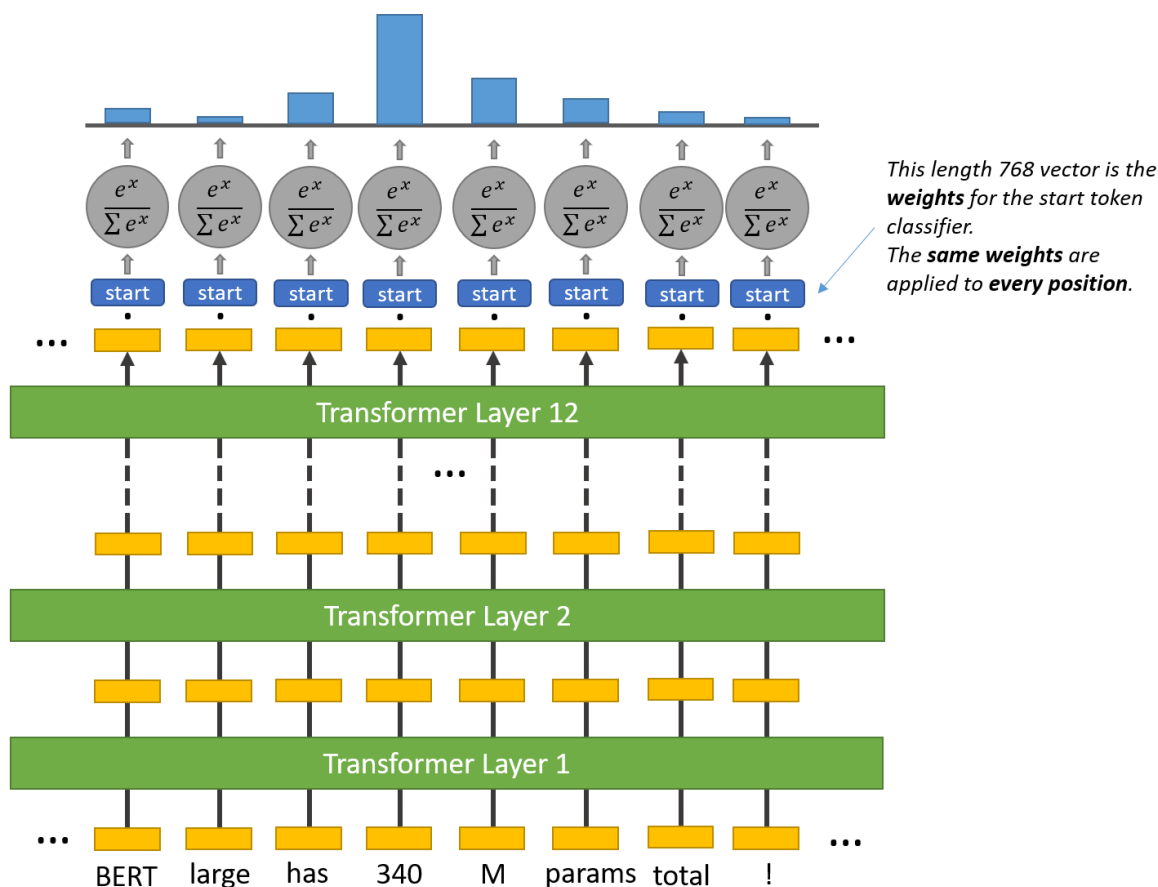


Figure 1: Start Token Classification

has its own weight vector.

2.3 HuggingFace's Transformers

The [Hugging Face Transformers package](#) is a well-known Python library that provides pre-trained models that may be used for a range of natural language processing (NLP) purposes. Text categorization is one of the most common and practical use cases for the library, which may be utilized for a variety of tasks ranging from Natural Language Inference (NLI) to Question-Answering.

3 Research Problems

- What it really means to apply BERT to QA and how to implement it?
- How to fine-tune BERT on your own dataset?
- What does it mean for BERT to achieve "human-level performance on Question Answering"?
- Is BERT the greatest search engine ever, able to find the answer to any question we pose it?

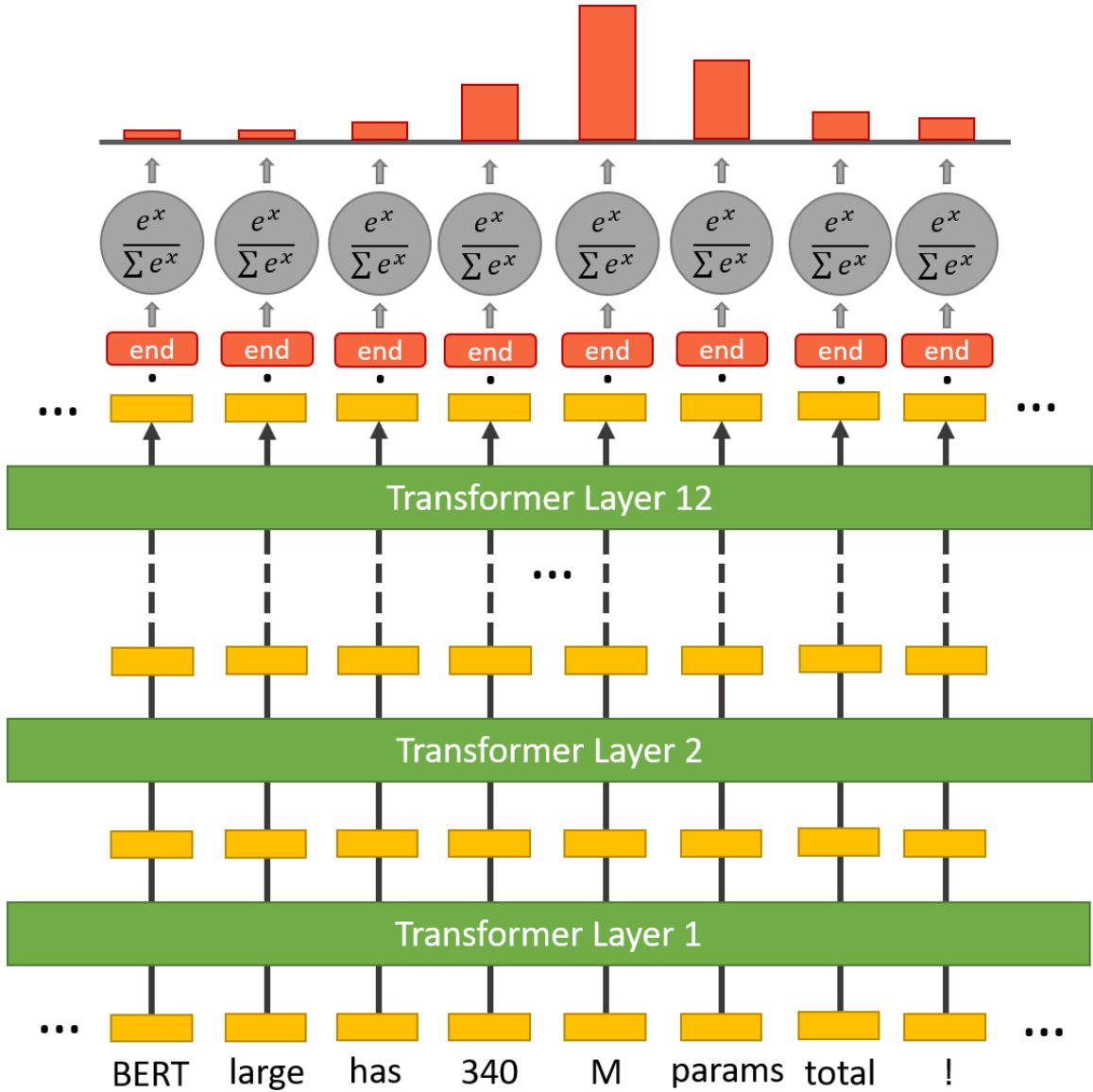


Figure 2: End Token Classification

4 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's reading sections are derived from high-quality Wikipedia pages and span a wide range of topics, from musicians to abstract concepts. A passage is a paragraph from an article that can be any length. 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.

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).