International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

Question Answering using BERT

Devansh Shukla ,Vishal Patil , Prof. Aju Rajan Palleri
Mumbai University , India

Abstract — **BERT is a machine learning framework for natural language processing that is open source (NLP).**
BERT is designed to help computers understand the meaning of ambiguous language in text by establishing context through the use of surrounding text to establish context. The BERT framework was trained using Wikipedia text and can be fine-tuned using question and answer datasets. BERT (Bidirectional Encoder Representations from Transformers) is based on Transformers, a deep learning model in which each output element is connected to each input element, and the weightings between them are dynamically calculated based on their connection. You should absolutely fine-tune BERT on your own dataset for anything like text categorization. For question answering, however, it seems like you may be able to get decent results using a model that's already been fine-tuned on the SQuAD benchmark. We'll do just that in this project, and test how well it works with text that isn't in the SQuAD dataset.

Keywords — Natural Language Processing, BERT, Question answering, SQuAD dataset.

1. Introduction

Question Answering systems are becoming more and more important and popular nowadays as it helps many domains in the industry as they can benefit from it effectively. When people talk about "Question Answering" as a BERT application, they're really talking about using BERT on the Stanford Question Answering Dataset (SQuAD).These systems contain special restrictions to ensure the quality of their material, in addition to standard user guidelines.Users can utilise Question Answering (QA) systems to get specific responses to inquiries given in natural language.The goal of this research is to discover QA methodologies, tools, and systems, as well as the metrics and indicators used to assess these approaches for QA systems, and to figure out how the link between Question Answering and Natural Language Processing is established. The Question answering model is developed by BERT. BERT's main technological breakthrough is the use of Transformer's bidirectional training to language modelling. Transformer is a popular attention model. BERT generally makes use of transformers. Transformer is an attention mechanism that learns contextual relationships between words (or sub-words) in a text.

2. Literature Survey

A.* Subjective Features of Questions Using BERT*:
Few applications have guidelines in order to maintain content quality which has major issues such as violation management, poor report quality, and feedback. As a result, in order to achieve the general goal of providing solutions for automating moderation operations on Q&A websites, the author's aim was to provide a model to predict 20 quality or subjective aspects of questions in QA websites and similar applications. The data was gathered from a crowdsource team at Google Research-2019 and a fine-tuned pre-trained BERT model.The model reached a value of 0.046 after two epochs of training, according to the Mean-Squared-Error (MSE) evaluation, which did not improve substantially in the next ones. The model achieved MSE with a value of 0.046 in predicting target values.

***B. BERT for Question Answering on SQuAD 2.0* :**
On the Stanford Question Answering Dataset, the BERT model is fine-tuned with extra task-specific layers to improve its performance (SQuAD 2.0). The authors created numerous output designs and thoroughly compared them to the BERT baseline model. On the dev set, our best-proposed single model developed an LSTM Encoder, an LSTM Decoder, and a highway network on top of the BERT basic uncased model and received an F1 score of 77.96.Model had achieved an F1 score of 77.96 on the dev set Good Performance for questions that have no answers. The only drawback of this model was the second power is the intricacy of self-awareness.

C. BERT for Question Answering on BioASQ :

Robert Slater. 2020 Bidirectional Encoder Representations from Transformers (BERT) have been very successful in language related tasks like question answering. The difficulty of the question answering task lies in developing accurate representations of language and being able to produce answers for questions. In this study, the focus is to investigate how to train and fine tune a BERT model to improve its performance on BioASQ, a challenge on large scale biomedical question answering. Our most accurate BERT model achieved an F1 score of 76.44 on BioASQ, indicating successful performance in biomedical question answering.

D. ALBERT: A Lite BERT For Self-Supervised Learning of Language Representations : GPU/TPU memory limitations and longer training time have made difficult to increase models related to natural language representations especially for downstream tasks.Keeping this in mind authors did two parameter reduction techniques to lower consumption and increase overall training speed.This is enhancement on top of original BERT. While ALBERT-xxlarge has fewer parameters and produces considerably better results than BERT-large, its larger structure makes it computationally more expensive.

E. Question Answering Using Deep Learning Stanford EDU : Recent developments in deep learning, neural network models have promise for QA. In this paper, authors have studied several deep learning models on similar lines starting with two RNN-based baselines and then focusing on memory networks which have provided state-of-the-art results on some of the important QA tasks. Two varieties of datasets have been employed in one; the output depends on general world knowledge in addition to any text provided in the dataset. In the second all information required for answering the question is provided in the dataset.

3. Proposed Work

When people talk about "Question Answering" as a BERT application, they're really talking about using BERT on the Stanford Question Answering Dataset (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.

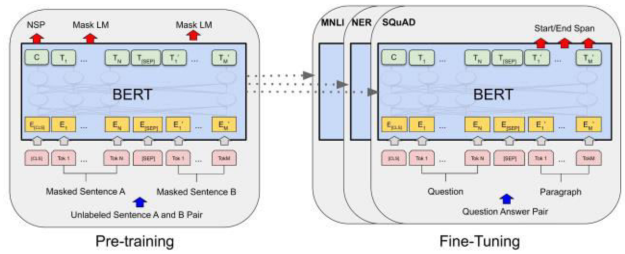

Fig. 1 : Overall pre-training and fine-tuning procedures

3.1 System Architecture

Our main idea is to add an encoder-decoder architecture on top of the PyTorch implementation of BERT baseline.. This idea may come from the computer vision area. For multi-view synthesis tasks, we always use a general auto-encoder to generate the sketch of other views and an additional auto-encoder for texture level reconstruction.

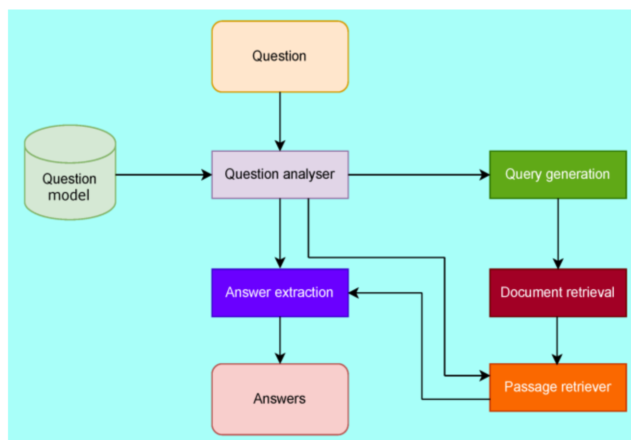

Fig. 2 : Question Answering System Architecture

To prevent the word in focus from "seeing itself," or having a fixed meaning independent of its context, BERT employs a masked language modelling technique. BERT is then compelled to recognise the masked word only on the basis of its context. Words in BERT are defined by their context rather than by a predetermined identity.

A. Input Block Description: To feed a QA task into BERT, we pack both the question and the reference text into the input.The two pieces of text are separated by the special [SEP] token. BERT also uses "Segment Embeddings" to differentiate the question from the reference text. These are simply two embeddings (for segments "A" and "B") that BERT learned, and which it adds to the token embeddings before feeding them into the input layer.

B. Output Block Description: Without making significant task-specific architectural changes, the BERT model may be finetuned with just one extra output layer to construct state-of-the-art models for a broad range of tasks, such as question answering and language inference.We pack both the question and the reference text into the input to feed a QA job into BERT and the answer is displayed along with the number of tokens. To distinguish the query from the reference text, BERT employs "Segment Embeddings." These are only two embeddings that BERT learnt (for segments "A" and "B") and inserts to the token embeddings before sending them into the input layer.

3.2. Packages and Data set :

Transformers - The Transformer in NLP is a novel architecture that aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease.The Transformer is the first transduction model relying entirely on self-attention to compute representations of its input and output without using sequence-aligned RNNs or convolution.

Torch - PyTorch-Transformers (formerly known as pytorch-pretrained-bert) is a library of state-of-the-art pre-trained models for Natural Language Processing (NLP).The library currently contains PyTorch implementations, pre-trained model weights, usage scripts and conversion utilities for the following models:

Squad 2.0 - SQuAD (Stanford Question Answering Dataset) is a 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, or the question might be unanswerable

4. Results and Evaluation

The SQuAD benchmark has a little different task than you may expect. BERT must highlight the "span" of text matching to the right 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.

4.1 Analysis of Data:

We were curious to see what the scores were for all of the words. We have visualized the scores of start and end scores.

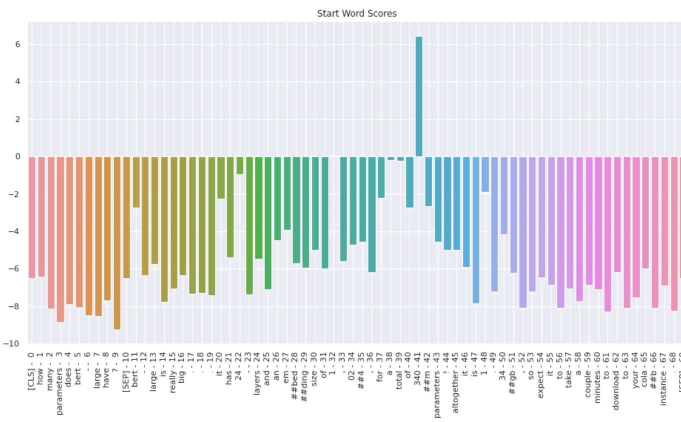

Fig. 3 : Start words Scores

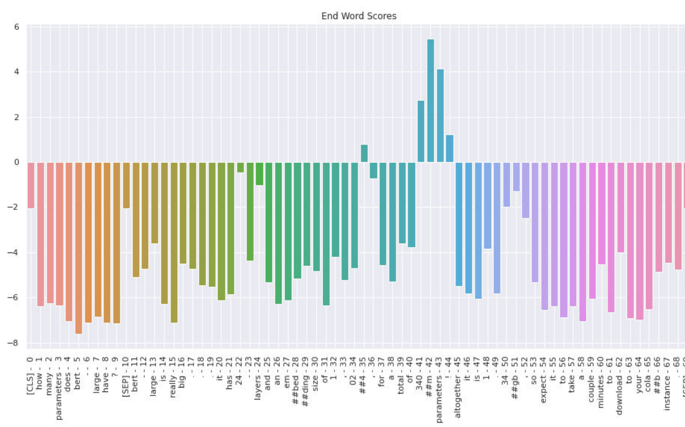

Fig. 4 : End words Scores

I also tried combining the start and finish scores into a single bar plot, but we believe that is more confusing than seeing them individually.

4.2 Evaluation:

We use two measures to assess our model's performance: the Exact Match (EM) score and the F1 score.Exact Match is a true/false binary measure of how closely the system output fits the n ground truth answer. This is a measure with a lot of requirements.F1 is the harmonic mean of accuracy and recall.

F1 = (2×precision×recall)/(precision+recall)

and it is a less stringent metric. If the system's response is a subset of the ground truth answer, it will have 100 percent precision and 50 percent recall if it only uses one of the two words in the ground truth output. When there is no response to a query, the F1 and EM scores are 1 if the model predicts no response, and 0 otherwise.

5. Future Scope

Without making significant task-specific architectural changes, the BERT model may be finetuned with just one extra output layer to construct state-of-the-art models for a wide variety of tasks, such as question answering and language inference.

ACKNOWLEDGMENT

It is our privilege to express our sincerest regards to our supervisor Prof. Aju Rajan Palleri of Pillai College of Engineering, New Panvel, for the valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of this work.

REFERENCES

1. *Predicting Subjective Features of Questions of QA Websites using BERT Issa Annamoradnejad, Mohammadamin Fazli, Jafar Habibi.*
2. *BERT for Question Answering on SQuAD 2.0 Yuwen Zhang, Zhaozhuo Xu.*
3. *Yiu-ming Chang and Qinmu Peng*, ”Eye Gaze Tracking With a Web Camera in a Desktop Environment”, IEEE Transactions on Human-Machine Systems, Volume 45, Number 4, pp. 419-430, August 2015.
4. *ALBERT: A Lite BERT For Self-Supervised Learning of Language Representations Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma Google Research Toyota Technological Institute at Chicago - 2020.*
5. *RoBERTa: A Robustly Optimized BERT Pretraining Approach Yinhan Liu, Myle Ott, Naman Goyal,Jingfei Du, Mandar Joshi Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle - 2019.*
6. BERT for Question Answering on BioASQ Eric R. Fu, Rikel Djoko, Maysam Mansor, Robert Slater Southern Methodist University - Fall 2020
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 2017-Decem(Nips), 5999–6009.
8. Question Answering Using Deep Learning by Eylon Stroh, Priyank Mathur Stanford EDU