How Much Knowledge Can You Pack Into The Parameters of a Language Model?

Shubham Gupta

February 18, 2020

1 Introduction

- This is a new paper which explores the limits of using their new T5 model in a context-free QA domain.

- As with the T5 model itself, it is very interesting to see these one-model-to-rule-them-all architectures as they exhibit some form of generalization.

- I found this paper from Adam Roberts twitter thread which is available here

- **Core Idea**: This paper will test two main things:

    - How well does the model create a knowledge base such that it can answer questions just based on this base and no other information.
    - Do model with more parameters store more information? Measuring knowledge retreiving ability is used to check this point.

2 Paper Points

- **Reading Comprehension**: Given a question and context, lookup and give the answer.

- **Open domain question answering**: Random context-independent questions. It is given entire context(all the information possible in the world) and the model is expected to deduce the answer. *Open book* exam.

- Here, problem is similar to open book exam + no context given at all. Model should retreive info from parameters and return the values. *Closed book* exam.

- T5: Treat every NLP task as text-to-text problem using encoder decoder Transformer.

- For natural questions dataset, evaluation is done as follows:

- First method:

    - Ignore all "unanswerable" and "long answer" type questions.

- model trained to output single answer
    - Questions with answers longer than 5 tokens are ignored
    - Answers normalized before comparsion
    - Answer is correct if it matches any of the annotated answers
- Second method:
    - Considered correct only if model predicts *all* the answers correctly
- For fine tuning, use AdaFactor Optimizer(need to read more about this one)

3   Results

- SOTA on Natural Questions(NQ) and WebQuestions(WQ) dataset. Worst performance on TriviaQA(TQA).
- Performance increases with model size.
- Guu et all(2020) performs better than T5 on NQ and WQ. Need to read this paper as well. It
    - Retreives Revevant documents
    - Answers questions in end-to-end fashion
- Closed-book model seem to perform on par with open-book models, leading to new research directions.
- For multiple answer type questions, T5 lower than SOTA BUT much better than baseline that was published with the paper. Therefore, T5 can perform well on these types of questions as well.

4   Shortcoming

- Model is far too expensive to train.
- Open-book models provide some indication of what information was used to answer the problem. HOWEVER, T5 just has a distribution over parameters that cannot be interpreted.
- MLE does not gurantee the model will learn a fact. Therefore, difficult to ensure the model learns specific information during pre-training
- Measure and improve performance on difficult QA tasks like DROP, which needs reasoning ability.