CSCI-GA 3033-091 IDLS Fall 2022 Project Proposal

Rushabh Musthyala, Charvi Gupta

October 2022

Title

Evaluating performance and training techniques of large sequence-to-sequence models on different QnA datasets

Team Members

Rushabh Musthyala (rm6416@nyu.edu), Charvi Gupta (cg4177@nyu.edu)

Objective

We aim to present a comparative analysis of metrics obtained by fine-tuning T5(3B) and T5-lm-adapt(250M) models on three different QA datasets using different training paradigms that take advantange of different forms of parallelism (model, data, task, etc.)

Using a multitude of metrics, we can hopefully objectively decide the best training technique for this task.

Challenges

We have chosen these two models to present an analysis of what kind of results we get for

1. Learning how to use open source implementations of transformer based models

2. Implementing fine tuning of these models on HPC NYU Greene

3. Learning how to use open source and built in functions for different types of training parallelism

Approach/Techniques

Some of the comparisons we aim to generate (but not limited to) are -

1. Train time vs dataset for each model and training technique

2. Accuracy, F1 score vs train time of model-dataset combinations

   If time permits, we will also work on other comparisons (which are not finalized yet)

Implementation Details

Hardware - at least 2 V100 GPUs on NYU Greene
Software - Pretrained models from the transformers library by HuggingFace

1. T5 (Text-to-Text Transfer Transformer): 2.8B parameters with 24-layers

2. T5-lm-adapt: 220M parameters with 12-layers

Datasets -

1. GSM8K - 8.5K high quality grade school math problems

2. TriviaQA: A Large Scale Dataset for Reading Comprehension and Question Answering

3. Natural Questions: a Benchmark for Question Answering Research

Demo Planned

We plan to present the comparisons in the form of graphs in addition to presenting the performance of the fine-tuned model on some questions asked and answers generated from the dev set. (Since some of them don't have a publicly available test set)

References

1. https://huggingface.co/transformers/v3.3.1/pretrained_models.html

2. https://arxiv.org/pdf/2110.14168v2.pdf

3. https://ai.google.com/research/NaturalQuestions

4. https://nlp.cs.washington.edu/triviaqa/