COSE474-2023F: Final Project Proposal
"Multi-Task Instruction Tuning of LLaMa with Adapter"

2020120036 Yoonjin Oh

1. Introduction

Large Language Models (LLMs) trained on large corpora of texts have attracted significant attention in recent years with their ability to solve tasks with few supervised examples. These few-shot models have shown state-of-the-art success across NLP tasks (Entity Recognition), standardized exams , as well as in subjective domains such as chatbots. The immense potential of LLMs has led researchers to explore their applicability across a multitude of industries and multitude of tasks. Numerous attempts have been made to harness LLMs in various vertical domains. However, the costs of accessing, fine-tuning and querying foundation models to perform new tasks are large. Given these costs, access to performant LLMs has been gated behind expensive and often proprietary hardware used to train models, making them inaccessible to those without substantial resources. In this study, I will explore LLaMA-Adapter as a Parameter-Efficient Fine-Tuning (PEFT) method for reducing the cost.

2. Problem definition & challenges

The problem defined in this study is, in limited time and computing resources, fine-tuning a pre-trained large language model to perform multiple tasks effectively. Tasks consist of several types of QA and summarization. It is expected to be more challenging in that it does not fine-tune just for one task, but also multiple tasks.

3. Related Works

Recent studies demonstrate that smaller foundational models, such as LLaMA (Touvron et al., 2023), can also display remarkable proficiency in tackling diverse tasks when fine-tuned using instruction-driven data. Instruction Tuning is a straightforward way to adapt LLMs for different tasks. For instance, Chung et al. (2022) train Flan-PaLM 540B, which is instruction-tuned on 1.8K tasks, and find it outperforms PaLM 540B by a large margin on unseen downstream benchmarks. Furthermore, Yue Zhang et al. (2023) demonstrate the importance of both generic and scenario-specific instruction tuning to harness LLMs' capabilities within a constrained scenario. Their study revealed that the appropriate training instructions contributed to improved performance on hold-out tasks. .

4. Datasets

For training and validation, 'databricks-dolly-15k' will be used. It is an open source dataset of instruction-following records generated by human including brainstorming, classification, closed QA, generation, information extraction, open QA, and summarization. For Evaluation, some task scenarios such as MMLU in HELM(Holistic Evaluation of Language Models) Benchmark will be used.

5. State-of-the-art methods and baselines

A family of LLaMA outperforms the existing large language models on most NLP benchmarks. The baseline model of the study is 'open-llama-7b-v2'.This is a permissively licensed open source reproduction of Meta AI's LLaMA large language model. A series of 3B, 7B and 13B models trained on 1T tokens is released. The baseline PEFT method of the study is 'LLaMA-Adapter-V2'.

6. Schedule

1. Select benchmarks for evaluation (-11/3)
2. Implement evaluation pipeline (-11/5)
3. Evaluate pre-trained model without fine-tuning (-11/10)
4. Implement data preprocessing pipeline (-11/15)
5. Implement model with LLaMA-Adapter-V2 architecture (-11/26)
6. Evaluate with HELM benchmarks (-11/29)
7. Write report (-12/3)

7. References

1. H. Touvron et al., "LLaMA: Open and Efficient Foundation Language Models," 2023.
2. H.W. Chung et al., "Scaling Instruction-Finetuned Language Models," 2023.
3. Y. Zhang et al., "Multi-Task Instruction Tuning of LLaMa for Specific Scenarios," 2023.
4. P. Gao et al., "LLaMA-Adapter V2: Parameter-Efficient Visual Instruction Model," 2023.