Parameter Efficient Fine-Tuning of LLMs towards Logical Reasoning from Images

Team: Maria Anson, Gugan Kathiresan, Aditya Shanmugham, Prabhleenkaur Bindra

Project Proposal

[High-level description of Task]
We aim to apply Parameter Efficient Fine Tuning(PEFT) or Quantized Low Rank Adaptation(QLoRA) to fine-tune LLMs as an adapter to VQA models. The goal is to achieve a fine-tuned LLM, capable of answering logical reasoning questions in the transferred domain.

[Motivation and Dataset Identified]
We have requested access to the MIT Recipe 1M+ dataset [1] and are yet to understand the dataset structure. However, the general idea from the official paper is to extract joint embeddings of image features, ingredients, and recipe instructions (Designed for image query ingredient + recipe retrieval).
Our motivation lies in the scope of using the dataset to understand the inherent relation between ingredients and recipes. LLMs can be fine tuned to achieve interactive query answering while applying logical reasoning, for eg, substituting ingredients, etc.
If time permits, we may consider enhancing the fine-tuning dataset by incorporating the VQA model to interact with images, thereby making it multimodal.

[Objective and Goals]

1. Achieve this project in the resources available to us. This would be key in supporting the claimed benefits PEFT or QLoRA to finetune LLMs in limited environments.
2. Employ the pretrained Llama 2–7b model [3] that supports a wide knowledge base (fine-tuning achievable on Colab Pro resources)
3. Perform EDA to understand how visual and textual features can be represented in a lower-dimensional space. And prepare it according to the Llama fine-tuning templates.
4. Prepare the learning paradigm in such a way that the LLM understands the relationship between ingredients and instructions/recipes.
5. We look towards PEFT or QLoRA methods for optimized finetuning and to incorporate regularization terms into the training objective of VQA models to encourage low-rank structures in the learned representations.

References

1. @article{marin2019learning, title = {Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images}, author = {Marin, Javier and Biswas, Aritro and Ofli, Ferda and Hynes, Nicholas and Salvador, Amaia and Aytar, Yusuf and Weber, Ingmar and Torralba, Antonio}, journal = {{IEEE} Trans. Pattern Anal. Mach. Intell.}, year = {2019} }
2. @inproceedings{gao2023lora, title={LoRA: A Logical Reasoning Augmented Dataset for Visual Question Answering}, author={Gao, Jingying and Wu, Qi and Blair, Alan and Pagnucco, Maurice}, booktitle={Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track}, year={2023}}
3. https://huggingface.co/meta-llama/Llama-2-7b