Overview of Large Language Models (LLMs) and Their Usage

By: Alisha Rath
SJSU ID: 016807362
Advisor: Dr. Chris Pollett

1. LLM Overview

- LLM (Large Language Model) is an AI model designed to understand and generate human language.
- Examples: GPT-3, BERT.
- Trained on massive text data.
- Tasks: Text generation, summarization, translation, etc.

2. LM vs LLM

- LM (Language Model): Traditional models, focus on smaller tasks.
- LLM: Utilizes huge datasets, more advanced, zero-shot learning capability.
- LLMs learn through self-supervised learning (next-word prediction).
- Autoregression task: Core task handled by LLMs.

3. Using LLMs

1. Prompt Engineering: Using LLM out of the box, no parameter changes.

2. Model Fine-tuning: Adjust internal parameters for specific tasks.

3. Build Your Own LLM: Create from scratch with specific datasets.

4. Prompt Engineering

- Craft inputs to maximize LLM performance.
- Easy way: Use ChatGPT, Bing, Bard.
- Programmatic way: Use OpenAI API or Hugging Face library.
- Examples: Structured text, chain of thought, chatbot personas.

5. Fine-tuning LLMs

- Adjust model parameters for better task-specific performance.
- Self-supervised learning: Predicts next word in text.
- Supervised learning: Labeled data (input-output pairs).
- Reinforcement learning: Feedback loop using human rankings.

6. Fine-tuning Approaches

- Retrain all parameters (expensive).
- Transfer learning: Freeze most parameters, fine-tune last layer.
- Parameter Efficient Fine-tuning (PEFT): Freeze parameters, add small trainable layers (LoRA).

7. Hugging Face Transformers Library

- Open-source Python library for downloading and training pre-trained models.
- Supports NLP tasks like summarization, translation, and text generation.
- Example: Sentiment analysis using pre-trained models.

8. Example: Supervised Fine-tuning

- Fine-tune a pre-trained LLM (e.g., using Hugging Face Transformers).
- Customize model attributes for specific tasks.
- Example: Sentiment analysis, text classification.

Code

https://colab.research.google.com/drive/1mAP8d-pjD0D4BKi0UVlcfie-kuLa4-QI#scrollTo=cZwUO9yMvJwY

References

1. Survey of Large Language Models : https://arxiv.org/abs/2303.18223
2. GPT 3 paper: https://arxiv.org/abs/2005.14165
3. Improving Language Understanding by Generative Pre-Training: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
4. https://huggingface.co/docs/transformers/main_classes/pipelines
5. https://arxiv.org/abs/2106.09685
6. Youtube: https://www.youtube.com/watch?v=tFHeUSJAYbE&list=PLz-ep5RbHosU2hnz5ejezwaYpdMutMVB0

Thank you