Project Proposal: Option 1

Pretrained Transformers As Universal Computation Engines: A Case Study with GPT-2

Group Members: Zihan Li, Yinggan Xu, Keling Yao, Changyi Yang

Backgroung knowledge: CNN & RNN; Transformer; Fine-tuning & Transfer Learning
Hardware Requirement: Computational power distributed on colab for regular homework is well sufficient
Proposed Homework Layout:
The core idea we hope to demonstrate in the homework is that large language model pre-trained on textual data can be used as the computation backbone for dealing with task in other domain as well.

GPT-2 Implementation

- Implement GPT-2 architecture, its loss function and training procedure(not necessarily train the entire model, we will load pretrained weights in latter sections). Get a feeling of how the GPT-2 model achieves its expressive power

Application of GPT-2 in a Simple Language Task

- Play with a simple language task (e.g., sentiment analysis or text classification). Learn about how to adapt and fine-tune GPT-2 for it.

Versatility of GPT-2 for Tasks in other Domains

- Introduce a completely different task (e.g. XOR, image classification) and solve using naive deep learning algorithms.
- Show how GPT-2 can be adapted for those tasks and discuss the results.

Advantages of Pre-trained Models: Speed and Accuracy

- Compare the performance of fine-tuning pretrained GPT-2 with naive models in terms of training speed and accuracy

Influence of Model Capacity on Accuracy and Training Time

- Examine how using GPT-2 of different model scales affects both accuracy and fine-tuning time
- Present a graphical or tabular comparison of results with different model capacities

Interpreting Attention Layers in GPT-2

- Explain the role of attention layers in the transformer architecture
- Illustrate how attention layers function in the context of the chosen tasks