CSE665: Large Language Models

Assignment 3

Fine Tuning Large Language Models

Aryan Dhull | 2021520

GitHub

In this task, I fine-tuned a language model using QLoRA (Quantized Low-Rank Adaptation), aiming to enhance its performance on a specific dataset. The process involved data preparation, model adaptation, quantization, and evaluation. Here's a step-by-step breakdown of my approach, with specific results where applicable.

1. Data Loading and Preprocessing

Began by importing the dataset and performing necessary preprocessing.

2. Model Initialization

Using QLoRA, initialized a pre-trained language model. QLoRA applies low-rank adaptation with quantization, making it efficient for fine-tuning tasks with limited memory resources. This setup prepares the model to incorporate new task-specific information without modifying the underlying structure significantly.

Code Insight: Initializing with QLoRA helps retain model efficiency by using low-rank matrices, which reduces the computational cost and memory load during adaptation.

3. Model Saving

After training, saved the fine-tuned model, including the quantization parameters. This setup allows for quick reloading and inference in future tasks without repeating the fine-tuning process.

Code Insight: Saving the model with quantization and adaptation layers makes it easily deployable and adaptable for new data without additional fine-tuning.

4. Inference and Model Comparison

Finally, conducted inference using both the fine-tuned model and the original base model. Assessed their outputs based on relevance and accuracy to the dataset.

Results:

- **Base Model Accuracy:** 31%
- **Fine-Tuned Model Accuracy:** 51%

The fine-tuned model showed marked improvement in producing relevant, task-specific responses, proving the effectiveness of QLoRA in achieving model adaptation with limited resource requirements.

Accuracy of Pretrained model : **31%**
Accuracy of Fine Tuned model : **51%**

Time taken to fine-tune the model using QLoRA : **3964.1522 seconds (66.07 mins)**

Total parameters in the model : **1544084480**
The number of parameters fine-tuned : **22691840 (1.47%)**

Resources used during fine-tuning : **GPU memory:5.8GB   DISK: 38.6 GB**

Failure Cases Corrected by Fine-Tuning

Example 1:
Premise: "A man is renovating a room."
Hypothesis: "A man is using a hammer in a room."
Correct Label: Neutral
Pretrained Model Prediction: Entailment
Fine-Tuned Model Prediction: Neutral (Correct)

Example 2:
Premise: "A shirt booth with a man printing a shirt."
Hypothesis: "The shirt is blue with black lettering."
Correct Label: Neutral
Pretrained Model Prediction: Entailment
Fine-Tuned Model Prediction: Neutral (Correct)

The pretrained model often predicted "entailment" when it should have predicted "neutral," as it hadn't learned the subtle difference between related but independent statements. Fine-tuning helped it recognize these distinctions—like knowing that "renovating a room" doesn't imply "using a hammer." This made the model better at correctly identifying "neutral" cases.

Failure Cases from Fine-Tuning

Example 1
Premise: "City street crowded with sports fans wearing orange."
Hypothesis: "The sports fans are wearing yellow."
Correct Label: Contradiction
Fine-Tuned Model Prediction: Neutral (Incorrect)

Example 2
Premise: "People are all standing together in front of a statue of an animal, and they are all wearing cool-weather clothing."
Hypothesis: "A beautiful statue of a man."
Correct Label: Contradiction
Fine-Tuned Model Prediction: Neutral (Incorrect)

In these cases, the fine-tuned model incorrectly predicted "neutral" instead of "contradiction." This suggests that fine-tuning may have made the model more likely to choose "neutral" when faced with statements that don't seem directly related. It may have learned to overlook contradictions that involve specific details (like color or subject differences) and instead assumed the statements were unrelated.

Fine-tuning improved the model's understanding of "neutral" examples, but it may have also caused it to miss clear contradictions in cases like these, where the statements conflict on specific details.

Consistent Failure Cases

Example 1
Premise: "Two women are observing something together."
Hypothesis: "Two women are looking at a flower together."
Correct Label: Neutral
Pretrained Model Prediction: Entailment
Fine-Tuned Model Prediction: Contradiction

Example 2
Premise: "A boy drags his sled through the snow."
Hypothesis: "A boy drags a sled up the hill."
Correct Label: Neutral
Pretrained Model Prediction: Entailment
Fine-Tuned Model Prediction: Contradiction

In these examples, both the pretrained and fine-tuned models struggled to identify "neutral." The pretrained model often assumed "entailment" when there was any overlap in action or context between the statements, while the fine-tuned model mistakenly labeled these as "contradiction."

This suggests both models had difficulty distinguishing subtle, neutral relationships where statements are related in theme but differ in specific details (e.g., "observing something" vs. "looking at a flower," or "dragging a sled" vs. "dragging it up a hill"). This could be due to limited examples in fine-tuning that highlight minor contextual differences without implying a clear relationship.

REFERENCES:

LINK 1

LINK 2

LINK 3