LoRA on Fine-tuning DistilBert for Text Classification

Jinhua Lyu

December 18, 2024

1 Base Case for Text Classification

Results

- Evaluation Loss: 0.3231
- F1 Score: 0.8506
- Training Time: 297 s

Model Description

- Pre-trained model: distilbert-base-uncased
- Dataset: IMDB benchmark dataset

Training and Evaluation Data

- Training data: 2000 samples
- Evaluation data: 500 samples

Training Hyperparameters
The following hyperparameters were used for training:

- Learning rate: 2e-5
- Training batch size: 16
- Evaluation batch size: 64
- Seed: 42
- Optimizer: AdamW
- Number of epochs: 5

LoRA Configuration

- r = 16
- lora_alpha = 16
- lora_dropout = 0.1

Training Results

Table 1: Training Results

Epoch	Training Loss	Evaluation Loss
1	0.6822	0.6613
2	0.6229	0.5571
3	0.4643	0.3820
4	0.3508	0.3313
5	0.3239	0.3231

Framework Versions

- PEFT 0.14.0
- Transformers 4.47.0
- Torch 2.5.1
- Datasets 3.2.0
- Tokenizers 0.21.0

Hardware Configuration

- CPU: Apple M1 Max, 10-core CPU
- GPU: 32-core GPU
- RAM: 32GB

2 Factors Affecting Results

Training Data Size

Increasing the size of the training dataset generally lowers the evaluation loss and improves the F1 score. However, it also substantially increases the required training time.

Table 2: Results for Different Training Data Sizes

Data size	Training time	Evaluation loss	F1 Score
2000 training data; 500 evaluation data	297 s	0.3231	0.8506
25000 training data; 5000 evaluation data	3724 s	0.2256	0.9105

Rank of LoRA

Changing the rank of LoRA does not affect the results too much on this task classification task. However, the sensitivity of LoRA rank may vary depending on the task.

Table 3: Results for Different Training Data Sizes

Rank of LoRA	Training time	Evaluation loss	F1 Score
1	199 s	0.3409	0.8542
8	308 s	0.3276	0.8512
16	297 s	0.3231	0.8506