Fine-tuning RoBERTa for Sentiment Classification on the IMDb Dataset

Abstract

This report presents the results of fine-tuning a RoBERTa model for sentiment classification on the IMDb movie review dataset. We explored four different fine-tuning approaches: Full Fine-tuning (Full FT), LoRA, QLoRA, and IA3. The performance of each method was evaluated in terms of accuracy, training time, and the number of trainable parameters.

Introduction

Sentiment classification is a fundamental task in natural language processing, with applications in various domains. This report focuses on fine-tuning RoBERTa, a powerful pre-trained language model, for sentiment classification on the IMDb dataset. We compare different parameter-efficient fine-tuning methods to assess their effectiveness and efficiency.

Methods

The following fine-tuning methods were employed:

Full Fine-tuning (Full FT): All parameters of the pre-trained RoBERTa model are updated during training. LoRA (Low-Rank Adaptation): This method injects trainable rank decomposition matrices into each layer of the transformer, significantly reducing the number of trainable parameters. QLoRA (Quantized LoRA): QLoRA combines LoRA with 4-bit quantization to further reduce memory footprint and training time. IA3 (Infused Adapter by Autoscaling, Automatic Layer Selection, and Alignment with Attention): This method automatically selects the most important layers for adaptation and aligns the adapters with the attention mechanism. Experimental Setup

Dataset: The IMDb movie review dataset was used, consisting of 3,000 training samples and 2,000 test samples. Model: RoBERTa-base was used as the pre-trained language model. Tokenizer: The RobertaTokenizerFast was used for tokenization. Training: The models were trained using the Hugging Face Trainer with a batch size of 8 and 3 epochs. Evaluation: Accuracy was used as the evaluation metric. Results

Method Accuracy Training Time (s) Trainable Parameters Full FT 0.8805 266.03 124,649,730 LoRA 0.868 116.27 3,584,002 IA3 0.854 121.35 3,584,002 The table shows that Full FT achieved the highest accuracy (0.8805) but required the longest training time (266.03 seconds) and had the largest number of trainable parameters (124,649,730). LoRA and IA3 demonstrated comparable accuracy while significantly reducing training time and the number of trainable parameters. QLoRA was excluded due to technical challenges.

The training time was significantly reduced by using parameter-efficient fine-tuning methods, such as LoRA and IA3.

The number of trainable parameters was significantly reduced by using LoRA and IA3.

Discussion

The results indicate that parameter-efficient fine-tuning methods like LoRA and IA3 can achieve comparable accuracy to Full FT while significantly reducing training time and memory requirements. This is particularly beneficial when working with limited computational resources. LoRA and IA3 offer a good balance between accuracy and efficiency for sentiment classification on the IMDb dataset.

Conclusion

This report evaluated different fine-tuning methods for sentiment classification using RoBERTa on the IMDb dataset. Full FT achieved the highest accuracy but was computationally expensive. LoRA and IA3 achieved comparable accuracy with reduced training time and memory footprint. We recommend using LoRA and IA3 for sentiment classification tasks where computational resources are limited.

Further Work

Further research could explore the impact of hyperparameter tuning on the performance of parameter-efficient fine-tuning methods. Additionally, investigating the performance of these methods on other sentiment classification datasets would be valuable.

