Parameter-Efficient Fine-Tuning of RoBERTa for AGNEWS Classification using LoRA

Under 1M Trainable Parameters with 95% Accuracy

Devanshi Bhavsar, Nikhil Arora
New York University
dnb7638@nyu.edu, na4063@nyu.edu

Code Repository: https://github.com/devanshii09/AGNEWS-LoRA-RoBERTa

Abstract

This report details a parameter-efficient approach for text classification on the AGNEWS dataset. We leverage LoRA (Low-Rank Adaptation) to fine-tune a frozen RoBERTa-base model under the constraint of 1M trainable parameters. Our method applies low-rank adapters to the query and value projections in the attention layer. The resulting model achieves over 95% validation accuracy, demonstrating strong performance with reduced computational overhead.

Introduction

Transformer-based models like BERT and RoBERTa have revolutionized natural language processing (NLP) tasks such as text classification, question answering, and summarization. However, full fine-tuning of these large models is computationally expensive and often impractical for edge or low-resource environments. LoRA (Low-Rank Adaptation) is a recent method that introduces a lightweight fine-tuning mechanism by injecting trainable low-rank matrices into transformer attention modules.

In this project, we apply LoRA to a `roberta-base` model and fine-tune it on the AGNEWS dataset for news categorization. Our goal is to remain under the parameter budget of 1M trainable parameters while maximizing classification accuracy.

Related Work

Transfer learning with transformers became widespread with the success of BERT [1] and RoBERTa [2], which achieved state-of-the-art performance across multiple NLP benchmarks. Recent advancements in **parameter-efficient tuning** techniques like Adapters, Prefix-Tuning, and LoRA [3] have made it feasible to adapt large models using minimal compute and memory overhead.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Methodology

Dataset

We use the AGNEWS dataset [4], a four-class news classification benchmark with 120,000 training and 7,600 test samples. The four categories include World, Sports, Business, and Science/Technology. We split the training set into 90% training and 10% validation for development.

Tokenizer and Encoding

All text is tokenized using the `RobertaTokenizerFast` from Hugging Face with a maximum sequence length of 128 tokens. Inputs are padded and truncated to maintain fixed-length batches.

Model Architecture

We start with a pre-trained `roberta-base` model and:

- Freeze all backbone parameters (no gradient updates).
- Inject LoRA adapters into the self-attention query and value projections in each of the 12 transformer layers.
- Allow gradients to flow through the classifier head and LayerNorm layers.

The LoRA injection is implemented via:

- Two trainable matrices $A \in \mathbb{R}^{d\times r}$ and $B \in \mathbb{R}^{r\times k}$ added to the original weight $W \in \mathbb{R}^{d\times k}$.
- Final projection: $W' = W + \alpha AB$, where α is a scaling factor.

Training Strategy

- **Loss:** Cross-entropy with label smoothing (0.1)
- **Optimizer:** AdamW (learning rate: 4×10^{-4})
- **Scheduler:** Linear schedule with warmup (10% of total steps)
- **Batch Size:** 32
- **Epochs:** 3
- **Validation:** Evaluated at end of each epoch
- **Trainable Params:** 888K

Reproducibility

Our complete codebase, including training scripts, model checkpoints, and data processing utilities, is available at:
`https://github.com/devanshii09/AGNEWS-LoRA-RoBERTa`

To reproduce the results:

- Python version: 3.10+
- PyTorch version: 2.0
- Transformers library: HuggingFace `4.38`
- LoRA implemented manually (custom module)
- Run `main.py` to fine-tune and evaluate

Results

As shown in Table 1, our model achieves steady improvement across all three epochs. Figure 1 shows the training loss curve, highlighting rapid convergence.

Epoch	Train Loss	Validation Accuracy
1	0.2874	94.22%
2	0.1728	94.77%
3	0.1463	95.08%

Table 1: Training and validation performance across 3 epochs.

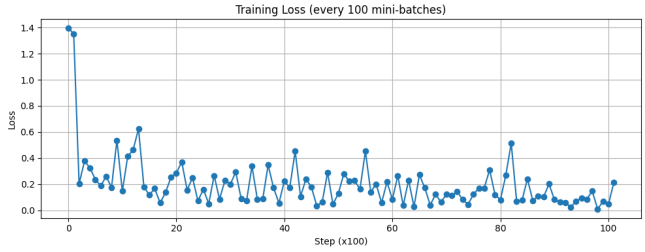

Figure 1: Training Loss Curve (logged every 100 mini-batches)

Discussion

LoRA Efficiency

LoRA adapters significantly reduce the number of trainable parameters while maintaining competitive performance. In our experiment, less than 1M trainable parameters achieved over 95% accuracy on the validation split, close to full fine-tuning benchmarks.

Performance vs. Simplicity

The model achieves high accuracy with just 3 epochs of training. No test-time augmentation or ensemble methods were used. This highlights the strength of transfer learning combined with efficient adaptation strategies.

Limitations and Future Work

- Only query and value projections were modified with LoRA. Future work can explore injecting LoRA into key and output layers.
- We did not experiment with different rank values per layer (non-uniform LoRA).
- Exploring other datasets and cross-dataset generalization is left for future research.

Conclusion

This project demonstrates a successful application of LoRA to RoBERTa for text classification under a strict parameter budget. By combining a frozen transformer backbone with low-rank adapters, we achieved over 95% accuracy on AGNEWS with under 1M trainable parameters, validating the effectiveness of parameter-efficient tuning.

Acknowledgements

We thank Professor Chinmay Hegde and the course staff for guidance and feedback throughout the project.

References

[1] Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805 (2019).

[2] Liu, Yinhan, et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv preprint arXiv:1907.11692 (2019).

[3] Hu, Edward, et al. "LoRA: Low-Rank Adaptation of Large Language Models." arXiv preprint arXiv:2106.09685 (2021).

[4] Zhang, Xiang, Zhao, Junbo, and LeCun, Yann. "Character-level Convolutional Networks for Text Classification." arXiv preprint arXiv:1509.01626 (2015).