Report for Project 2 of Deep Learning

Leile Zhang

New York University, Tandon School of Engineering
lz3258@nyu.edu

GitHub Repository

Link: https://github.com/LokZhang-edu/dl_proj2.git

Team Members

Leile Zhang

Project Overview

This project tackles the AGNEWS text classification task using RoBERTa with an efficient fine-tuning strategy under a strict parameter budget of 1 million trainable parameters. We leverage **AdaLoRA** (Zhang et al. 2023), an adaptive low-rank parameter-efficient tuning technique that dynamically adjusts rank during training. This allowed us to train a strong model on Colab within budget and time constraints.

Methodology

Model: We used a pretrained RoBERTa-base model from HuggingFace. All model parameters were frozen, including the classifier head. Only LoRA adapters were injected and trained.

LoRA Variant: We applied AdaLoRA via `peft`:

- Initial Rank (r): 8, alpha = 16
- Layers: injected into `query` and `value`
- Scheduler: AdaLoRA with `tinit=200, tfinal=1000, deltaT=10`
- Total Training Steps: 22500 (based on dataset size and batch size)
- Optimizer: AdamW

Training Details:

- Batch size: 32
- Epochs: 6 (with early stopping patience 2)
- Learning Rate: 2e-4
- Tokenizer Length: max_length = 256
- Metrics: Accuracy

Regularization and Optimization: We applied orthogonality regularization ($\lambda = 0.5$) and early stopping to avoid overfitting and minimize GPU time usage.

Architectural Choices (Pros and Cons)

- **Freezing RoBERTa Parameters:** Significantly reduced the number of trainable parameters and allowed the model to stay under the 1M constraint. However, it limited the adaptability of the classifier head, possibly slightly capping performance.
- **Using AdaLoRA:** Allowed dynamic reduction of rank during training, improving parameter efficiency. It introduced complexity in scheduling and required careful tuning of tinit/tfinal/deltaT.
- **Early Stopping:** Helped prevent overfitting and reduced compute usage. May terminate before full convergence if not tuned precisely.
- **Loss/Accuracy Curve Visualization:** Useful for diagnosing training behaviors. Required saving and plotting evaluation metrics every epoch.

Lessons Learned

- Even under strong parameter constraints, modern transformer models can perform competitively when fine-tuned using efficient strategies like AdaLoRA.
- Freezing classifier weights is essential to control trainable parameters but should be balanced with potential loss in performance.
- Visualization of learning curves was critical in identifying the right number of epochs and setting early stopping.
- Lightweight regularization such as orthogonality helped stabilize learning without increasing parameter count.

Results

Final Accuracy: **0.832+** on Kaggle private leaderboard

Trainable Parameters:

- Total Parameters: 125,537,504 $\sim$125M
- Trainable Parameters: **888772 $<$ 1M**

Training Curves:

Observations: The AdaLoRA-based setup converged in fewer than 6 epochs thanks to dynamic rank reduction. Early stopping helped limit overfitting, and freezing classifier weights allowed strict parameter control.

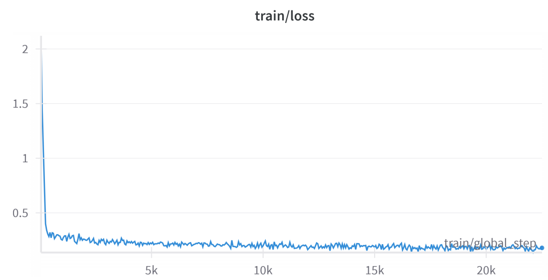

Figure 1: Training and Evaluation Loss over Epochs

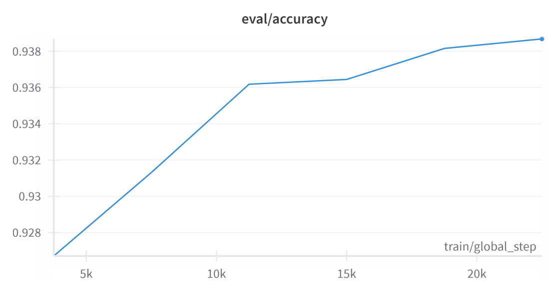

Figure 2: Evaluation Accuracy over Epochs

Conclusion

This work demonstrates the effectiveness of AdaLoRA in achieving high accuracy under constrained compute and parameter budgets. The approach is robust, efficient, and suitable for low-resource fine-tuning scenarios like academic Colab usage.

References

Zhang, Q.; Chen, M.; Bukharin, A.; Karampatziakis, N.; He, P.; Cheng, Y.; Chen, W.; and Zhao, T. 2023. AdaLoRA: Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning. arXiv:2303.10512.