Sentiment Analysis with DistilBERT

Aliman Alibek, Dauletbek Yergali, Karabaliyev Yerlan

December 12, 2024

Abstract

This document describes a project aimed at performing sentiment analysis on movie reviews using DistilBERT, a lightweight transformer model. Techniques such as gradient checkpointing and mixed precision training were used to optimize performance and efficiency. The Transformer architecture was introduced in [Vaswani et al., 2017]. DistilBERT, a lighter version of BERT, was presented in [Sanh et al., 2019]. You can find the project code at: `https://github.com/AlibekWarBoss/sen-an-w-dbert`.

1 Introduction

Sentiment analysis is an essential task in Natural Language Processing (NLP) with applications in diverse fields such as market research, product analysis, and customer feedback evaluation. This project focuses on classifying movie reviews as positive or negative using the IMDb dataset and a transformer-based architecture, DistilBERT. The project emphasizes model efficiency and performance through advanced training strategies such as gradient checkpointing and mixed precision training .

1.1 Team

Aliman Alibek, Dauletbek Yergali, Karabaliyev Yerlan: Implemented the entire project, including data preprocessing, model training, and evaluation.

2 Related Work

Traditional approaches to sentiment analysis, such as Naive Bayes and Support Vector Machines (SVMs), rely heavily on handcrafted features. Modern transformer-based architectures, like BERT and its derivatives, have outperformed these methods. DistilBERT offers a computationally efficient alternative with comparable accuracy. This project builds on the advantages of DistilBERT and integrates optimization techniques for better efficiency.

3 Model Description

The model used in this project is DistilBERT, fine-tuned for sentiment classification with two output classes (positive and negative).

Key features of the model:

- **Gradient Checkpointing**: Reduces memory usage by recomputing intermediate layers.
- **Mixed Precision Training**: Speeds up computation by using 16-bit floating-point operations.
- **AdamW Optimizer**: Helps prevent overfitting by regularizing the learning process.

4 Dataset

The IMDb dataset [Maas et al., 2011] was used for training and evaluation. It consists of 50,000 reviews, equally split between positive and negative sentiment. The dataset was preprocessed using the DistilBERT tokenizer to handle tokenization and padding.

The statistics for the dataset are shown in Table 1.

	Train	Valid	Test
Reviews	25,000	N/A	25,000
Classes	2 (Positive, Negative)		

Table 1: Statistics of the IMDb dataset.

5 Experiments

5.1 Metrics

The primary evaluation metric was accuracy, defined as:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Additionally, the F1-score was used to measure the balance between precision and recall.

5.2 Experiment Setup

The model was trained on the following configuration:

- **Batch Size**: 8

- **Learning Rate**: $2e-5$
- **Epochs**: 3
- **Hardware**: NVIDIA RTX 3050 Mobile (4 GB)

5.3 Baselines

A baseline logistic regression model was trained on TF-IDF features for comparison.

6 Results

The results of the experiments are summarized in Table 2.

Model	Accuracy	F1-Score
Logistic Regression (Baseline)	85.2%	84.5%
DistilBERT (No Augmentation)	88.5%	88.1%
DistilBERT (With Augmentation)	89.8%	89.5%

Table 2: Performance comparison of models on the IMDb dataset.

7 Conclusion

This project demonstrated the effectiveness of DistilBERT for sentiment analysis tasks. By incorporating data augmentation, gradient checkpointing, and mixed precision training, the model achieved significant improvements in accuracy and efficiency. Future work will explore hyperparameter tuning and domain-specific pretraining for further enhancements.

References

[Maas et al., 2011] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning word vectors for sentiment analysis. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150.

[Sanh et al., 2019] Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

[Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.