Progress Report: Improving Transformer-Based Text Classification Using Contrastive Loss

Emre Kuru
emre.kuru@ozu.edu.tr

Merve Keleş Doğan
mervekeles@ozyegin.edu.tr

Berkecan Koçyiğit
berkecan.kocyigit@ozu.edu.tr

May 7, 2025

1 Completed Work

To extend the scope of our analysis, we decided to add another dimension to our project and evaluate the proposed contrastive loss against the baseline cross-entropy loss across different BERT-based transformers. On this end, we've completed the following:

1.1 Reusable Training Workflow

We built a unified script that handles data loading, tokenization, checkpointing, and reproducibility.

1.2 Evaluation Framework

An evaluation framework has been established to track classification performance (accuracy, precision, recall, f1-score) across datasets for all evaluated models.

1.3 Initial Experiments

Baseline experiments were run on one of our datasets (financial-phrasebank) using the standard cross-entropy loss on various BERT-based models (DistilBERT, BERT-base, LoRA BERT, ModernBERT). This was done to assess the correctness of the general framework. The results are summarized in Figure 1.

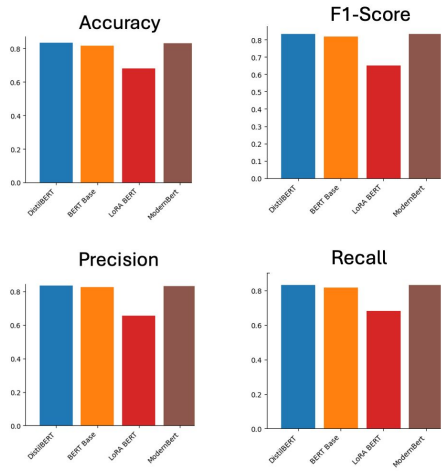

Figure 1: Baseline Experiment Results on financial-phrasebank dataset across all evaluated models.

2 Remaining Tasks

2.1 Contrastive Loss Integration

The existing training workflow will be extended to support a supervised contrastive loss. To accomplish this goal, we will implement a dataset class specifically designed to facilitate in-batch sampling of pos-

itives (same label) and negatives (different labels). This is a crucial aspect of contrastive learning based training workflows.

2.2 Full-Scale Experiments

We plan to execute the planned grid of runs across all three datasets (Financial PhraseBank, Emotion, AG News subset) and models (DistilBERT, BERT-base, LoRA BERT, ModernBERT). Afterwards, we will compare the proposed method's classification performance with the current baseline of cross-entropy loss across all settings.

2.3 Embedding Space Analysis

For all datasets, 2D projections with t-SNE and UMAP will be generated to visually compare cluster separability and observe the expected effect of contrastive learning. Furthermore, silhouette scores will be computed for each embedding projection to quantify improvements in class clustering when applying contrastive loss.