

Finetuning minBERT Model for Multiple Downstream Tasks

Yuan Wang | Winter 2023 CS224N Default Final Project | Stanford University

Problem

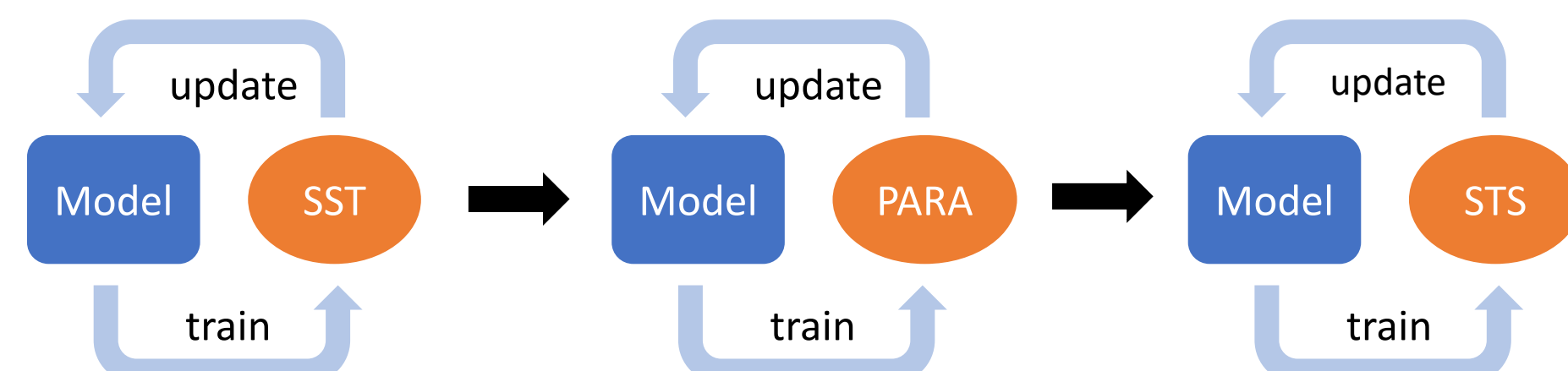
- Pre-trained Large Language Models (LLMs) contain rich token embeddings that are useful for various downstream tasks. Instead of building a separate model for each individual task, it is more resource-efficient to build one model that could perform multiple tasks.
- We attempt to finetune and extend a minBERT model to perform multiple downstream tasks.

Background

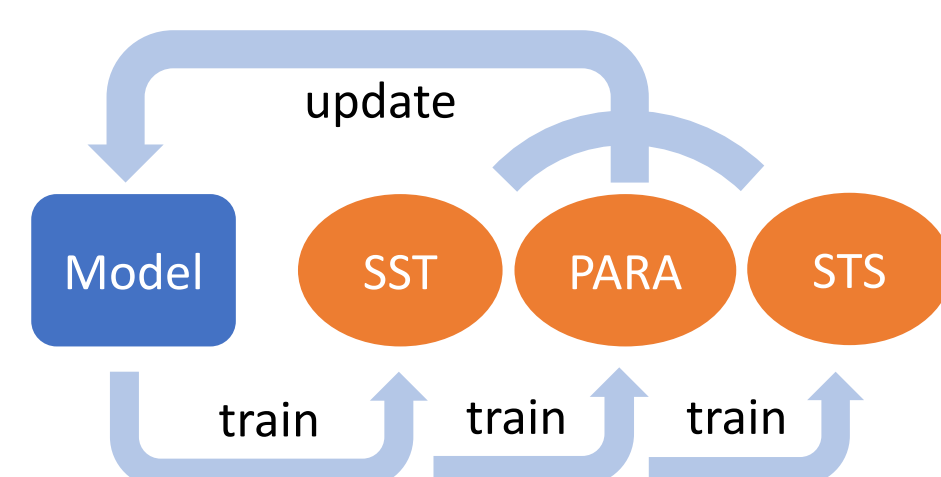
- With a pre-trained minBERT model, we would like to develop a model to perform three downstream tasks: Sentiment Analysis, Paraphrase Detection, and Semantic Textual Similarity (STS) analysis.
- We are provided three datasets to use for training: the Stanford Sentiment Treebank (SST) [8.5k], the Quora dataset [141.5k], and the SemEval STS dataset [6.0k].

Methods

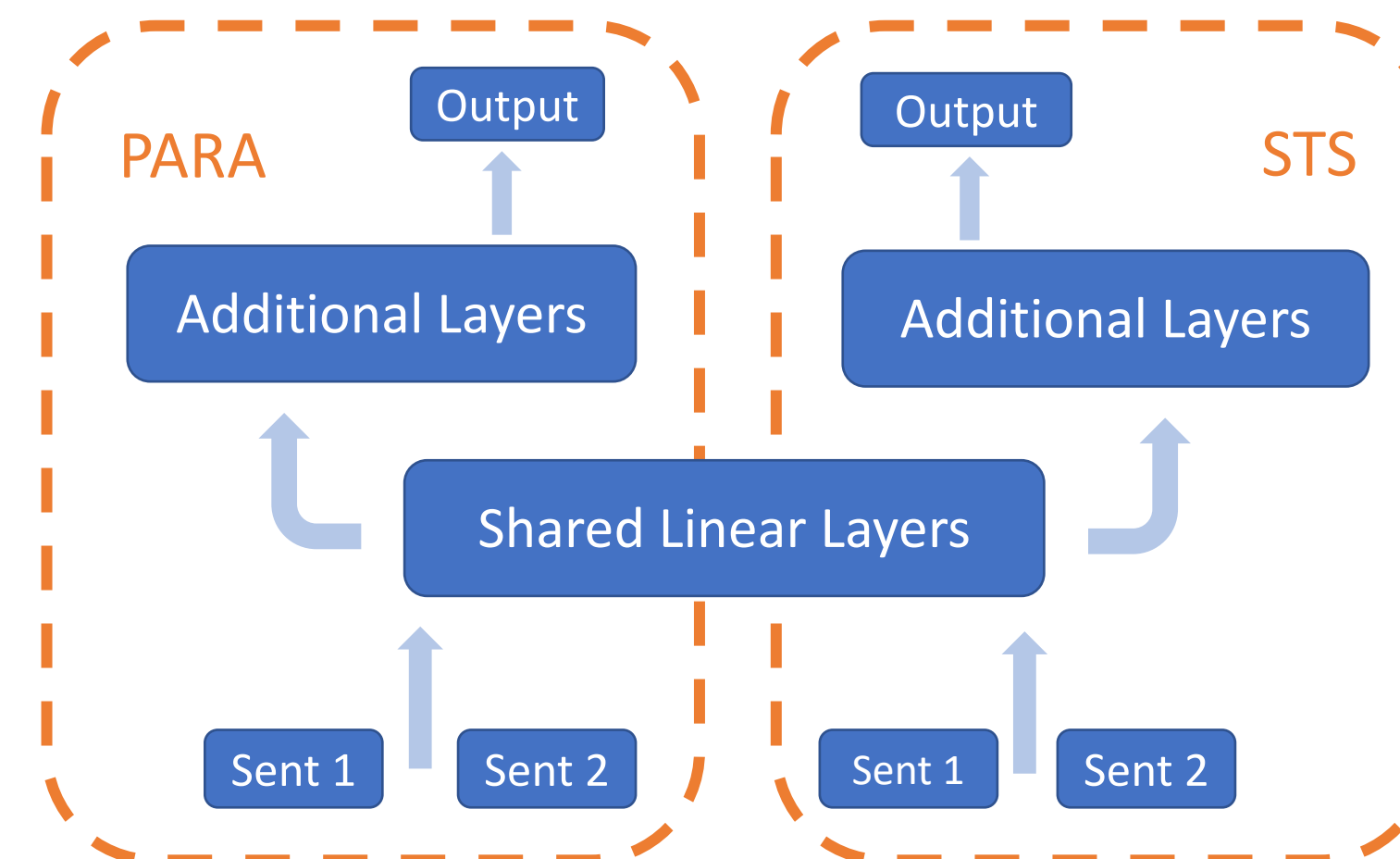
- Baseline: task-by-task sequential training



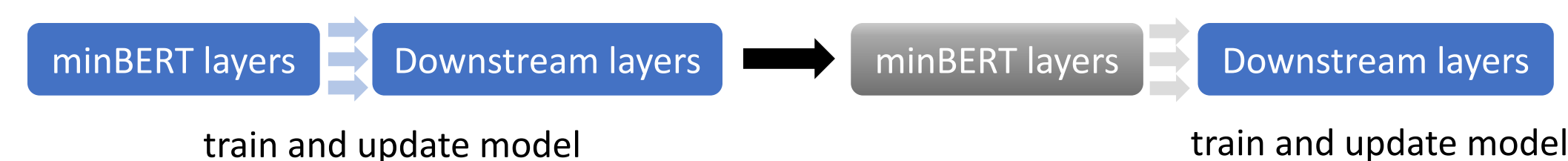
- Round robin multitask training



- Layer sharing

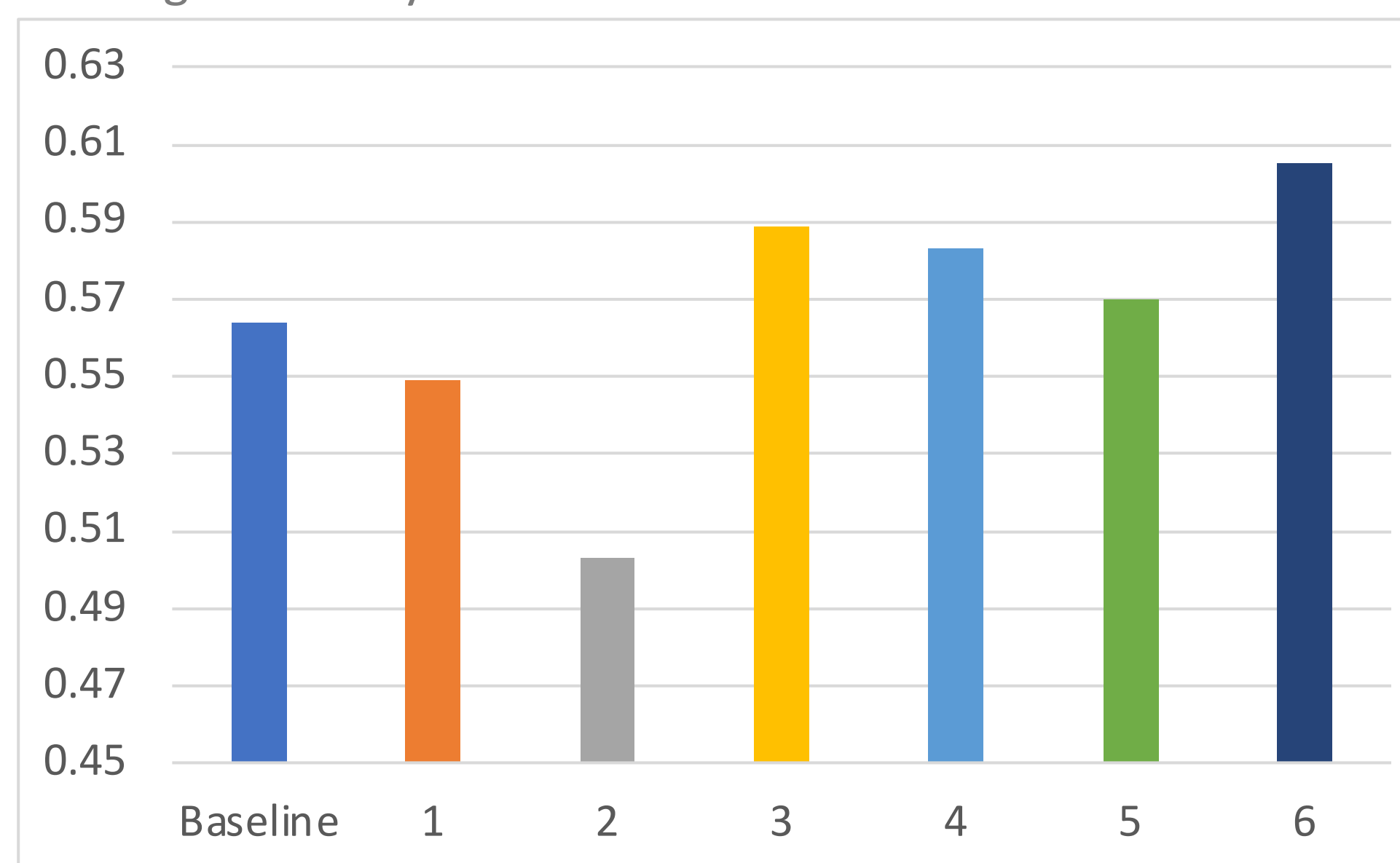


- Additional training with minBERT parameters fixed



Experiments

Average Accuracy

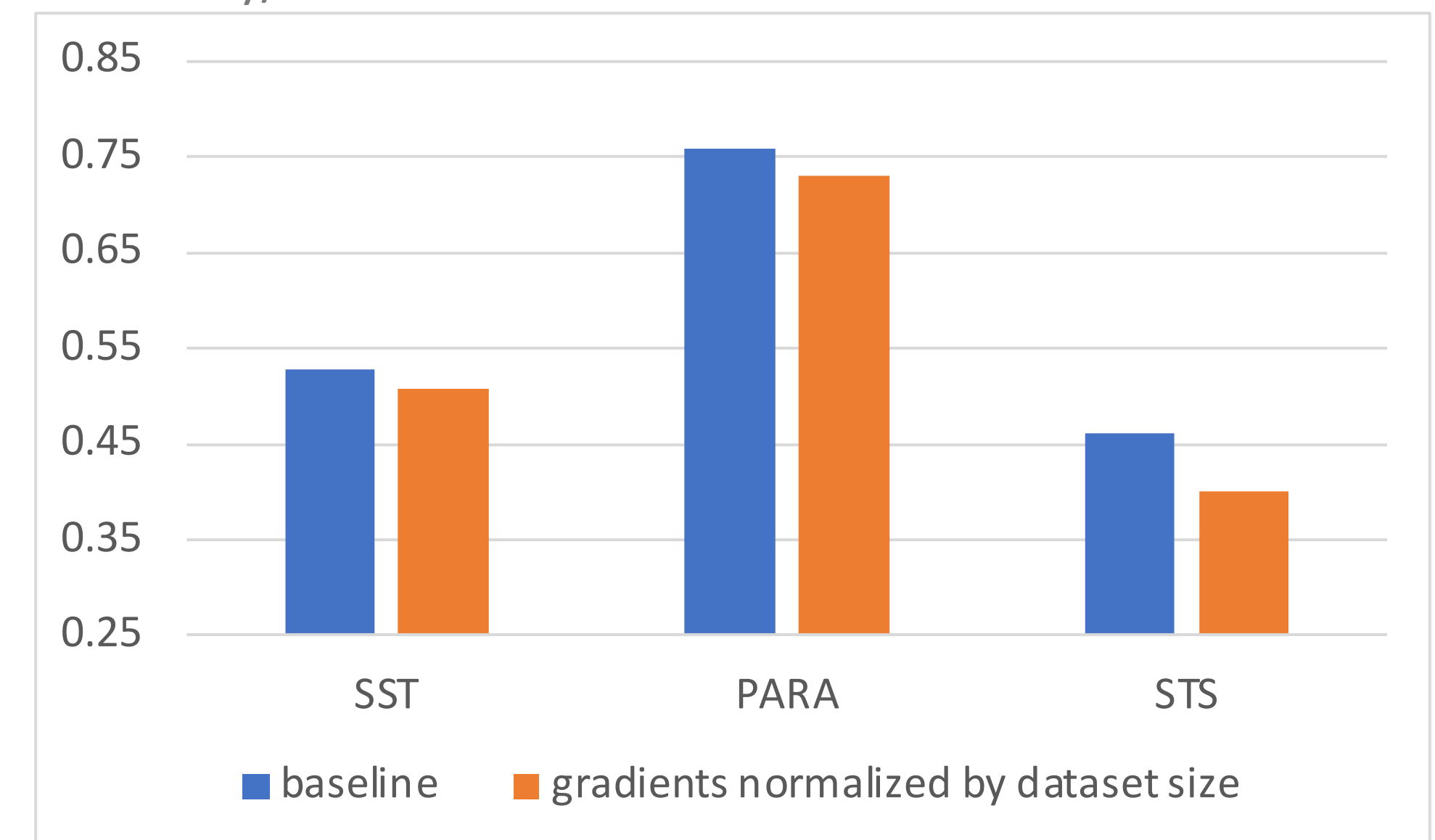


- Baseline: sequential training
- 1: Share layers between PARA and STS
- 2: Cut down on PARA training
- 3: Additional SST and STS training data
- 4: Round robin multitask training
- 5: Additional training with minBERT parameters fixed
- 6: 3, 4, 5, and additional training epochs

Analysis

- Round robin multitask training: focused training on the (high-volume) Quora dataset benefits other tasks

Accuracy/Correlation



Conclusions

- Adapted additional training data helps improve performance
- Round robin multitask training leads to better overall results than sequential training
- Additional training with minBERT parameters fixed further improves performance on each individual task
- Training on a large volume of Quora data (paraphrase detection) tends to improve performance on SST and STS

Reference

Qiwei Bi, Jian Li, Lifeng Shang, Xin Jiang, Qun Liu, and Hanfang Yang. Mtrec: Multi-task learning over bert for news recommendation. In Findings of the Association for Computational Linguistics: ACL 2022, pages 2663–2669, 2022.