CausalMixFT Accelerates Convergence and Accuracy in TabLM-2B on TabBench

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: How does CausalMixFT impact the convergence rate and final accuracy of TabLM-2B compared to standard mixup augmentation on the TabBench suite. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Survey on Mixup Augmentations and Beyond. Research question: How does CausalMixFT impact the convergence rate and final accuracy of TabLM-2B compared to standard mixup augmentation on the TabBench suite?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

11 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 3.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
MixUp is gradually replaced by basic data augmentation to shift from exploration to exploitation.	×	0.04
RegMixup found that different α values have an effect on the entropy resulting from the model.	×	0.02
RegMixup found that the MCE loss brings complementary effects when it is used in combination with the one-hot CE loss.	×	0.02
AdaMixup proposed a method to learn λ and introduced two neural networks to generate the mixing ratio λ and to judge whe	×	0.04
MetaMixup used Meta-Learning to optimize the mixup with a bi-fold optimization: the model fθpq and the mixing ratio λ.	×	0.03
MixPUL rewards unsupervised consistency between unlabeled samples by alleviating the supervised problem by mining reliab	×	0.01
P3Mix performs more accurate supervision by mixing unlabeled samples and positive samples from the decision boundary tha	×	0.02
MixRL proposed using the validation set to learn, for each sample, 'how many nearest neighbors should be mixed to obtain	×	0.04
C-Mixup adjusts the sampling probability based on the similarity of the labels and then selects pairs of samples with si	×	0.06
C-Mixup uses a symmetric Gaussian kernel to calculate the sampling probability of another sample, where the closer sampl	×	0.02
Deep Neural Networks (DNNs) have powerful feature representation ability and have been successfully applied to a variety	×	0.06
DNNs employ a large number of learnable parameters, and without numerous training data, models could easily overfit and	×	0.03
Data Augmentation (DA) techniques prevent over-fitting by being a 'data-centric' regularization technique.	×	0.05

References

- http://arxiv.org/abs/2409.05202v2
- http://arxiv.org/abs/2107.12246v2
- http://arxiv.org/abs/2312.02405v1