Scaling Synthetic Datasets in CausalMixFT for Tabular Foundation Model Performance on TabMWP

Assignee Research

June 12, 2026

Abstract

Fine-tuning tabular foundation models (TFMs) under data scarcity is challenging, as early stopping on even scarcer validation data often fails to capture true generalization performance. We propose CausalMixFT, a method that enhances fine-tuning robustness and downstream performance by generating structurally consistent synthetic samples using Structural Causal Models (SCMs) fitted on the target dataset. This approach augments limited real data with causally informed synthetic examples, preserving feature dependencies while expanding training diversity. Evaluated across 33 classification datas

1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: What is the impact of scaling the size of the synthetic dataset generated by CausalMixFT on the fine-tuning performance of tabular foundation models, measured by validation accuracy and downstream task generalization on TabMWP?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

7 papers retrieved. 15 claims extracted; 11 independently verified. Quality review score: 7.5/10.

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
Experiments were conducted on the Mitra model across 33 classification datasets from the TabArena benchmark suite.	✓	0.18
The experimental setup involved 10 folds for each of the 33 datasets, totaling 2,310 fine-tuning runs.	×	0.12
Model performance was reported as normalized ROC-AUC relative to the pre-trained model.	✓	0.27
CausalMixFT achieved a median improvement of +0.12$\pm$0.63 over the pre-trained model.	✓	0.18
The default fine-tuning baseline achieved a median improvement of +0.10$\pm$0.98 over the pre-trained model.	✓	0.22
Purely synthetic augmentation methods (CTGAN, SCM, TabEBM, TableAugment, and MixedModel) showed negative median improvem	✓	0.21
CausalMixFT demonstrated lower performance variability ($\pm$0.63) compared to the default fine-tuning baseline ($\pm$0.98).	✓	0.15
In average rank analysis across datasets, CausalMixFT ranked first overall.	×	0.08
The normalization strategy used sets the base model's zero-shot performance as the baseline to measure fine-tuning impro	×	0.07
The normalization formula is defined as: score_normalized = metricsign $\times$ (score_method / score_baseline - 1) $\times$ 100%.	×	0.00
CausalMixFT generates synthetic data using Structural Causal Models (SCMs) fitted to the target dataset.	✓	0.20
SCMs encode causal dependencies among features through a directed acyclic graph (DAG) and structural equations.	✓	0.21
Structural relations between features are estimated using the PC and FCI algorithms.	✓	0.16
DAGs are sampled and fitted using DoWhy's SCM framework with additive noise models.	✓	0.24
In the SCM framework, numerical features are modeled with regressors and categorical features with classifiers.	✓	0.20

References

- http://arxiv.org/abs/2601.04110v2
- http://arxiv.org/abs/2407.21523v1
- http://arxiv.org/abs/2507.05904v1