

CausalMixFT Robustness in Tabular Foundation Models Under Distribution Shifts on TabMWP and TABFACT

Assignee Research

June 11, 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 datasets

1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: What is the effect of CausalMixFT on the robustness of tabular foundation models against distribution shifts when evaluated on out-of-domain subsets of TabMWP and TABFACT?.

2 Methodology

Systematic literature search across multiple databases yielded 13 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

13 papers retrieved. 14 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
CausalMixFT achieves the highest median improvement of $(+0.12 \pm 0.63)$ over the pre-trained model on 33 classification data	✓	0.27
CausalMixFT outperforms the default fine-tuning baseline $(+0.10 \pm 0.98)$ and all purely synthetic augmentation methods, in	✓	0.30
Default fine-tuning has a variability of ± 0.98 , while CausalMixFT has a variability of ± 0.63 , indicating greater stability	×	0.13
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline, with pur	✓	0.25
The normalization strategy used to compare performance across different data generators is based on Gorishniy et al. [12]	×	0.06
The normalized performance is computed as $\text{score}_{\text{normalized}} = \text{metric}_{\text{sign}} \times (\text{score}_{\text{method}} / \text{score}_{\text{baseline}} - 1) \times 100\%$, where	×	0.08
CausalMixFT extends the fine-tuning framework of Bhlér et al. [5] by mixing real and causally grounded synthetic sample	✓	0.29
SCM-Based Synthetic Augmentation (CausalMixFT) uses Structural Causal Models (SCMs) fitted to the target dataset to gene	✓	0.21
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	✓	0.27
The structural relations between features are estimated using the PC and FCI algorithms, producing a probabilistic adjac	✓	0.25
DAGs are sampled and fitted using DoWhy’s SCM framework with additive noise models.	✓	0.24
Numerical features are modeled with regressors, and categorical features with classifiers in the SCM framework.	✓	0.20
The complexity of the internally used model types in SCM can be controlled through a quality hyperparameter.	✓	0.17
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	✓	0.22

References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2307.05284v6>
- <http://arxiv.org/abs/2512.03307v1>