What is the trade-off in inference latency between ciDATGAN and non-conditional tabular GANs when generating l

Assignee Research

June 10, 2026

Abstract

Conditionality has become a core component for Generative Adversarial Networks (GANs) for generating synthetic images. GANs are usually using latent conditionality to control the generation process. However, tabular data only contains manifest variables. Thus, latent conditionality either restricts the generated data or does not produce sufficiently good results. Therefore, we propose a new methodology to include conditionality in tabular GANs inspired by image completion methods. This article presents ciDATGAN, an evolution of the Directed Acyclic Tabular GAN (DATGAN) that has already been sh

1 Introduction

This paper examines: ciDATGAN: Conditional Inputs for Tabular GANs. Research question: What is the trade-off in inference latency between ciDATGAN and non-conditional tabular GANs when generating large-scale synthetic datasets with high-cardinality features, and how does this impact scalability?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

3 Results

10 papers retrieved. 9 claims extracted; 1 independently verified. Quality review score: 4.7/10.

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
ciDATGAN was evaluated using a modified version of the LPMC dataset, which consists of 18 variables and 16,904 trips.	×	0.03
The conditional inputs used in ciDATGAN for the LPMC dataset are the variables age, female, and hh_region.	×	0.10
ciDATGAN and DATGAN were trained on the LPMC dataset five times, and five different synthetic datasets were generated fo	×	0.07
The performance of ciDATGAN was assessed using frequency lists for each variable and each combination of two variables,	×	0.06
A LightGBM model was trained on all but one column to predict the corresponding column in the original data, and results	×	0.06
The SRMSE metric was used to evaluate the performance of ciDATGAN, with values reported as 0.103 and 0.095 in one table,	×	0.03
ciDATGAN aims to generate synthetic tabular data based on conditional inputs, inspired by image completion.	✓	0.19
ciDATGAN can be used for removing bias in a dataset using unbiased variables as conditional inputs or combining informat	×	0.14
ciDATGAN can generate detailed synthetic datasets covering a whole population by learning how to generate detailed varia	×	0.05

References

- http://arxiv.org/abs/2501.17324v1

- http://arxiv.org/abs/2104.11797v1
- http://arxiv.org/abs/2210.02404v1