

TTVAE: Transformer-Based Generative Modeling for Tabular Data Generation (Abstract Reprint)

Alex X. Wang¹, Binh P. Nguyen^{1,2}

¹School of Mathematics and Statistics, Victoria University of Wellington, New Zealand

²Faculty of Information Technology, Ho Chi Minh City Open University, Viet Nam

Abstract Reprint. This is an abstract reprint of the journal article by Wang and Nguyen (2025).

Abstract

Tabular data synthesis presents unique challenges, with Transformer models remaining underexplored despite the applications of Variational Autoencoders and Generative Adversarial Networks. To address this gap, we propose the Transformer-based Tabular Variational AutoEncoder (TTVAE), leveraging the attention mechanism for capturing complex data distributions. The inclusion of the attention mechanism enables our model to understand complex relationships among heterogeneous features, a task often difficult for traditional methods. TTVAE facilitates the integration of interpolation within the latent space during the data generation process. Specifically, TTVAE is trained once, establishing a low-dimensional representation of real data, and then various latent interpolation methods can efficiently generate synthetic latent points. Through extensive experiments on diverse datasets, TTVAE consistently achieves state-of-the-art performance, highlighting its adaptability across different feature types and data sizes. This innovative approach, empowered by the attention mechanism and the integration of interpolation, addresses the complex challenges of tabular data synthesis, establishing TTVAE as a powerful solution.

References

Wang, A. X.; and Nguyen, B. P. 2025. TTVAE: Transformer-based generative modeling for tabular data generation. *Artificial Intelligence*, 340: 104292.