Generative Models for Time-Series and Fair Data Generation

Md Fahim Sikder[1], Resmi Ramachandranpillai[2], Daniel de Leng[1], and Fredrik Heintz[1]

[1]Reasoning and Learning Lab (ReaL), Department of Computer and Information Science (IDA), Linköping University, Sweden
[2]Institute for Experiential AI, Northeastern University, USA

Abstract

Machines and software powered by artificial intelligence are rapidly growing and making our lives much easier. This is possible due to the availability of a vast amount of data. Unfortunately, most of the time, these data are not processed, and it takes a lot of resources to make it usable. On top of that, most of the processed datasets that are available are kept private due to various limitations. One possible solution to this problem is generating synthetic data statistically like the original using generative models. Computer vision and natural language processing have seen advancement in generative models in recent years. However, little work has been done in the time-series generation domain due to its volatile nature. Furthermore, most of the data in real life is full of technical/human biases. Applications powered using these biased data might give unfair outcomes towards various demographics. So, generative models should also generate fair synthetic data. From these motivations, in this PhD thesis, we aim to find ways to create generative models that can generate Time-Series and Fair synthetic data.

Introduction

Research Challenges

- RQ-1: How can we capture the temporal dynamics of time-series using generative models?
- RQ-2: How to generate bias-free data using generative models?
- RQ-3: How can generative models generate fair-time-series data?

Our Contributions

- We introduce **TransFusion** [1], a Transformer and Diffusion-based generative model that can generate long-sequenced high-fidelity time series data. We also introduce two evaluation metrics called **Long Discriminative Score (LDS)** and **Long-Sequenced Predictive Score (LPS)** to evaluate the quality of the synthetic data as well as its predictive characteristics [RQ-1].
- We present a novel formulation **(FLDGMs)** [2] of a fair latent generative framework common for both GANs and Diffusion models, which introduces the concept of syntax-agnostic, model-agnostic fair latent vectors [RQ-2].
- We introduce Bt-GAN [3], GAN-based fair synthetic data generation framework, specifically designed for healthcare domain. We use score-based weighted sampling techniques to capture the sub-group representations also, we present problem definition of how generative models are affected by various bias [RQ-2].

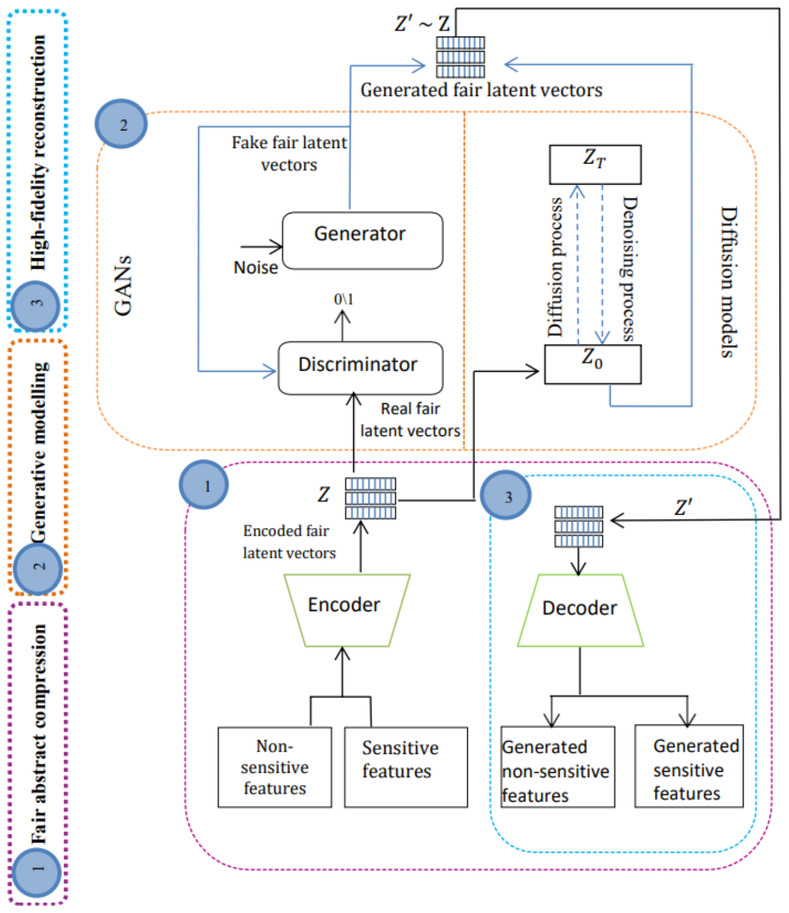

Fig 1: Architecture of FLDGMs [2]

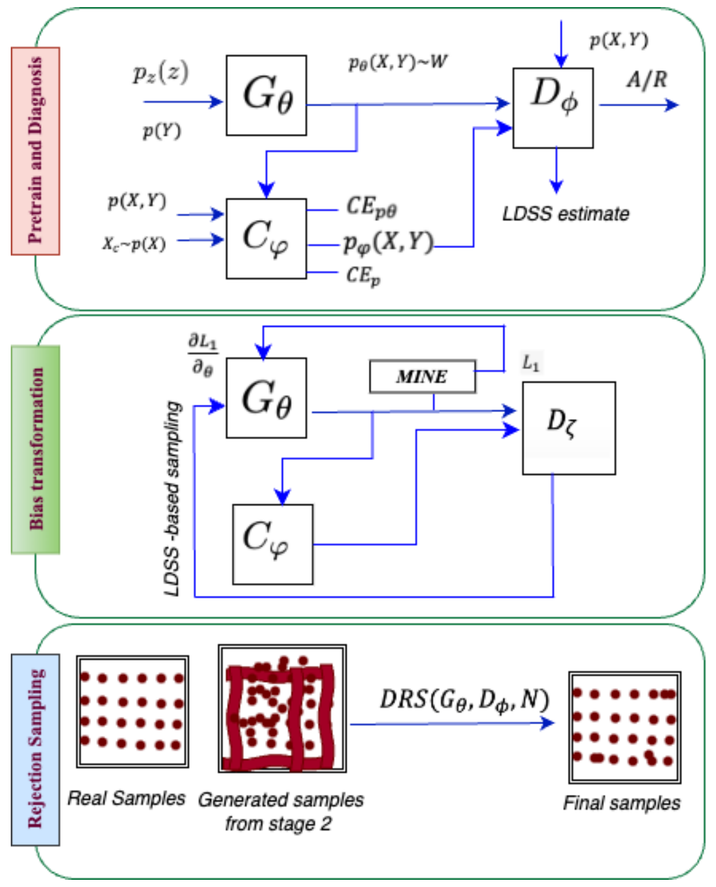

Fig 3: Architecture of Bt-GAN [3]

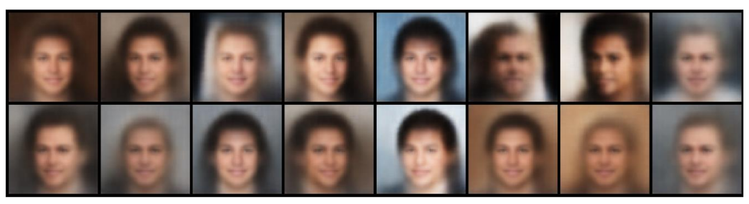

Fig 4.a: 'Female' as the sensitive sub-group (Generated by FLDGMs [2])

Fig 4.b: 'Male' as the sensitive sub-group (Generated by FLDGMs [2])

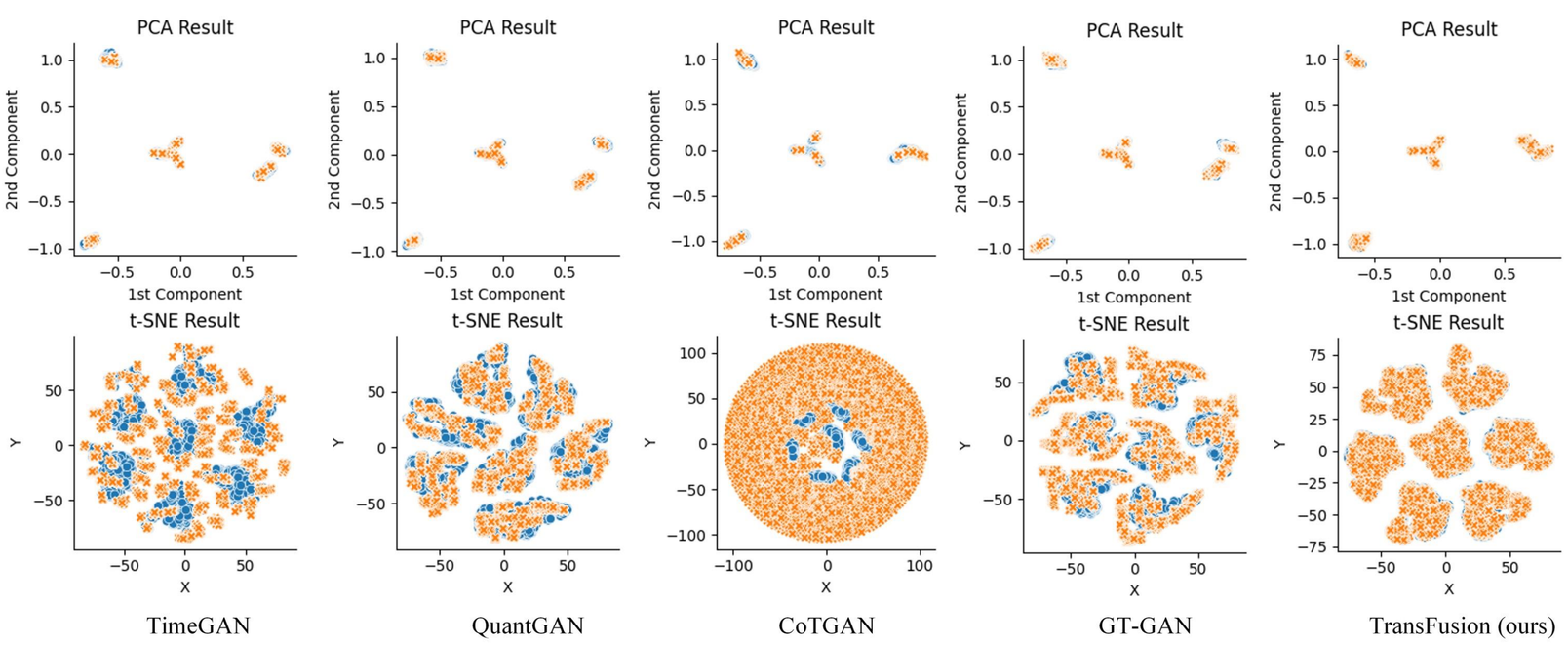

Fig 2: PCA and t-SNE plots of the real data (electricity consumption dataset, blue dots) and synthetic data (orange dots) generated by state-of-the-arts generative models and TransFusion [1], each dots represents a sequence of time-series, if the generative models learns the original data distribution, orange and blue dots should overlap, sequence length: 100

Acknowledgment

This work was funded by the Knut and Alice Wallenberg Foundation, the ELLIIT Excellence Center at Linköping-Lund for Information Technology (portions of this work were carried out using the AIOps/Stellar), and TAILOR - an EU project with the aim to provide the scientific foundations for Trustworthy AI in Europe. The computations were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre.

References

[1] Sikder, Md Fahim, Resmi Ramachandranpillai, and Fredrik Heintz. "Transfusion: Generating long, high fidelity time series using diffusion models with transformers." arXiv preprint arXiv:2307.12667 (2023). [Under Review]

[2] Ramachandranpillai, Resmi, Md Fahim Sikder, and Fredrik Heintz. "Fair Latent Deep Generative Models (FLDGMs) for Syntax-Agnostic and Fair Synthetic Data Generation." ECAI 2023. IOS Press, 2023. 1938-1945.

[3] Ramachandranpillai, Resmi, et al. "Bt-GAN: Generating Fair Synthetic Healthdata via Bias-transforming Generative Adversarial Networks." Authorea Preprints (2023). Accepted at JAIR.

Poster TransFusion FLDGMs Bt-GAN

LINKÖPING UNIVERSITY

Department of Computer and Information Science (IDA)
Artificial Intelligence and Integrated Computer Systems (AIICS)