ఇండియన్ ఇన్స్టిట్యూట్ ఆఫ్ ఇన్ఫర్మేషన్ టెక్నాలజీ,
డిజైన్ అండ్ మాన్యుఫ్యాక్చరింగ్, కర్నూలు

भारतीय सूचना प्रौद्योगिकी, अभिकल्पना
एवं विनिर्माण संस्थान, कर्नूल

Generative Adversarial Networks in Time Series: A Systematic Literature Review

Authors:

- Eoin Brophy, Dublin City University
- Zhengwei Wang, Trinity College Dublin
- Qi She, ByteDance AI Lab
- Tomás Ward, Dublin City University

Summary

Aditya Shinde

120AD0018 (AI&DS)

1. Introduction:

- This paper focuses on the application of Generative Adversarial Networks (GANs) to the time series domain, in contrast to their predominant use in computer vision.

2. GAN Architecture and Training:

- GANs follow a minimax objective, where the generator aims to maximize the failure rate of the discriminator while the discriminator aims to minimize it.

- historical development of GANs- the use of multi-layer perceptrons to deep convolutional generative adversarial networks (DCGAN) in 2015.

3. Challenges in Time Series GANs:

- **Training stability** poses a challenge in time series GANs due to the sequential dependencies present in time series data.

  - mode collapse problem, where the generator produces a limited variety of samples or fails to capture the entire data distribution.

  - vanishing gradient problem, where small gradients hinder the learning process of the generator and discriminator.

- **Evaluation** of time series GANs is crucial for analyzing their quality and performance.(qualitative & quantitative)

  - Evaluation metrics T-distributed Stochastic Neighbourhood Embedding (for higher dimensions) and Principal Component Analysis (PCA) for dimension reduction are used.

- **Privacy risks** associated with generating time series data, such as violating privacy regulations and exposing sensitive information.

4. Datasets and Applications:

- Unlike image-based datasets, standardized or commonly used benchmark datasets for time series generation are currently lacking.

- Datasets used in time series are signals made up of highly complex waveforms (physiological and audio)

- Relevant repositories like the CR Time Series Classification/Clustering database and the UCI Machine Learning Repository.

- Challenges with discrete time series generation = The distribution on discrete objects are not differentiable with respect to their parameters , so zero gradient.This limitation makes the generator untrainable using backpropagation alone.
Generator makes random sampling and deterministic transform from gradient loss of discriminator push closer to desired output.

-Challenges with continuous time series generation. =Complex correlations exist between the temporal features and their attributes eg. ecg of person related to age/health
As per image reduction which has standard dimension there is no such thing for time series

5. Notable Time Series GAN Models:

- Recurrent Generative Adversarial Network (RGAN) proposed in 2016.(LSTM)

- Sequential GAN for discrete time series data, utilizing recurrent neural networks (RNNs) with LSTM cells and a convolutional neural network (CNN) discriminative model.

- Quant GAN, designed for financial time series data, particularly volatility clusters.

- C-RNN-GAN, capable of learning characteristics of continuous sequential data and generating music.

- NR-GAN, focusing on noise reduction in continuous time series, especially EEG data(brain signals).

- TimeGAN - recovery loss-which encourages the generator to preserve the autocorrelation structure of the real data. This ensures that the generated time series data closely resembles the original data in terms of temporal patterns and dependencies.

- Conditional Sig-Wasserstein GAN (SigCWGAN) (June 2020) - autoregressive feed-forward neural network (AR-FNN) is used for higher dimensionality requirements problem(complex)

- Decision-Aware Time Series Conditional GAN (DAT-CGAN) (Sept. 2020) = used mainly in financial portfolio ,supports decision processes by end users due to incorporating a decision-aware loss function

6. Evaluation Metrics and Privacy Preservation:

- Evaluation metrics used in assessing the performance of time series GANs- qualitative and quantitative metrics.

- empirical estimation of Maximum Mean Discrepancy (MMD) as a suitable metric for evaluating GAN performance across domains.MMD provides a quantitative measure of the dissimilarity between the distributions, helping to assess how well the synthetic data captures the characteristics of the real data.

- Work is ongoing to develop machine learning methods with privacy-preserving mechanisms such as differential privacy.

Decentralized/Federated Learning is one option for it However, it should be noted that they did not experiment with differential privacy in this study but it as an avenue of future work.

Models for privacy preservation - Synthetic Biomedical Signals GAN (SynSigGAN) (Dec. 2020).

Sequentially Coupled GAN (SC-GAN) (April 2019).

Recurrent Conditional GAN (RCGAN) (2017).