Federated Time-Series Generative Adversarial Networks (FeTGAN)

Marcus Plesner, under supervision of Lydia Chen, Zilong Zhao, and Aditya Kunar

M.K.Plesner@student.tudelft.nl

Abstract

The key to producing high-fidelity time-series data is to preserve temporal dynamics. To this end, a novel federated framework (FeTGAN) is proposed, which generates realistic time-series data, by combining supervised and unsupervised training. The framework is based on the work in TimeGAN and Federated GAN (FeGAN). Using an embedded learning space, TimeGAN encourages the network to mimic the structure of the training data. FeGAN allows the results of TimeGAN to be combined at a central server, which has benefits for both throughput, and potential to improve data privacy. The novel framework also shows the ability to integrate cross-domain data from different nodes. The novel framework proposed in this paper demonstrates the ability to produce equivalent quality synthetic time-series data compared to the original TimeGAN.

Introduction

Generative adversarial networks are used to create high fidelity synthetic data.

GAN's have been shown to perform well at tasks, such as text to image synthesis, drug discovery, and privacy maintenance.

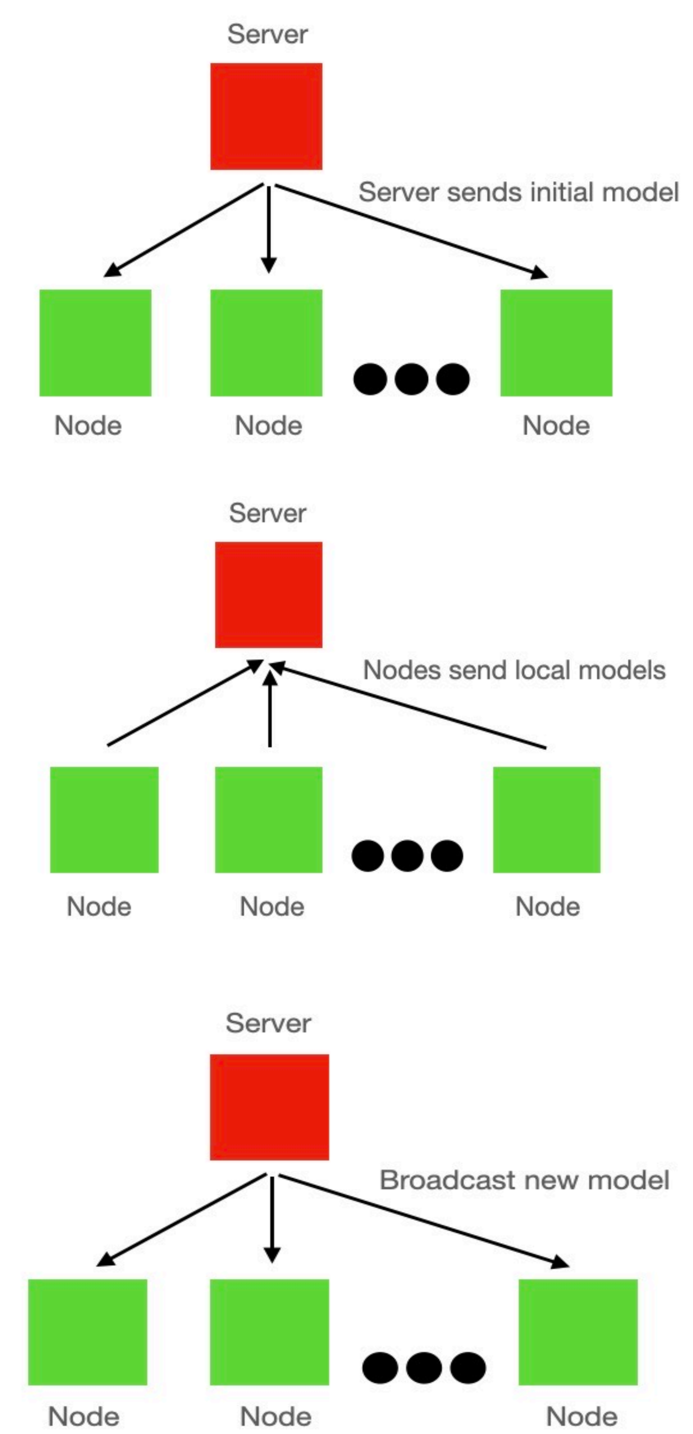
A generative adversarial networks (GANs) is composed of two neural networks. The generator, which produces fake data, and the discriminator which tries to distinguish fake data from real data.

Methodology

Every node in a federated learning system runs TimeGAN, and the server aggregates the results.

The weights of the models are combined using loss based weighting.

The following figure demonstrates the FeTGAN model. The initial model is broadcasted, then the training is done for 100 iterations. The worker models then submit their model to the server, it is aggregated, and then broadcasted to nodes.



Results

The following results are based on the Google stock dataset, which was used in the original TimeGAN.

The discriminative score is the error score for a post-hoc time-series classification model classifying sequences from the original and generated datasets

The predictive score is the means squared error of a post-hoc sequence-prediction model which predicts next-step temporal vectors over each input sequence.

	Centralised	Two Workers	Eight Workers
Discriminative	0,12285	0,14536	0,16834
Predictive	0,04184	0,037822	0,038169

t-SNE and PCA analyses are done on both the original and synthetic datasets, in order to flatten the temporal dimension.

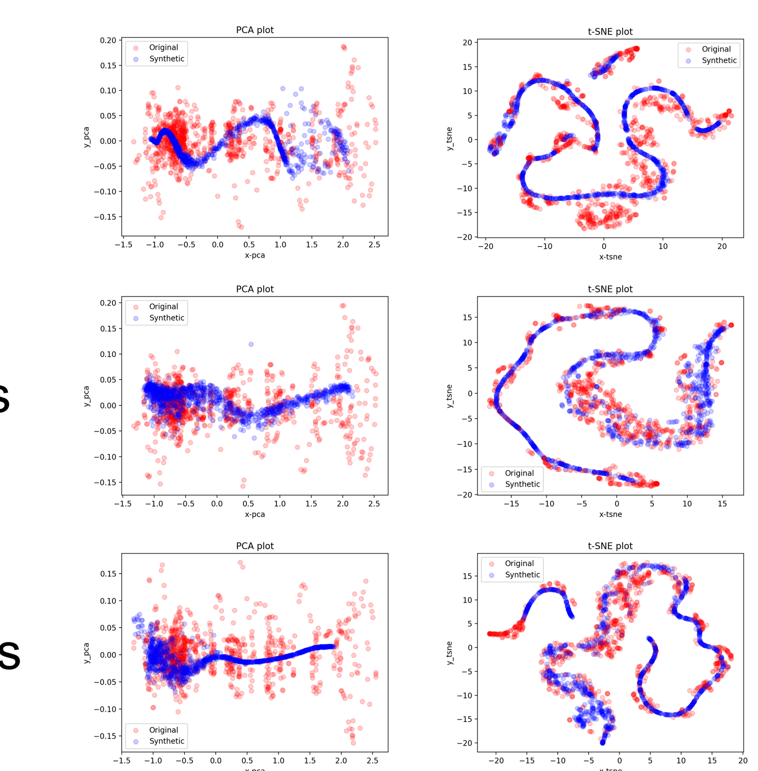
The following diagrams show the PCA and t-SNE analysis for a the original (centralised) TimeGAN, two workers, and eight workers in FeTGAN.

PCA t-SNE

Centralised

Two Workers

Eight Workers



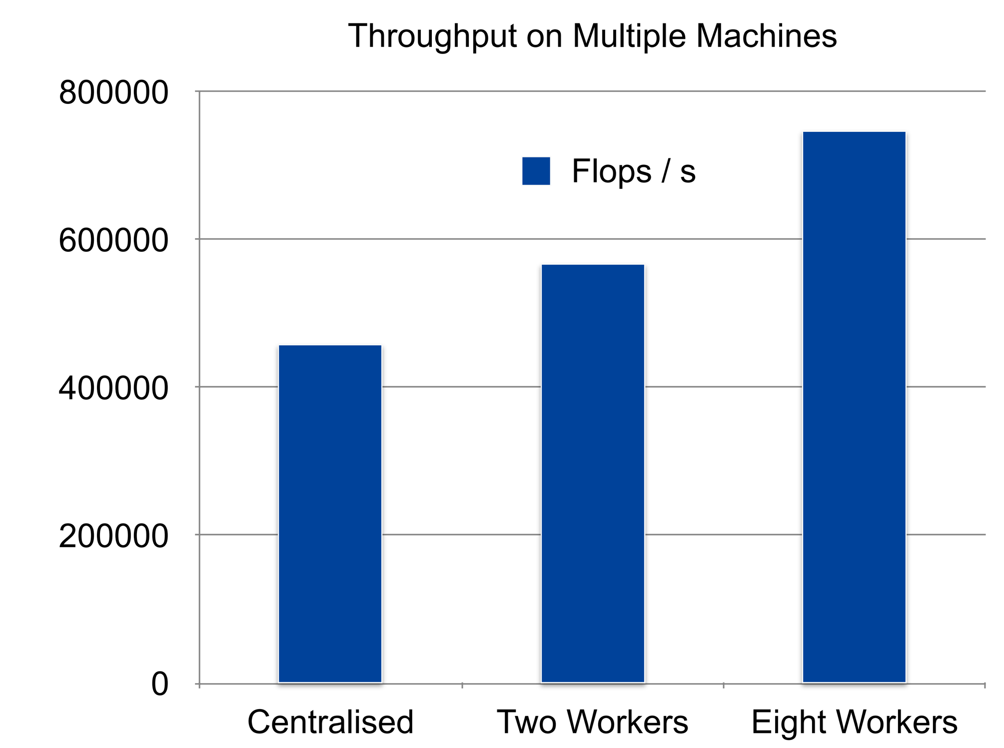
The following shows the results of FeTGAN with one node trained on Jacksonville weather data, and another node trained on Charlotte data (Aggregated), compared with TimeGAN trained on Jacksonville data only (Centralised). This also simulates cross domain data usage.

	Centralised	Aggregated
Discriminative	0.32497	0.29115
Predictive	0.27293	0.28251

The following compares the FeTGAN model to TimeGAN trained on Charlotte data.

	Centralised	Aggregated
Discriminative	0.41439	0.30593
Predictive	0.30675	0.25921

Throughput is measured in FLOPS per second. The following graph shows the FLOPS per second of TimeGAN, and FeTGAN.



Conclusion & Further Work

In this paper, FeTGAN is introduced as an extension of the original TimeGAN implementation, but adapted for federated learning using parts of FeGAN. This combination leverages the embedding and recovery network developed in TimeGAN to create high fidelity time-series data, while using part of the aggregation process described in FeGAN. Federated TimeGAN demonstrates significant throughput improvements in generating high fidelity time-series data.

Further work should seek to integrate differential privacy, and improve the currently naive sampling used in Federated TimeGAN. Furthermore, a weighting scheme that is more appropriate for time-series data should be found. Future research should also be directed at cross domain data usage.

References

1] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. Technical report, University of Cambridge and University of California Los Angeles, 2019.

[2] Rachid Guerraoui, Arsany Guirguis, Anne-Marie Kermarrec, and Er-wan Le Merrer. Fegan: Scaling distributed gans. Technical report, EPFL and Univ Rennes, Inria CNRS Irisa, 2021.