GANs4RE

GANs for Renewable Energy

Kognitive Energiesysteme

M. Jürgens, C. Roelofs, B. Hoppmann, F. Rehwald

Kontakt: Björn Hoppmann| +49 561 7294-1606 | bjoern.hoppmann@iee.fraunhofer.de

Motivation

Getting reliable and anonymous data often is a crucial issue in research on wind turbines and their electric load profile. For this reason, this project aims to generate synthetic but realistic time series of wind turbine data using a GAN architecture. The main purpose is to improve anomaly detection due to the additional amount of realistic data. Besides that, the generated data can also be used for various other training purposes.

Generative Adversarial Networks (GANs) form a powerful framework for synthetic data generation and have predominantly been used to generate extremely realistic images. The usage of them in time series generation, however, has not been fully explored yet.

Generative Adversarial Networks

- Two neural networks, the Generator and the Discriminator perform a min-max game through simultaneous training
- Usage of additional conditions (time features, labels, locations), that are fed into both networks, improve the performance of the model as well as enable to generate more specified samples

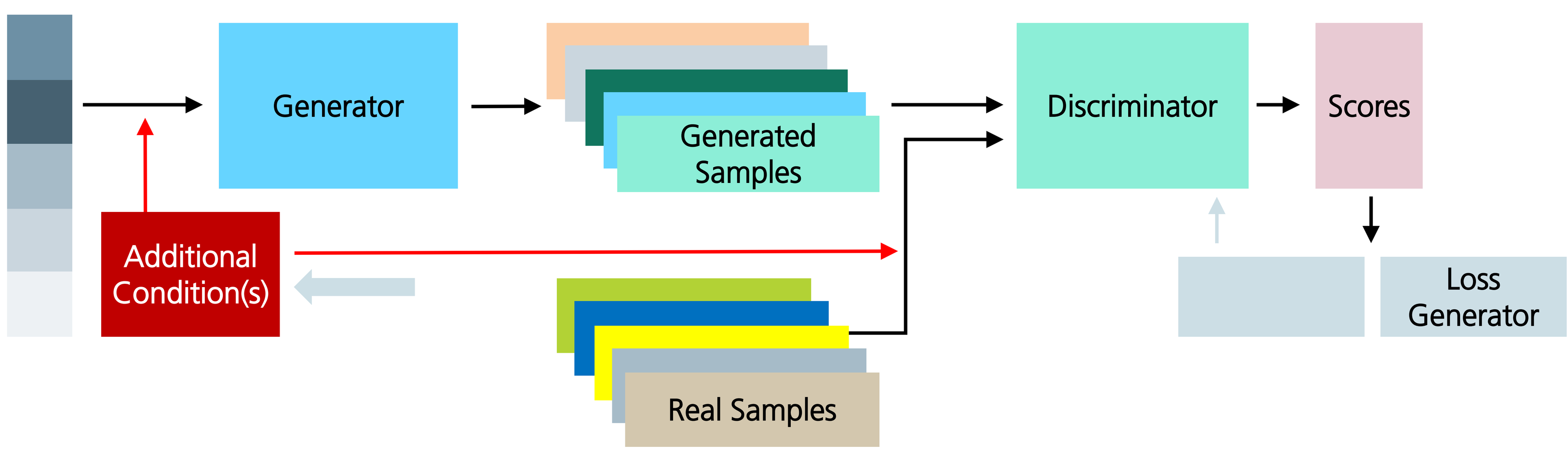

Figure 1: General Structure of a Conditional GAN (CGAN).

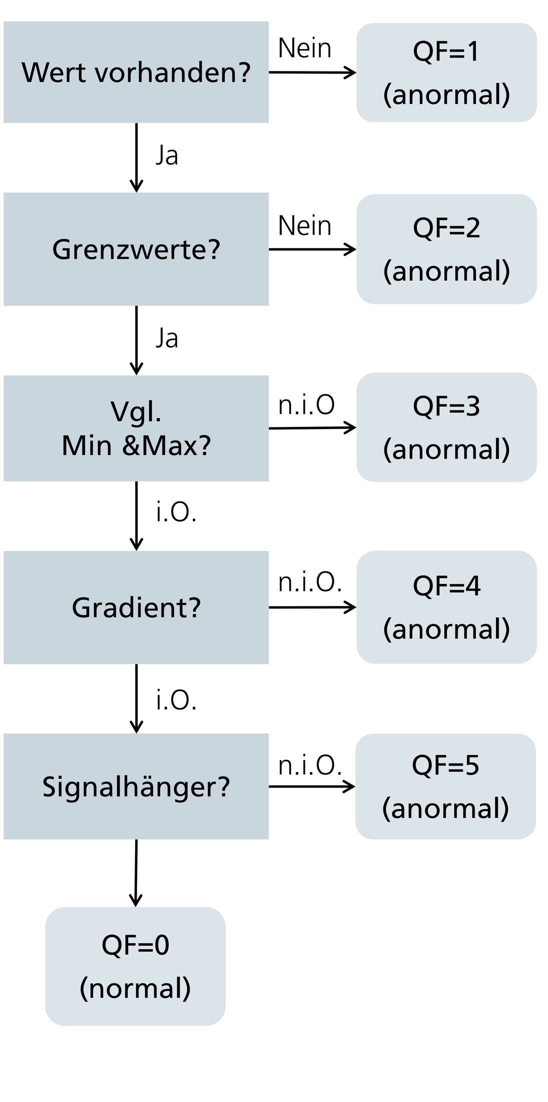

Figure 2: Preprocessing pipeline for flagging time samples.

Data

The data used in the experiments all originate from the SCADA system of an offshore wind farm. Seven characteristics were used (wind speed, output power, wind direction, ambient temperature, pitch angle 1-3). All features were available as 10-minute averages and were preprocessed with the following steps:

1. Flag samples as normal/anomalous (see Figure 2)
2. Create sequences with 32 timestamps and an overlap of 31 timestamps
3. Add conditions: anomalous/normal, timestamp embedding and location of wind turbines

Evaluation

In order to evaluate the „realness“ of the synthetic time series samples, we made use of different criteria:

- Visual comparison: comparison of correlation heatmaps (auto-correlation as well as correlation between pairs of features), real and synthetic samples, histograms, and lower dimensional embeddings (t-SNE, PCA)

Figure 3: Correlation heatmaps between pairs of features for real data and data generated by a CGAN model.

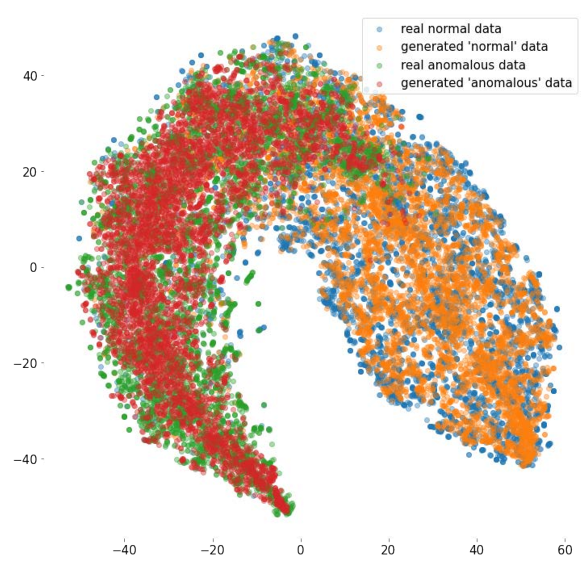

Figure 4: t-SNE of real and synthetic samples labeled as normal (blue, orange) and anomalous (green, red).

- **Maximum-Mean-Discrepancy**: Usage of a kernel-based statistical test measure to compare the distributions of real and synthetic data
- **Train on Synthetic - Test on Real (TSTR)**: Usage of external models (LSTM, Autoencoder) for prediction and classification tasks that are trained on the synthetic data, and then evaluated on the real data

a)

	MMD
Train data (real vs synth.)	0.05 ± 0.006
Validation data (real vs synth.)	0.31 ± 0.005
Validation and train data (real vs real)	0.28 ± 0.005

b)

	LSTM Predictive Score	AE Reconstruction Loss
TRTR	0.0306	0.003
TSTR	0.0531	0.012

Figure 5: Different evaluation scores based on a three months training horizon, validated on a wind turbine unknown to the model. In a) the MMD between real and generated data is shown for the train data and validatiaon data. In b) the prediction error of an LSTM, as well as the reconstruction error of an autoencoder is given for both TSTR and TRTR.

In this project, we developed a widely applicable CGAN model using an architecture specifically adapted to time series. Evaluation metrics and scores based on the TSTR principle indicate that the model is able to learn the distribution of the training data almost perfectly. However, when tested with features from data that the model has not seen before, there is a significant discrepancy in performance. It remains a challenging task to evaluate whether this is caused by overfitting to the training data or if the underlying structure was not actually learned.

Future work might include:

- Conditioning the CGAN to weather as a next step, which yields the potential to generate data for a whole wind park based on only two features.
- Using the already developed autoencoder model to further investigate anomaly generation and detection.

Gefördert durch:

HESSEN
Hessisches Ministerium für Wissenschaft und Kunst

SPOTLIGHTS

Fraunhofer IEE