CIA | ARCANA

XAI – Explainable AI for Anomaly Detection in Wind Turbines

C. Roelofs, F. Rehwald, R. Heinrich, A. Lutz, I. Ghosh

Contact: Cyriana Roelofs | +49 561 7294 1575 | cyriana.roelofs@iee.fraunhofer.de

Kognitive Energiesysteme

Motivation & Goals

- Highly complex blackbox models are used for anomaly detection in wind turbines, making it hard to interpret detected anomalies.
- Explainable AI (XAI) is needed to identify underlying root causes of anomalies.
- Interpretable anomalies can be used to improve the models and in turn reduce false positives and increase recall.
- CIA helps to identify modeling issues regarding anomaly concepts. The method shows which kind of anomalies can be detected and which cannot.
- ARCANA provides human interpretable explanations of detected anomalies.

CIA – Concept-based Interpretable Anomaly Detection

- Use Explainable AI methods to identify which concepts AI-based methods have learned for detecting anomalies in performance time series
- Demonstrate the usability of concept-based XAI methods for anomaly detection in time series data.

Data

- As a first approach data quality flags were used as anomaly concepts.
- As a second experiment, concept vectors were formed using an AE-based (see Autoencoder-based Anomaly Detection) anomaly detection algorithm by clustering the detected anomalies

TCAV

- Concept-datasets are created for each concept
- Concept weights are compared to the model-weights

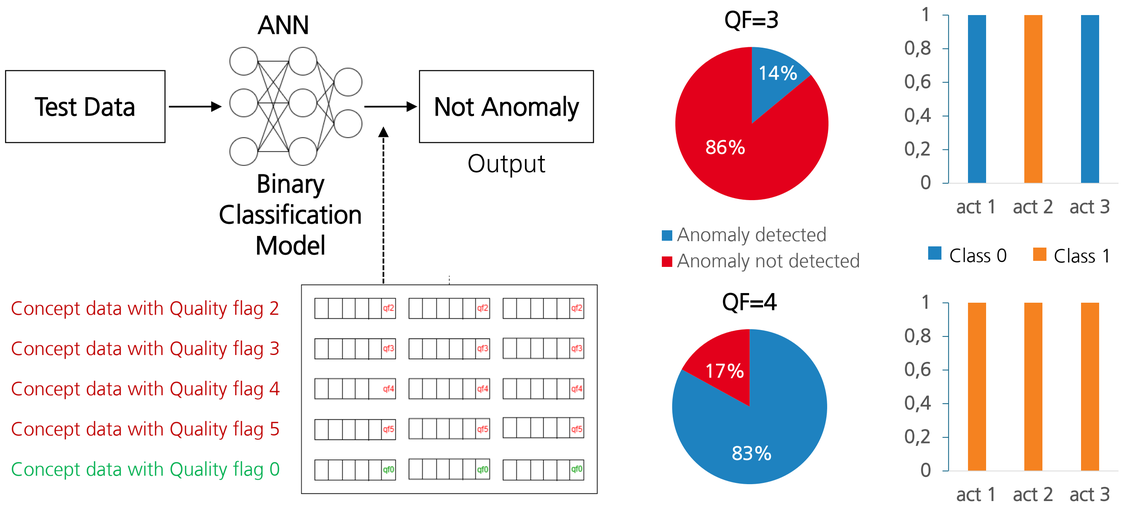

Figure 1: Concept of the TCAV and evaluation

Conclusion

- The functionality of the TCAV method for time series could be shown.
- However, the TCAV method is only conditionally suitable for unsupervised anomaly detection algorithms.
- Deriving concepts from the already detected anomalies is only partially useful. (It can be shown that certain concepts are harder to learn.)
- A possible approach here could be to check concepts for normal behavior with experts.

Autoencoder-based Anomaly Detection

Autoencoders are a type of neural network that reconstructs the input data. They consist of an encoder and a decoder. By training an autoencoder only on data without anomalies, the model learns to encode and reconstruct normal behavior only. The autoencoder will fail to reconstruct anomalous input data correctly, resulting in a large reconstruction error (RE).

ARCANA – Anomaly Root Cause Analysis

- Autoencoder models have proven to be very successful in detecting anomalous behavior in wind turbine sensor data, yet cannot show the underlying cause directly. Such information is necessary for the implementation of these models in the planning of maintenance actions.
- For this problem, a novel method for autoencoder-based anomaly detection root cause analysis – ARCANA – was developed.
- ARCANA is an optimization algorithm that identifies only a few, but highly explanatory anomalous features.

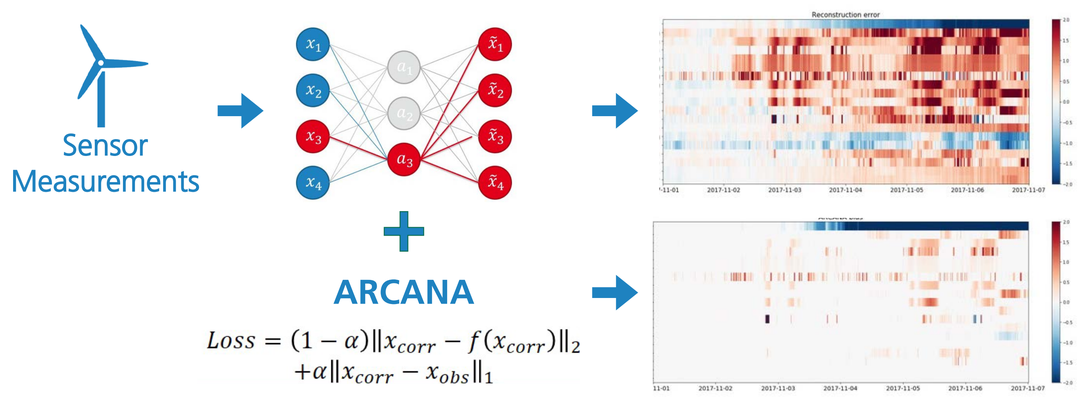

Figure 3: An autoencoder model is trained using 1 year of wind turbine sensor data. If only one feature of the input data shows an error (sensor failure), the error is propagated throughout the network and perturbs all outputs (upper right). ARCANA shows that the actual cause of the anomaly is one specific feature (lower right).

Case Study

Due to a component change in the wind turbine, the converter water conductance changed and was detected as an anomaly. The features with the largest reconstruction errors do not explain this anomaly, whereas ARCANA attributes more importance to the water conductivity, which directly influences the converter water conductance.

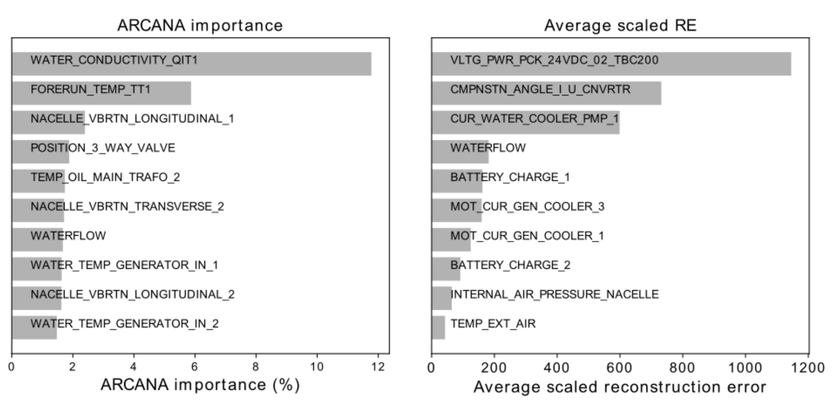

Figure 4: Time window between 2020-05-30 08:40 and 2020-06-02 00:00

Gefördert durch:

HESSEN
Hessisches Ministerium für Wissenschaft und Kunst

Scan me to read more!

SPOTLIGHTS

Fraunhofer IEE