CNNs for Fault Detection: a Wind Turbine Use Case

Lilach Goren Huber
Markus Ulmer
Jannik Zgraggen

lilach.gorenhuber@zhaw.ch

2) Transfer Learning Approaches for Fault Detection (Jannik, 6.5.2021).
Fault Detection

• Anomaly detection: detect deviations from a familiar pattern ➔ anomaly score.
• Fault detection: detect early patterns of deviations from normal behavior of machines ➔ Health Index.
• Common machine data: multivariate time series.

Complexities
• Anomalies are rare ➔ class imbalance
• Unknown fault nature (time dependence, distribution)
• Diversity of anomaly classes/types (abrupt, slow degradations, point anomalies, different severity)

Challenges
• Thershold setting ➔ False positives vs. Missed detections.
• Model evaluation.
• Detection + Explanation => diagnostics.
• Noise resilience under diverse operating conditions.
Anomaly Detection

- Need labeled faults
- Few clear outliers
- Train on healthy data

**Supervised**
- Classification
  - Clustering
  - OCSVM
  - Isolation forest

**Unsupervised**
- Distance/density based
  - PCA
  - NNs

**Semi-supervised**
- Dimension Reduction
- Reconstruction
- Regression/prediction
  - Stat. models
  - AR models
  - Reg. Trees
  - NNs

**Dimension Reduction**
- PCA
- Autoencoder

**Reconstruction**
- PCA
- NNs

**Regression/prediction**
- Stat. models
- AR models
- Reg. Trees
- NNs

**Feature engineering**
- X1
- X2
- X3
- X4

**Model**
- Model

**Large residual ➔ fault**
Trends in Fault Detection

**Academic Research**

- Method driven → classification mainly
- Encode ts data as images → limited applicability.
- RUL predictions (simulated data)
- Transfer Learning/ Domain Adaptation
- Hybrid models (physics + ML)
- Interpretability («XAI»)
Trends in Fault Detection

Academic Research

• Method driven ➔ classification mainly
• Encode ts data as images ➔ limited applicability.
• RUL predictions (simulated data)
• Transfer Learning/ Domain Adaptation
• Hybrid models (physics + ML)
• Interpretability («XAI»)

Practical R&D

• Use existing data (little, no faults, low freq)
• How to set robust thresholds?
• How to evaluate algorithms without labels?
• How to scale up algorithms (data selection, run times)?
• Robustness under changing conditions.
• Diverse operating conditions (TL/DA?)
Who are we?

- Smart Maintenance Team
- Data intelligence platform for renewable energy assets
- > 7 GW of monitored assets
- Electricity provider of Canton of Zurich

Intelligent fault detection and diagnosis algorithms for wind turbines

Added Value for Park Operators

Original Equipment Manufacturer (OEM):

- Designated Condition Monitoring hardware.
- High frequency data.
- Use for design improvement.

Not accessible for operators/owners.
Added Value for Park Operators

Cost effective solution:
- Use available data.
- No need for additional hardware.

10-minute averaged SCADA data is already stored.

Sensor data = time series

- Power
- Wind speed (ws)
- Temperature (Temp)
- Generator bearing, rpm
Challenge

High performance fault detection with existing data

Practical solution:
- Detect early
- Accurately
- Diverse Fault types
- Transferable
- Scalable

10-minute averaged SCADA data is already stored.
Semi-supervised Anomaly Detection

✓ Faults are rare and unique ➔ use only healthy data for training, detect anomalies online.
✓ Fault localization ➔ regression

Input:
- SCADA sensor data

Output:
- Predicted component temperature

Prediction Model

Train with normal data

- \( \hat{y}(t) \)
- \( y(t) \)
- \( \delta(t) = y(t) - \hat{y}(t) \)

Fault localization ➔ regression

Large residual ➔ fault

\[
\begin{align*}
\text{Power} & : X_1(t) \\
\text{Wind speed} & : X_2(t) \\
\text{Amb. T} & : X_3(t) \\
\text{rpm} & : X_4(t)
\end{align*}
\]
Prediction model
Single Output CNN

✔ Learn time dependent patterns ➔ CNN

Commonly used: Multi Layer Perceptron (MLP)

\[ x_1(1) \cdots x_1(t) \]
\[ x_2(1) \cdots x_2(t) \]
\[ x_3(1) \cdots x_3(t) \]
\[ x_4(1) \cdots x_4(t) \]
Diverse fault types ➔ multi-target regression
Model Evaluation and Selection

- Regression models: metric = prediction error MSE.
- AD: training on healthy data ➔ we can only evaluate prediction error and minimize it.
- **Problem: model selection.** reducing the error on the healthy data does not imply increased error for fault data.
- Example: we can select a predictor with perfect correlation to the target. We then have good predictions also during abnormalities.

➔ select a minimal set of «exogenic» predictors.

Without nacelle T as input

With nacelle T as input
Over fitting

prediction

training

training
Over fitting

Prediction error[°C]

Over fitting: Prediction errors at testing do not grow training get smaller ➔ False positives.
Example
Detection of Abrupt Faults

Turbine A₀
Abrupt faults in the generator bearing

Turbine shutdowns

(b)
Results
Residuals and Health Index: Abrupt Faults

Turbine $A_0$
Abrupt faults in the generator bearing

$$\delta(t) = y(t) - \hat{y}(t)$$

Health Index

Turbine shutdowns
Results
Health Indices with Multi-output CNN

(a) Generator bearing (b) stator phase (c) gearbox bearing (d) gearbox oil (e) slip ring
(f) rotor spinner (g) grid transformer (h) controller top (i) controller hub (j) hydraulic oil.

Lilach Goren Huber - ZHAW
Results
Model Comparison: Abrupt Faults

Both CNNs perform better than the MLP
Results
Model Comparison: Slow Degradation

Turbine $B_0$
Slow degradation of the main bearing

<table>
<thead>
<tr>
<th>Model</th>
<th>First detection ($\alpha=0.0001$)</th>
<th>First detection high confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>16.10.17 2:00</td>
<td>25.4.18 00:00</td>
</tr>
<tr>
<td>CNN</td>
<td>15.8.17 17:00</td>
<td>10.9.17 19:00</td>
</tr>
<tr>
<td>CNNm</td>
<td>15.8.17 17:00</td>
<td>9.9.17 14:00</td>
</tr>
</tbody>
</table>

- CNN detects 2 months earlier than MLP.
- CNNm detects even earlier than CNN.
Results

Sensitivity Analysis: Slow Degradation

Modify the detection threshold and calculate the date of first detection:


The CNNm is more robust against threshold selection
Scaling up fault detection algorithms

Motivation for transfer learning

- New turbines/farms.
- New operational conditions.
- Speed up fleet-wide detection.

Can we train on one turbine and predict on another?
Solution
Cross-Turbine Training Scheme

**Goal:** predict faults on turbine $T$.

**Problem:** only 3M data from $T$.

**Simple Solution:**
- Train CNN on 9M from turbine $S$.
- Predict $\hat{y}_t$ for new data from $T$.
- Train a regression model $y_t \sim \hat{y}_t + X_t$ on 3M from $T$.
- Adapt new predictions on $T$ using the same regression.
Cross-turbine training scheme allows early fault detection with scarce data

(a) Trained on $T$ limited

(b) Trained on $T$

(c) Trained on $S$
Results
Source domain comparison: abrupt fault

Summary

- Fault detection of rare and diverse fault types ➔ semi-supervised AD
- Easy fault localization ➔ regression
- Robust capturing of time dependent patterns ➔ CNNs
- Scalable multi-component detection ➔ multi-output CNN.
- Robust threshold setting ➔ use error distributions for anomaly scores.
- Model evaluation ➔ MSE not enough.

- Scarce data ➔ TL approaches.
- Preliminary: LR-based cross turbine training scheme with promising results.
- Next week: overview of TL approaches, comparison of selected solutions.