
Model-based Data Exploration

Hans-Ulrich Kobialka

Daniel Paurat

Lisa Schrader



Learn from Data



Check your sources!

Industry 4.0:

Offset printing machines



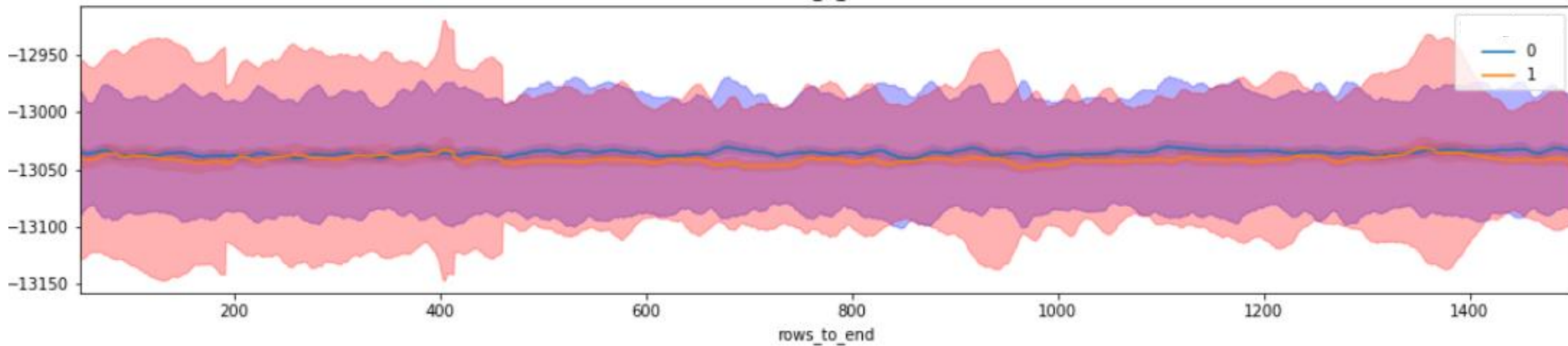
- Multi-dimensional time series
34 sensors (50 Hz) + events recorded in log files
- Labeled *failure* and *non-failure*
- Task: **Root Cause Analysis**
The number of root causes is unknown.

Industry 4.0 data: Common Problems



- unsynchronized clocks
- unknown values
- gaps in time series
- bad signal-to-noise ratio compared to the amount of data available
- differences between machines:
sensors, software versions, calibration ...
- changes over time:
material used, quality criteria, ... sensors, software versions, calibration

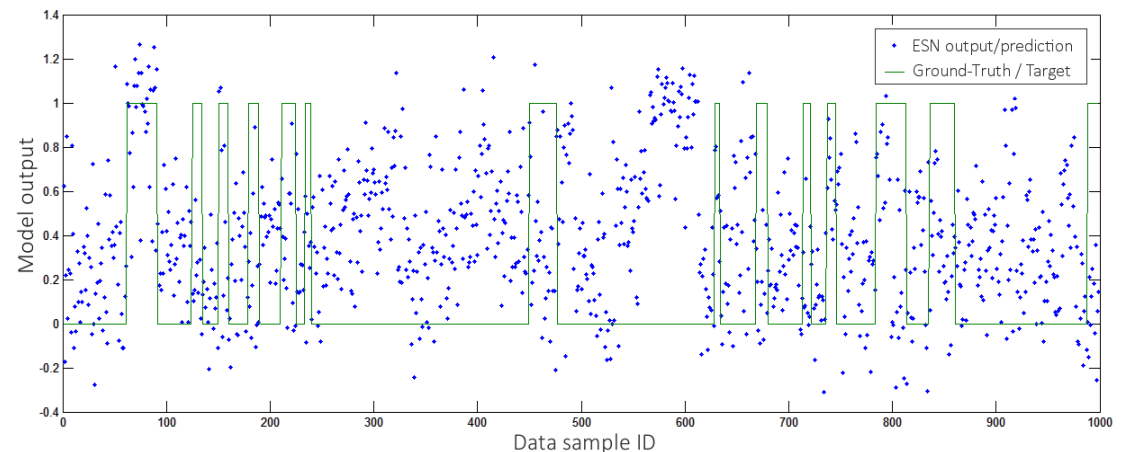
Visual Data Exploration



Industry 4.0 : Offset printing machines



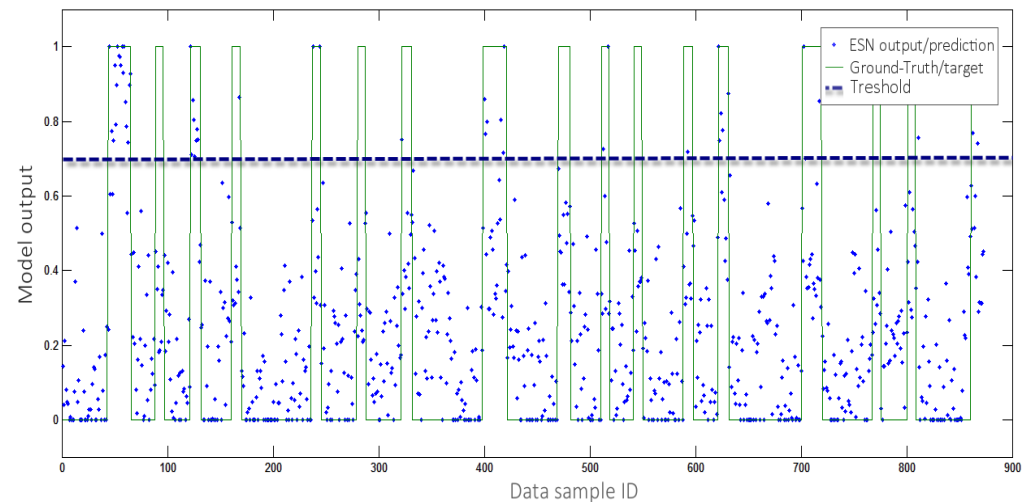
- Multi-dimensional time series
34 sensors (50 Hz) + events recorded in log files
- Labeled *failure* and *non-failure*



Model-based Data Exploration

Conducting Experiments to obtain Insights into the Data

- ease the learning task by selecting a subset of data
 - Iteratively exclude 5% of all failure samples having largest training error
 - **CAUTION:** Removing non-failure samples to would yield perfect but meaningless results.



Model-based Data Exploration

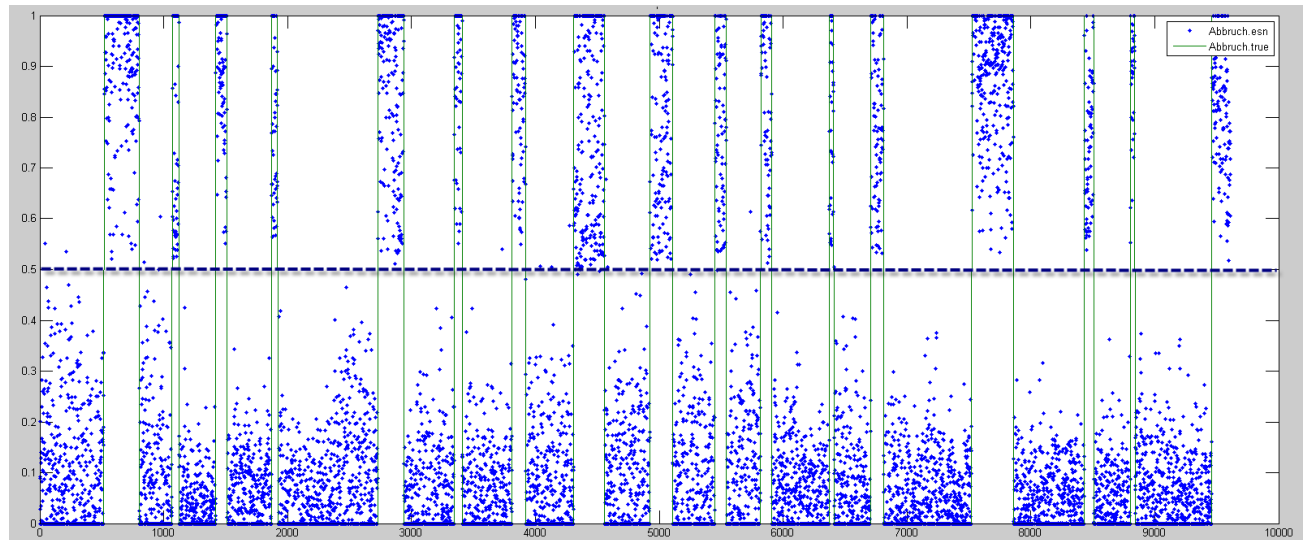
Conducting Experiments to obtain Insights into the Data

- ease the learning task by selecting a subset of data
 - use this subset to decide e.g. on preprocessing choices.
 - group failure samples into cluster
 - label failure samples according to the earliest time-to-failure at which they can be predicted
 - define groups of sensors and determine the failure samples which be labeled by a sensor group
- Domain Experts
- ⇒ 2207 (47%) of 4690 failures can be detected by paper sensors between 500 and 1000 milliseconds before the failure occurs.

Model-based Data Exploration

Intermediate Results

2207 (47%) of 4690 failures can be detected by paper sensors between 500 and 1000 milliseconds before the failure occurs.



- To the other 3 groups of sensors, significantly less samples (232 to 246) can be associated to.

=> We trained a model to detect failures caused by paper supply.

Model-based Data Exploration

About building Models

- Models can be built by any ML method.
- In our experiments, we used Echo State Networks (ESN)
 - ESNs are invented & patented by Fraunhofer IAIS
 - ESNs are large RNNs which can be trained fast
 1. all time series are processed by a large RNN
=> one large state vector for each time series
 2. On the state vectors of all training time series, linear Regression is performed to train the output weights.
 - State vectors can be reused, e.g. during cross validation.

Industry 4.0 data:

Outlook



- We trained a model to detect failures caused by paper supply.
- What about the root causes of the other 53% failures?
- Project is on hold while waiting for new data having better quality.
- Then we will have to discuss with the domain experts
 - promising groups of sensors
 - refinement of the paper supply root cause

Conclusions

- In Industry 4.0 projects, we face bad signal-to-noise ratios.
- Model-based Data Exploration is on conducting Experiments to obtain Insights into the Data.
- Domain experts can give hints to do the right experiments first. This can speed up the process significantly!
- Clustering of failure samples to root causes requires intensive interaction with domain experts.