Air pump fault / anomaly detection

Antonín Hruška

April 4, 2024

Task definition

What we want?

Formal Definition:

Task: Fault detection / isolation (offline)

Objective: An (optimal) classifier $h : \mathbb{R}^n \to \{0, 1\}$ to predict fault occurrence

Supervised classification on time series

Informal Description:

Pump broke / pinpoint its cause (after it happend)

A system that examines a time series of measurments and determines if a specific fault happend

We have labeled data (even those, were fault occured)

Dataset $\mathcal{D}_{\mathcal{T}} = \left\{ (\{x_1, x_2, ..., x_{n^i}\}, y^i) |\, i \in \{1, ..., m\} \right\}.$

Theoretical background and our knowledge

What we know?

Fault / anomaly detection:

- Traditionaly we have only working model $p_{\text{working}}(\boldsymbol{x})$
- Anomalies are measurments highly unlikely under the model.

Fault isolation approach:

- Assumes models for both: $p_{\text{working}}(\boldsymbol{x})$ and $p_{\text{faulty}}(\boldsymbol{x})$
- Based on hypothesis testing and likelihood ratio tests, i.e.

$$\log(p_{\text{working}}(\boldsymbol{x})) - \log(p_{\text{faulty}}(\boldsymbol{x})) > \theta$$

Exploiting physics and environment:

- Air pump is a dynamical system. Use shadow box modelling (State space based on physics, Kalman filter) or use data driven methods (Min Squares, ARMA)
- We would need more details about measurments, the air pump architecture, the enviroment and the reference signal.

Model

Simplistic generative model based on heuristics and examination of the time series.

Naive Bayes Gaussian classifier on first quarter of trimmed sequence.

$$p_{\text{working}}(\boldsymbol{x}) = \pi_{\text{working}} \prod_{j=1}^{n_i} p(x_j|\text{working})$$

$$p_{\text{faulty}}(\boldsymbol{x}) = \pi_{\text{faulty}} \prod_{j=1}^{n_i} p(x_j|\text{faulty})$$

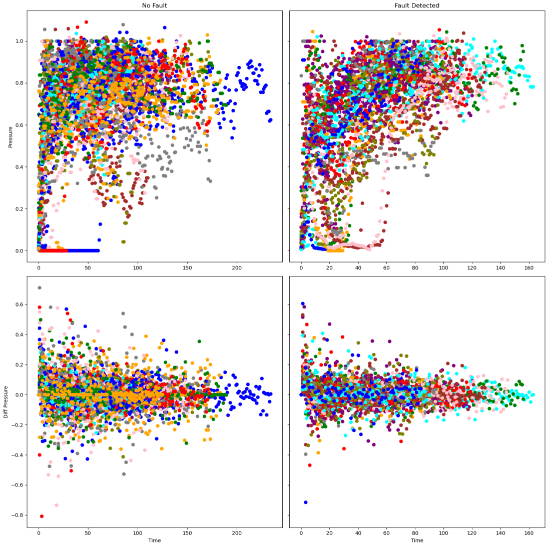

Figure: Trimmed sequences with min-max normalization. Upper row: Pressure, Lower row: Pressure difference. Left column: Regular, Right column: Faulty.

Results and future work

	Actual Positive	**Actual Negative**
Predicted Positive	445	1245
Predicted Negative	128	3562

Table: Confusion Matrix (All predictions: 5380, Accuracy 0.74) on random split of 0.7 trn / 0.2 val / 0.1 test (one run only)

Try more complex model with knowledge of dynamical system and physics.

Or use very simple Deep Learning or Random Forest and hope for strong regularization.