Power Wind Mill Fault Detection via one-class ν-SVM Vibration Signal Analysis

David Martínez-Rego, Oscar Fontenla-Romero and Amparo Alonso-Betanzos

Overview

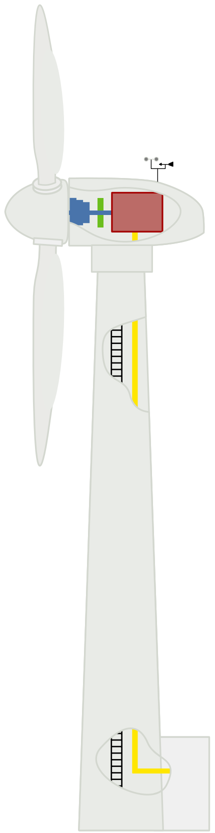
1. Goals
2. Methodology
3. Background
4. Results
5. Review

Vibration Signal Analysis

Goals

- Detection & diagnosis of failures
- Discern between failure and normal data
- Capture evolution of defect
- Accurate results under noisy conditions

- Cost effective scheduling of maintenance
- Avoid fatal breakdowns



„Windkraftanlage“ by Arne Nordmann
http://commons.wikimedia.org/wiki/File:Windkraftanlage.svg

Methodology

- Extract data from possible fault points
- Dimensionality reduction of raw data
- Novelty detection using One-Class SVM
- Three tier evaluation:
  - Simulation
  - Laboratory
  - Real data
- Comparison to ANN

One-class Support Vector Machine (ν-SVM)

Motivation:

- Negative training data for detecting faults is hard to come by
- Train SVM to distinguish between “normal” data and “not normal” data
  - Given data of normal behaviour: Separate probability distribution from origin
    - ➢ Fault is recognized as outliers of SVM decision plane
    - ➢ Novelty detection

One-class Support Vector Machine (ν-SVM)

Data

$$\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$
$$y_i \in \{-1, 1\}$$

Hyper-plane

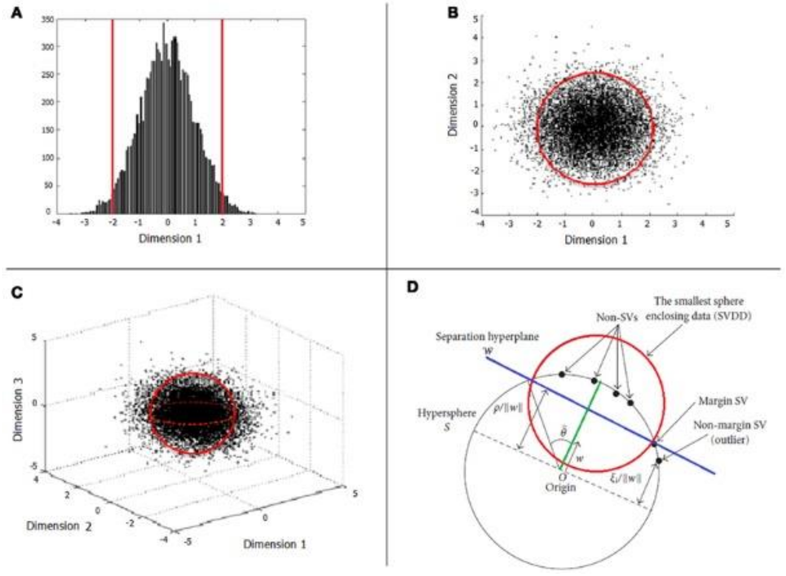
$$w^T x + b = 0 \qquad w \in F \qquad b \in R$$

Convex optimization problem

$$\min_{w,\, \xi_i,\, \rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho$$

subject to:

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i \qquad \text{for all } i = 1, \ldots, n$$
$$\xi_i \geq 0 \qquad \text{for all } i = 1, \ldots, n$$



Sato et al. “Measuring abnormal brains: building normative rules in neuroimaging using one-class support vector machines”

One-class Support Vector Machine (ν-SVM)

TECHNISCHE UNIVERSITÄT DARMSTADT

Data

$$\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$
$$y_i \in \{-1, 1\}$$

Hyper-plane

$$w^T x + b = 0 \qquad w \in F \qquad b \in R$$

Convex optimization problem

$$\min_{w,\, \xi_i,\, \rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho$$

subject to:

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i \qquad \text{for all } i = 1, \ldots, n$$
$$\xi_i \geq 0 \qquad \text{for all } i = 1, \ldots, n$$

ν

- upper bound on the fraction of outliers
- lower bound on the fraction of support vectors

Can be adjusted to account for spurious or abnormal data
Here: $\nu = 0.01$, assuming 1% of spurious data

One-class Support Vector Machine (ν-SVM)

Data

$$\Omega = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$
$$y_i \in \{-1, 1\}$$

Hyper-plane

$$w^T x + b = 0 \qquad w \in F \qquad b \in R$$

Convex optimization problem

$$\min_{w,\, \xi_i,\, \rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho$$

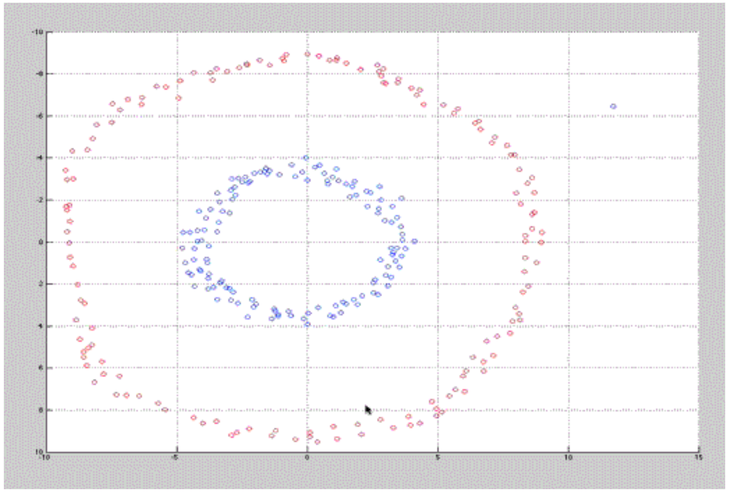
subject to:

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i \qquad \text{for all } i = 1, \dots, n$$
$$\xi_i \geq 0 \qquad \text{for all } i = 1, \dots, n$$

Kernel Function

$$\phi(x_i)$$

here: RBF Kernel with $\sigma = 20$



http://rvlasveld.github.io/blog/2013/07/12/introduction-to-one-class-support-vector-machines/

One-class Support Vector Machine (ν-SVM)

Data

$$\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$
$$y_i \in \{-1, 1\}$$

Hyper-plane

$$w^T x + b = 0 \qquad w \in F \qquad b \in R$$

Convex optimization problem

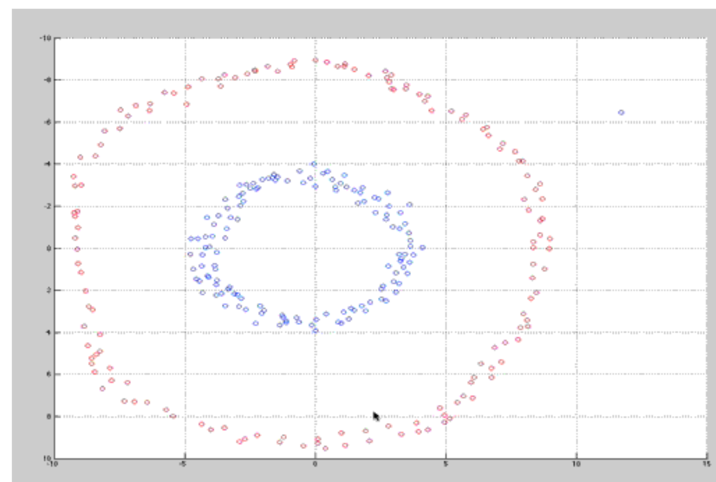
$$\min_{w,\,\xi_i,\,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho$$

subject to:

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i \qquad \text{for all } i = 1, \ldots, n$$
$$\xi_i \geq 0 \qquad \text{for all } i = 1, \ldots, n$$

Kernel Function

$$\phi(x_i)$$



http://rvlasveld.github.io/blog/2013/07/12/introduction-to-one-class-support-vector-machines/

One-class Support Vector Machine (ν-SVM)

TECHNISCHE UNIVERSITÄT DARMSTADT

Data

$$\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$
$$y_i \in \{-1, 1\}$$

Hyper-plane

$$w^T x + b = 0 \qquad w \in F \qquad b \in R$$

Convex optimization problem

$$\min_{w,\, \xi_i,\, \rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho$$

subject to:

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i \qquad \text{for all } i = 1, \ldots, n$$
$$\xi_i \geq 0 \qquad \text{for all } i = 1, \ldots, n$$

$$f(\mathbf{x}) = sgn(\underbrace{\mathbf{w} \cdot \phi(\mathbf{x}) - \rho}_{w \cdot \phi(x_i) - \rho})$$

$$w \cdot \phi(x_i) - \rho$$

“measure of normality”

One-class Support Vector Machine (ν-SVM)

TECHNISCHE UNIVERSITÄT DARMSTADT

Convex optimization problem

$$\min_{w,\,\xi_i,\,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho$$

subject to:

$$\begin{aligned} (w \cdot \phi(x_i)) \geq \rho - \xi_i \qquad & \text{for all } i = 1, \ldots, n \\ \xi_i \geq 0 \qquad & \text{for all } i = 1, \ldots, n \end{aligned}$$

Lagrangian of primal:

$$L(\mathbf{w}, \xi) = \tfrac{1}{2}\|\mathbf{w}\|^2 + \tfrac{1}{\nu l}\sum_i \xi_i - \rho \quad - \sum_i \alpha_i((\mathbf{w}\cdot\phi(\mathbf{x}_i)) - \rho + \xi_i) \quad - \sum_i \beta_i \xi_i$$

Derivatives:

$$\mathbf{w} = \sum_i \alpha_i \phi(\mathbf{x}_i),$$

$$\alpha_i = \tfrac{1}{\nu l} - \beta_i \leq \tfrac{1}{\nu l},$$

$$\sum_i \alpha_i = 1$$

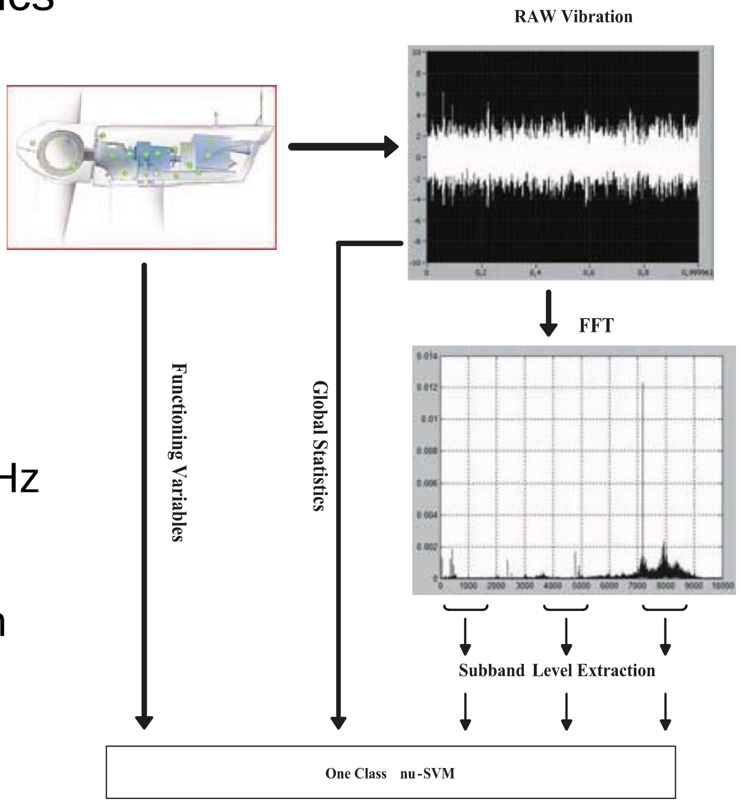
Wolfe dual:

$$\begin{array}{ll} \max\limits_{\alpha} & -\frac{1}{2}\sum_{ij}\alpha_i\alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} & 0 \leq \alpha_i \leq \frac{1}{\nu l} \\ & \sum_i \alpha_i = 1 \end{array}$$

Training Data

Input training data: Normal state vibration statistics

- Root mean square (RMS)
  - Analysis of global vibration level
  - Grows linearly in first stage of failure
  - Grows exponentially in later phases
- Fast Fourier Transform
  - Analysis of sub-bands (50Hz width for 0-2000Hz spectrum)
  - Change in one sub-band can indicate failure in specific component
- Revolutions per minute (RPM)
  - Helps to determine different normal states



Fault detector architecture
Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

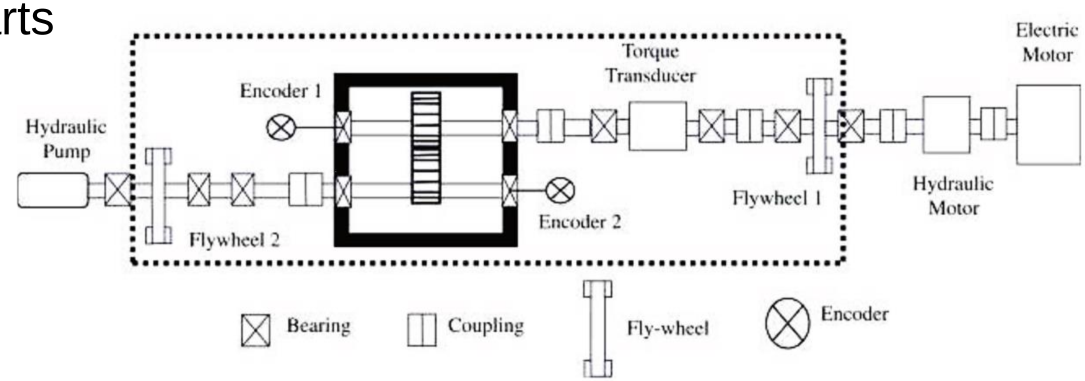
Experimental Results

1) Simulated data scenario
   - Detection accuracy and capacity to present qualitative indicator of defect's evolution
2) Laboratory experimental setting
   - Real vibration data captured in a controlled test-to-failure scenario
3) Real data
   - Application on power wind mill data extracted from production machine in north western Spain

Experimental Results

1) Simulated data scenario

- Simulation model for a gearbox text rig from the University of New South Wales (UNSW)
- Model of whole system of gears and shafts supported by bearings with 34-DOF
- Simulated signals showed similar pattern to actual measured counterparts



Simulation Test Rig Scheme
Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

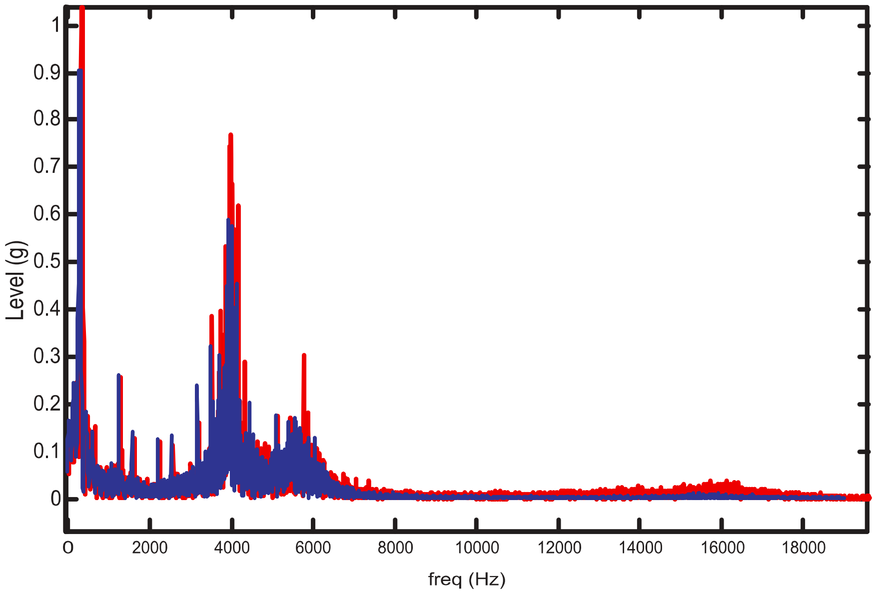
Experimental Results

1) Simulated data scenario
   - 65 captures in normal state
   - 65 captures under linear increasing fault
     - Training using the first half of normal state captures
     - Model subsequently applied to the remaining captures

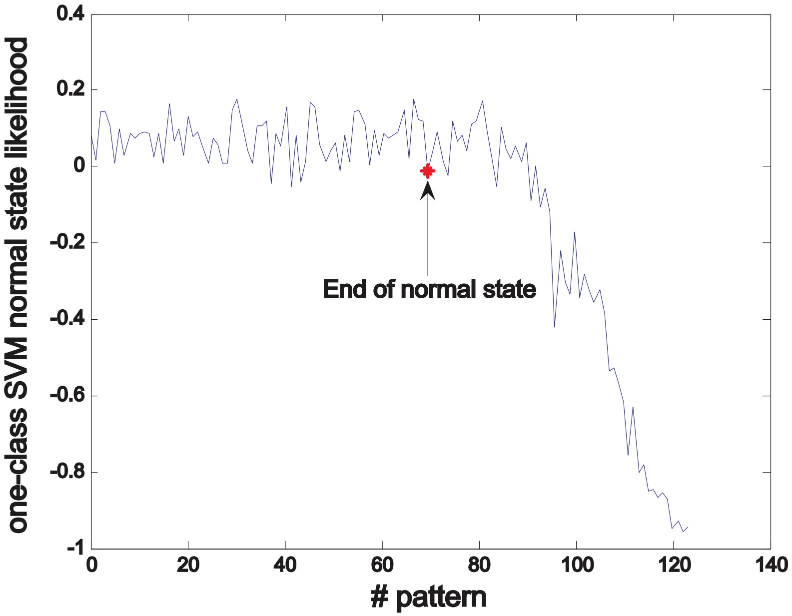
Experimental Results

1) Simulated data scenario

- Fault: $0 - 200\mu m$ depth, $0 - 0.5\ mm$ width
- Incipient fault: $0.3\ mm$ width



Normal state spectrum of simulated data
Incipient fault spectrum of simulated data



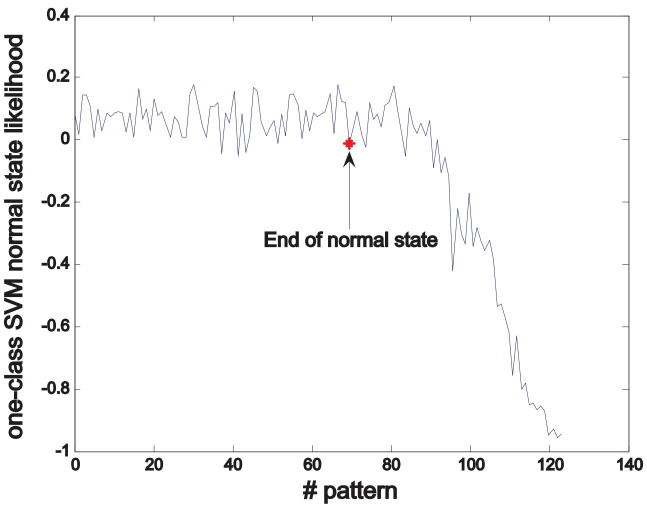
Fault detection for the simulated fault growth case.

Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

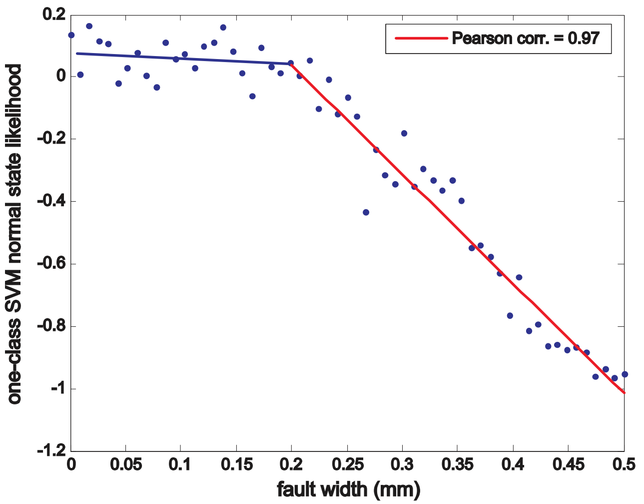
Experimental Results

1) Simulated data scenario

- Fault: $0 - 200\mu m$ depth, $0 - 0.5\ mm$ width



Fault detection for the simulated fault growth case.



Correlation between nomal state assessment of one-class ν-SVM and fault width.

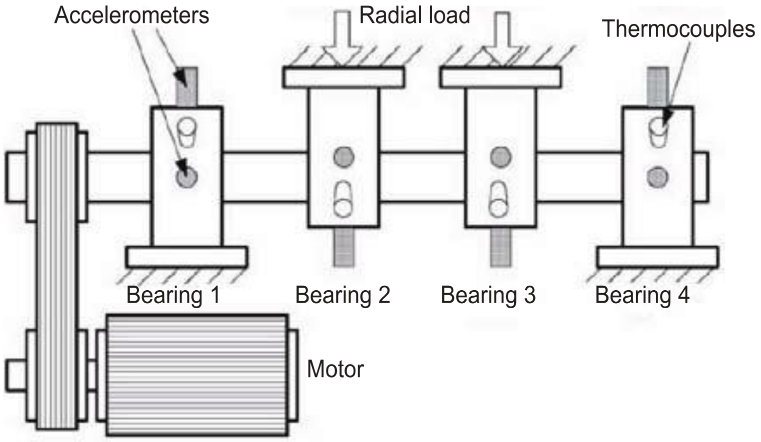
$$Pearson\ Correlation: \quad \rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

Experimental Results

2) Laboratory experimental setting

- Bearing dataset provided by the Center for Intelligence Maintenance System (IMS), University of Cincinnati
- Four bearings in one shaft
- $2000\ rpm$ rotation speed
- $6000lbs\ (\sim 2722kg)$ radial load
- Vibrational data collected every 10 Minutes during 8 days



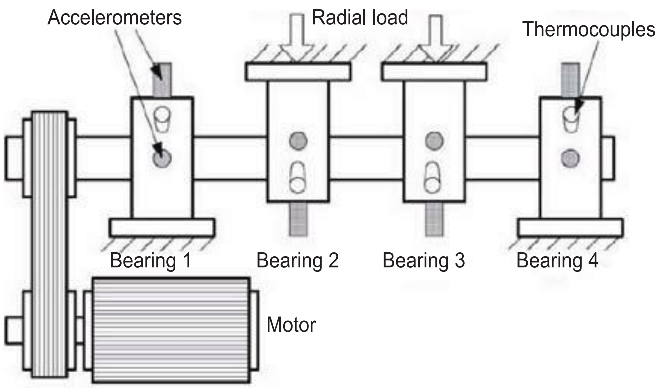
Front Scheme of the installation for the test-to-failure experiment.

Martinez-Rego et. al. "Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis"

Experimental Results

2) Laboratory experimental setting

- Vibrational data collected every 10 Minutes during 8 days
- Nature of induced fault was “slow”
  - Goal:
    - No false alarms under normal operation
    - Early detection of vibration signature change
    - Qualitative indication of evolution until machine stops working



Front Scheme of the installation for the test-to-failure experiment.

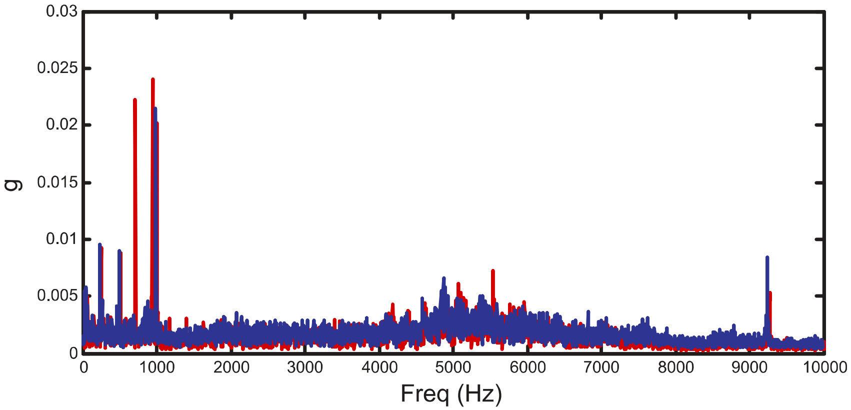
Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

- 200 captures of normal state used for training (40 hours)
  - Rest was applied for detection of possible deviations

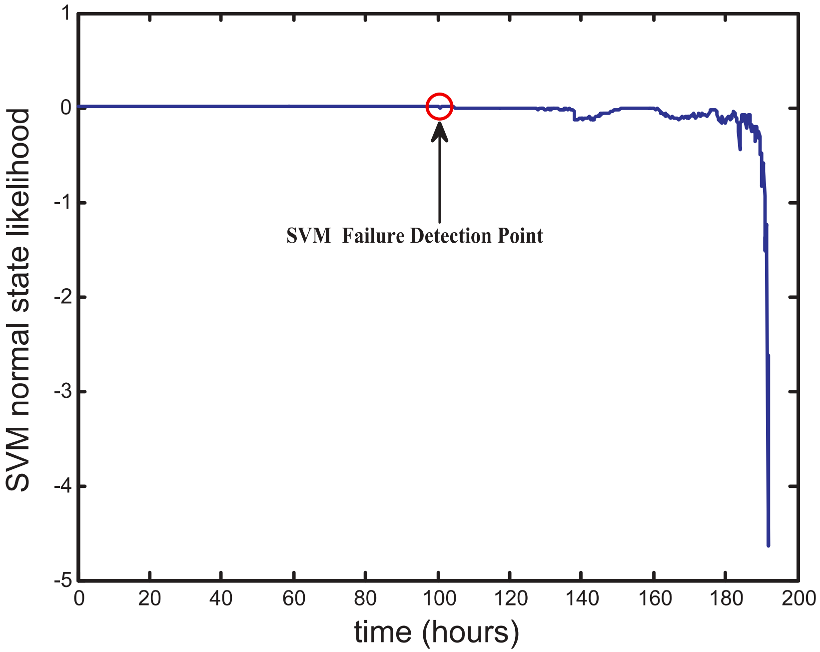
Experimental Results

2) Laboratory experimental setting

- Change of behaviour was detected 92 hours before machine failure



Normal state spectrum of test-to-failure experiment.
Incipient fault spectrum of test-to-failure experiment.



Failure detection on test-to-failure experiment

Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

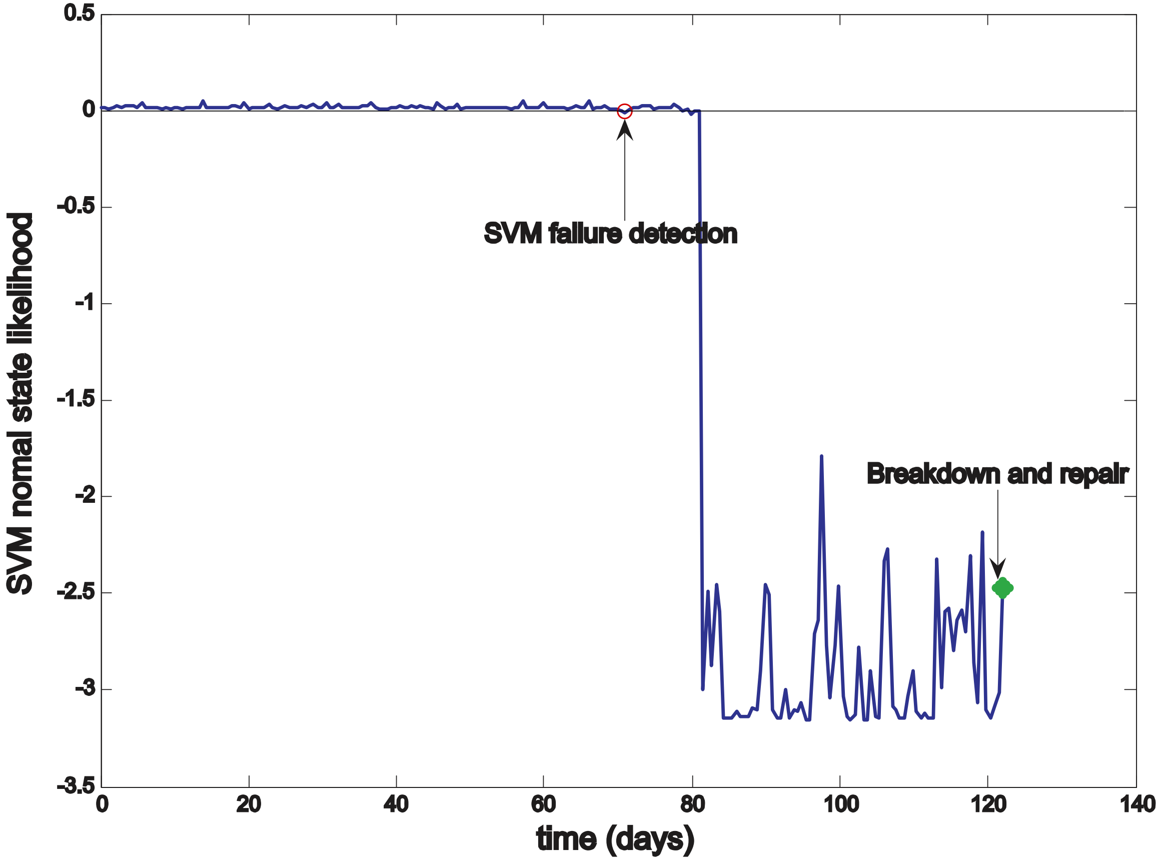
Experimental Results

3) Real Data

- Monitored in a 660KW wind turbine over the duration of March 2010 to July 2010
- Extra input data: revolutions per minute
- 80 captures of normal state used for training

Experimental Results

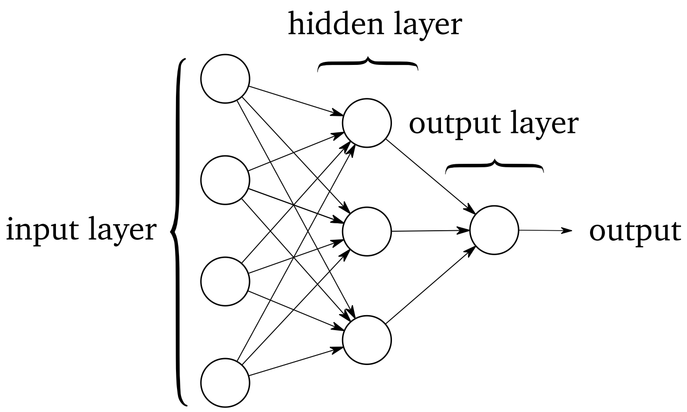
3) Real Data



Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

Comparison to Neural Network Approach

Auto-associative Neural Network



- Feed forward neural network
- Reduction of input dimensionality in hidden layers
- Learns to reconstruct normal state patterns with minimum error
- Reconstruction error at output indicates fault

Comparison to Neural Network Approach

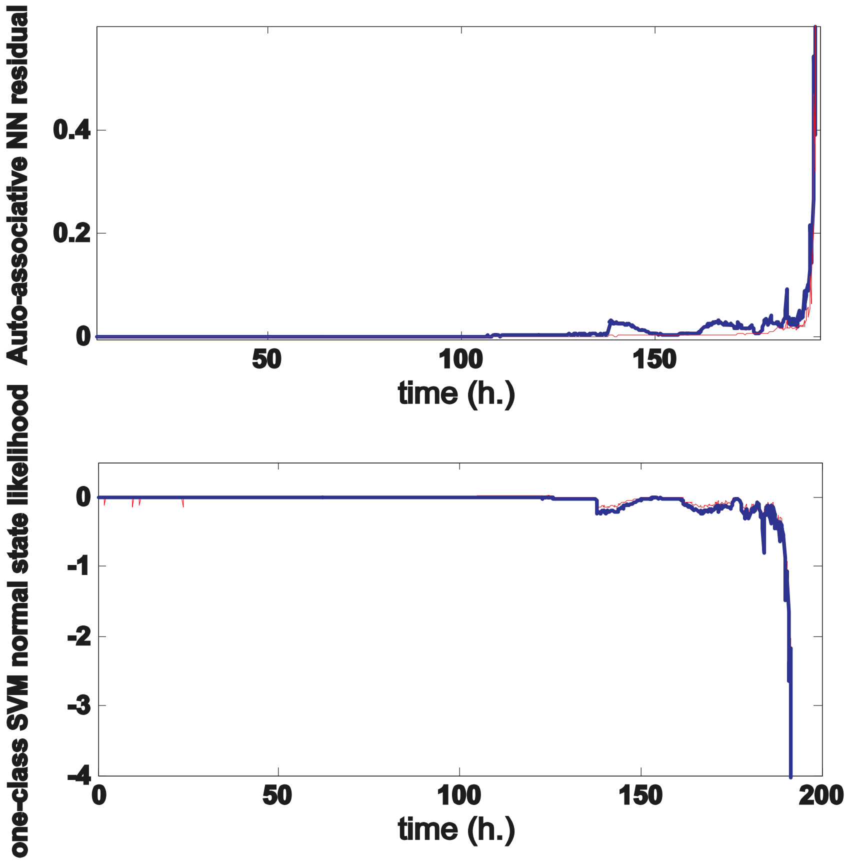
- Comparison using previously tested data
  1. Data generated from laboratory setting
  2. Data generated from simulation

- Trained network structures:
  - Two hidden nodes up to “minimum input space reduction”
  - Results presented using best behaviour

- Comparison with and without extra generated noise

Comparison to Neural Network Approach

1) Results generated from Laboratory data



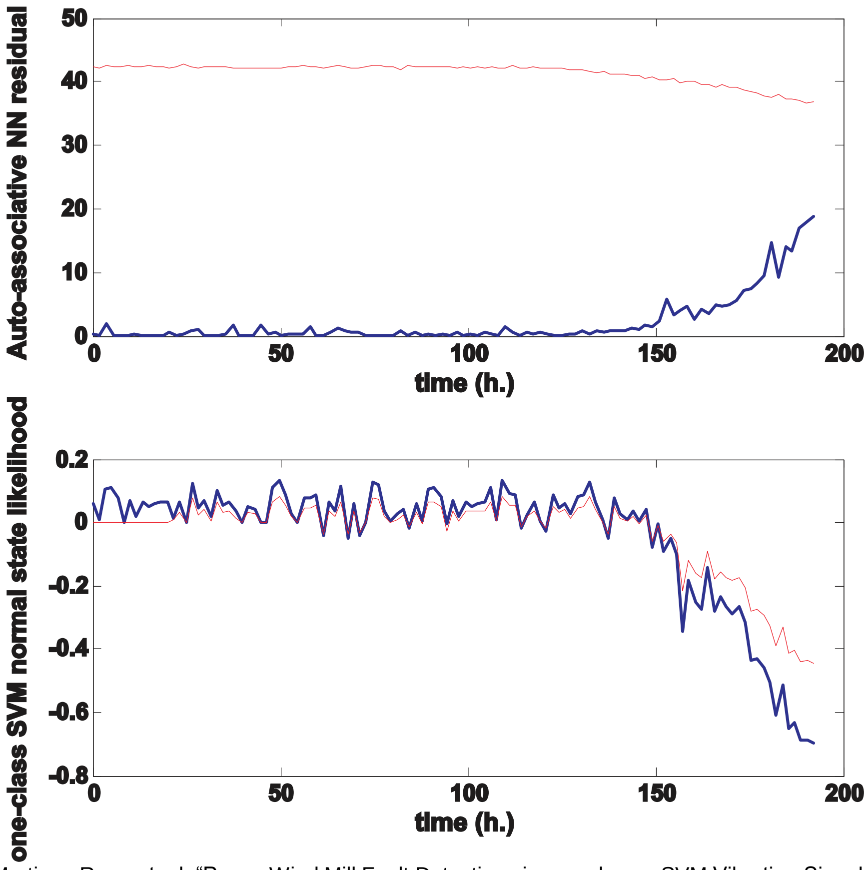
- Blue line represents same conditions as before
- Red line relates to data including additional 2% of data captured under fault

- Similar results, but ANN delays detection of fault until last stage

Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

Comparison to Neural Network Approach

2) Results generated from simulated data



- Blue line represents same conditions as before
- Red line relates to data including additional 5% of noisy patterns following a normal distribution

- ANN shows improper behaviour
- One-class SVM maintains detection accuracy

Martinez-Rego et. al. “Power Wind Mill Fault Detection via one-class v-SVM Vibration Signal Analysis”

Comparison to Neural Network Approach

Conclusions:

- Successful application beyond laboratory analysis
- One-class SVM comply with requirements
- Qualitative assessment of failure and evolution thereof
- One-class SVM proved to be more robust under noise than ANN

Review

TECHNISCHE UNIVERSITÄT DARMSTADT

- Results:
  - Successful novelty detection using one-class SVM
  - Missing more detailed explanations
    - Model parameters – SVM are quite sensitive to this
    - Exact decision procedure
  - Evolution of failure?

- ANN Comparison
  - Missing detailed information
    - Exact network structure
    - Training details

Larry M. Manevitz et al. “One-Class SVMs for Document Classification”
“The SVM approach as represented by Schölkopf was superior to all the methods except the neural network one, where it was, although occasionally worse, essentially comparable. However, the ***SVM methods turned out to be quite sensitive to the choice of representation and kernel*** in ways which are not well understood; therefore, for the time being leaving the ***neural network approach as the most robust***.”