PO.148a

Scalable systems for early fault detection in wind turbines: A data driven approach

Martin Bach-Andersen[1,2], Bo Rømer-Odgaard[1] and Ole Winther[2]
[1]Siemens Diagnostic Center
[2]Technical University of Denmark

Abstract

In order to leverage the complex data streams from modern wind turbines fully in large scale monitoring platforms, further automatization is required. This paper addresses this issue first by establishing a learning framework on which algorithms can learn from the diagnostic process performed by an experienced human expert, and second, establishing a structured methodology for evaluating the diagnostic performance of such algorithms. Building on this work and vibration data from 80 actual main bearing failures, deep networks are demonstrated to outperform shallow counterparts and rival human expert performance for this complex diagnostic task.

The robustness of these methods make them suitable candidates for implementation in large scale monitoring platforms, and the study will provide a solid stepping stone for further research into data driven turbine diagnostics.

Objectives

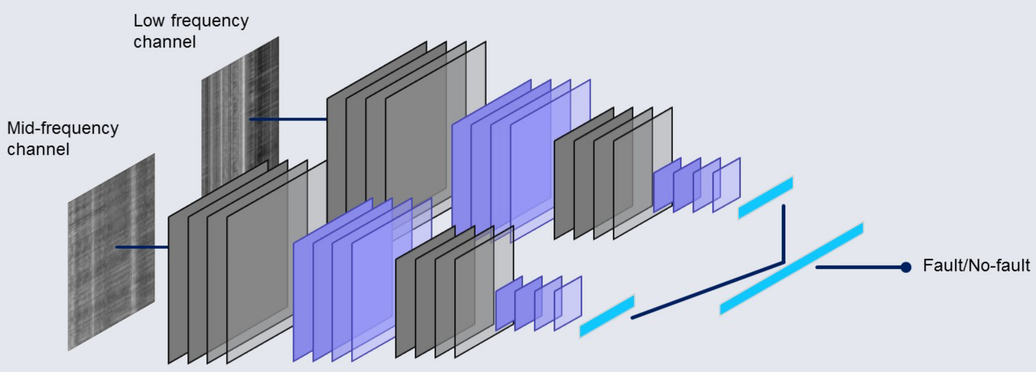
- To facilitate learning of data driven diagnostic models based on actions and outputs from a human diagnostic expert.
- To provide a structured methodology for evaluating the performance of such models.
- To demonstrate the diagnostic capabilities of shallow and deep learning archictectures for a main bearing diagnostic application.

Methods

- Vibration data was collected from 80 turbines with known main bearing failures. A diagnostic expert was tasked with specifying the point of change from no-fault to fault conditions using the full time history. This point in time is denoted the Expert Retrospective Changepoint (ERC).
- 6 month data windows containing low-frequency and mid-frequency autospectra from a single sensor was randomly collected around the ERC as fault-state data points. Each data point is then a pair of 100x198 and 100x1342 spectral images.
- Data set split: 40 turbines for training, 15 turbines for validation and 25 turbines for testing.
- Model 1: A logistic regression model [1] where the posterior fault probability given input **x** is modeled as

$$p(\text{Fault}|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T\boldsymbol{x}) = \frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}}}$$

- Model 2: A deep convolutional network [2] using a number of nonlinear processing layers inspired by the biological system of the visual cortex:



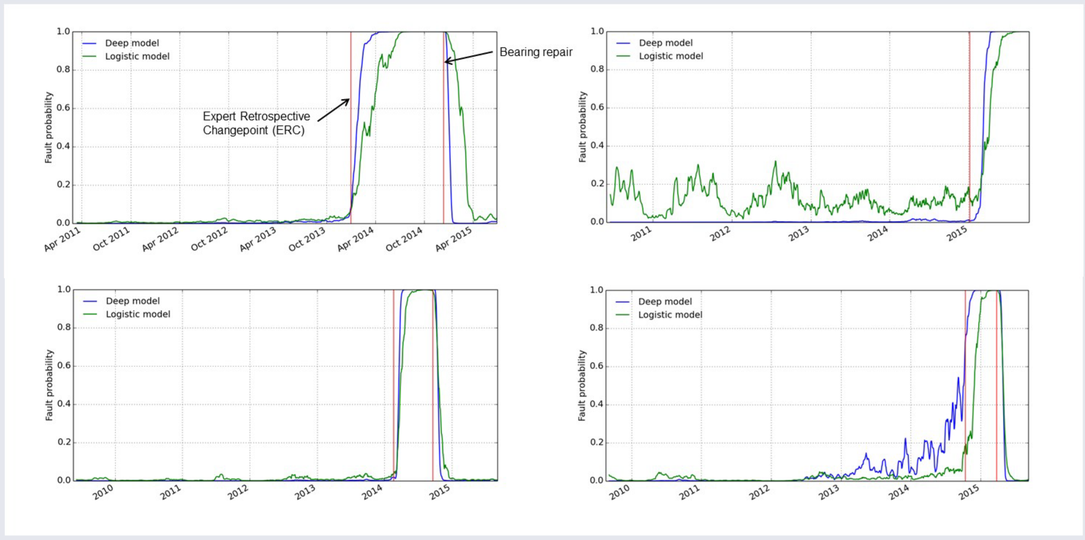
Schematic of the deep learning architecture employed in this study. Convolutional layers are marked in grey, max-pooling layers in purple and fully connected layers in blue. Individual recognition networks are assigned to each input image before being merged in two final fully connected layers.

Architecture for mid-frequency channel: 16 kernels of size 32x4, 3x2 pooling, 32 kernels of size 16x4, 3x2 pooling and a fully connected layer with 1024 units.
Architecture for low-frequency channel: 16 kernels of size 32x4, 4x2 pooling, 32 kernels of size 16x4, 4x2 pooling and a fully connected layer with 1024 units.
Individual networks are joined using a fully connected layer with 512 units and a sigmoidal output.

Except for the output unit, a ReLu-nonlinearities was applied to all units and overfitting was controlled by using dropout on all fully connected layers during training. The network contains more than 67M parameters.

Results

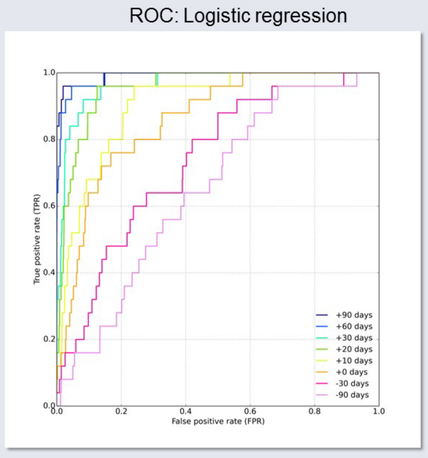
Examples of posterior fault probabilities for test turbines:



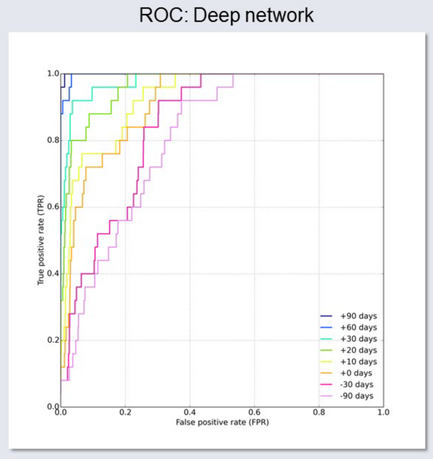
- Models generally demonstrate excellent separation of fault and no-fault states, but the output from the logistic model during no-fault states is more noisy.
- The deep network is more certain of its fault predictions at an earlier stage compared to logistic model.
- In the last case the deep network seemingly predicts an impending fault in an online setting more than a year earlier than a human expert is able to do in a retrospective setting.

How do the diagnostic models perform on entire test dataset? ROC and AUC metrics measured *relative* to the ERC provides an answer:

ROC: Logistic regression

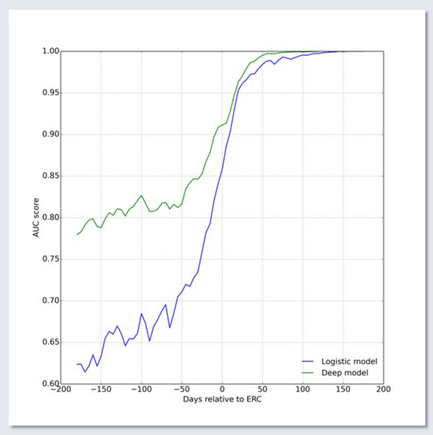


ROC: Deep network



- Receiver Operating Characteristics [3] define classifier performance in terms of true positive rate and false positive rates.
- AUC scores provides a condensed view of classifier performance measured relative to the ERC.
- The deep network outperforms the shallow model throughout entire relative time range: Higher detection rate with less false positives.

AUC scores



Conclusions

- Both shallow and deep learning architectures were succesfully trained to perform accurate detection of main bearing faults based on complex vibration signatures.
- Deep networks provided the earliest and most robust detection with an online performance that rivals human expert performance.
- The detection models and training procedures presented are very general in nature and the models provides probabilistic outputs suitable for decision support systems and direct implementation in large scale monitoring platforms..

References

1. C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
2. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
3. T. Fawcett, “Roc graphs: Notes and practical considerations for researchers,” *Machine learning*, vol. 31, pp. 1–38, 2004.