

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

Martin Bach-Andersen^{1,2}, Bo Rømer-Odgaard¹, and Ole Winther²

¹Siemens Diagnostic Center, Denmark

²Cognitive Systems, Technical University of Denmark

Presenting author: Martin Bach-Andersen

1 Introduction

The world has changed dramatically for wind farm operators and service providers in the last decade. Organizations whose turbine portfolios was counted in 10-100s ten years ago are now managing large scale operation and service programs for fleet sizes well above one thousand turbines. A big challenge such organizations now face is the question of how the massive amount of operational data that are generated by large fleets are effectively managed and how value is gained from the data. A particular hard challenge is the handling of data streams collected from advanced condition monitoring systems. These data are highly complex and typically require expert knowledge to interpret correctly resulting in poor scalability when moving to large Operation and Maintenance (O&M) platforms.

In this paper we present a purely data driven fault detection method that has the potential to vastly improve the scalability of fleet wide condition monitoring systems as much of the complex diagnostic process can be put in algorithmic form. A feasibility study has been performed on data from 61 actual main bearing failures on both on- an offshore turbines. The study will provide a solid stepping stone for further research into data driven turbine diagnostics, but will also provide the diagnostic performance metrics required for practitioners that may wish to implement such technology in a large scale monitoring setting.

2 Data driven fault detection

Recent reviews of condition monitoring applications in wind turbines can be found in [1], [2] and [3]. The diagnostic task we focus on in this work is the detection of spalling on the inner- and outer raceways of the main bearing of geared turbines. Vibration based diagnostics [4] is a mature technology for monitoring turbine drive trains and here we will restrict ourselves to vibration-based fault detection.

Two fundamentally different approaches exist for developing fault detection systems. One is a model-based approach typically using first principles. The second, is a data driven approach where the configuration of the diagnostic system is inferred from data. Data driven approaches for turbine diagnostics are typically unsupervised in the sense that models are fitted to no-fault conditions whereafter faults are identified as deviations from this model of normality [5] [6] [7] [8] [9] [10] [11]. When moving to large scale monitoring platforms, a sufficient number of actual failure data is now available to approach the diagnostic problem using fully supervised classification, where models are trained to distinguish between the fault/no-fault state in the dichotomous case. This is the approach we adopt here.

In order to frame the learning process we start by recognizing that state-of-the-art large scale vibration monitoring systems are based on a suite of automated outlier detection algorithms to generate warnings of potential faults, but the end diagnosis still heavily relies on human interaction as experts sift through the available data to provide early fault detection. Inspired by the diagnostic process performed by these analysts we will therefore explore whether the task can be fully or partly automated using deep and shallow learning architectures.

3 Methods

By observing how an experienced vibration analysis expert solves the task of detecting faults on a main bearing, a number of key elements was identified in this process, namely 1) feature selection, 2) time frame selection and 3) learnings from past observed faults. We will address these elements as follows:

Features A subset of two measurements was selected for this study, namely low-frequency acceleration autospectra (10-62 Hz) and low to mid-frequency acceleration autospectra (10-1000 Hz) from a single sensor mounted on the bearing housing. It is important to stress that we only perform feature selection on this high level. No attempts will be made to reduce these data to low dimensional diagnostic features using hand-crafted failure signal models or unsupervised methods. The aim is to have the models infer features of diagnostic relevance automatically from the high level feature set.

Time frame Based on input from the expert, a time window spanning back six months for both measurements was selected as input.

Experience Full vibration data records were collected from 61 turbines with known main bearing faults. The model was trained to detect faults using this data set as background knowledge.

3.1 Data augmentation and processing

The full data set was split in three: 1) A training set containing data from 49 turbines used for training the models. 2) a validation set containing data from 6 turbines used for monitoring the generalization properties of the model during training and finally 3) a test set of 6 turbines which was used to evaluate the performance of the models performing best on the validation set.

For each of the selected turbines, the expert was tasked with specifying the change point from fault to no-fault condition using whatever data source he or she had available. This will be named as the Expert Retrospective Changepoint (ERC), as it is inferred from the full time record. In an online setting, the time of detection will generally be later compared to the ERC as more data will have to be observed before an accurate detection can be made. From this specified ERC for each turbine, a number of random time windows was extracted around this date, with random width scaling factors. This data augmentation scheme has been implemented to increase the generalization properties of the models by adding a stochastic term to the ERC as well as a stochastic term to the failure rate. The available vibration data was not sampled equidistantly, so the data was further processed using linear interpolation on an uniform time grid with 100 samples in the time dimension. The total number of input dimensions in a given window for both channels totals $100 \times 2684 + 100 \times 396 = 30800$.

3.2 Models

The first model to be considered is a logistic regression model [12], where the posterior fault probability is parameterized as

$$p(\text{Fault}|\mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x}), \quad (1)$$

where σ is the sigmoid function, \mathbf{w} is the parameters of the model and \mathbf{x} is the input data vector.

As recent work on deep learning structures [13][14] has provided state-of-the-art results on complex image recognition tasks, a deep learning convolutional fault recognition model was implemented to investigate if a deeper architecture could provide increased diagnostic performance in the given learning setting. The deep network is composed of two distinct recognition channels as depicted in 1. Rectified linear units were used for all hidden units and the networks were trained using AdaGrad [15]

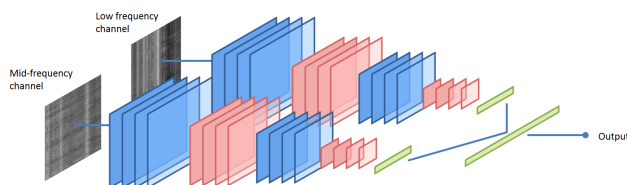


Figure 1: Schematic of the deep convolutional network investigated in this work. Convolutional layers are marked in blue, max-pooling layers are marked in red and fully connected layers are marked in green. Only four kernels are shown for each convolutional layer, the actual network uses 32 kernels.

4 Results

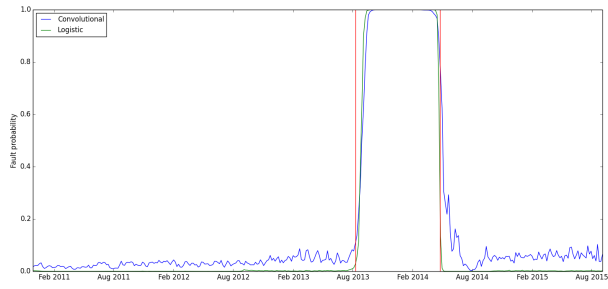
From the training and validation stage the best performing models were tested on the test set of turbines. The results are listed in table 1. Both models provide high accuracies on the test set, but the deep network performs better measured both in terms of cross entropy error and accuracy. These results indicate that both models generalize well to new observations. A detailed view of the fault probability output from the two models have been plotted in figures 2a and 2b for two of the test turbines. These plots show a strong signal separation from the no-fault to the no-fault state and back again when the bearing is replaced.

When implementing such model in large scale monitoring platforms it is not the performance on individual failed turbines that is most relevant, but the performance across the entire fleet, including the turbines in no-fault states. We therefore evaluate the models, based on the data from all 61 turbines, in terms of Receiver Operating Characteristics (ROC) [16]. These curves will be heavily dependent on how early the fault must be detected. Again using the ERC as a fixed reference point, ROC-curves can be generated for each classifier using a sliding window approach. The results can be observed in figures 3a and 3b. When moving beyond +30 days relative to the ERC, both models become perfect classifiers. These results indicate that the better fit of the deep network observed on the test set does not translate directly into better classifier performance.

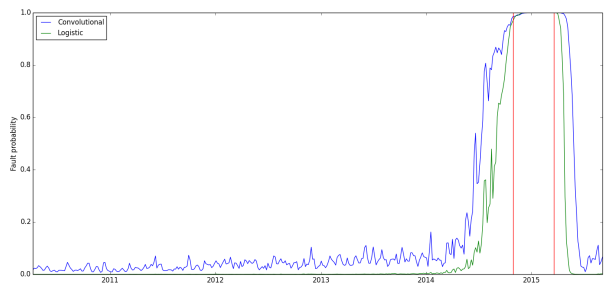
The overall classifier performance as a function of required detection lead time can be condensed to a single value, namely the Area-Under-Curve (AUC) which is the area below the given ROC-curve. The AUC-values are plotted in figure 4 and show a marginal better performance by the logistic model in the later stages of failure development, and marginally better performance by the deep model during the very early stages.

	Average test error	Test accuracy
Logistic	0.118	94.60%
Convolutional network	0.084	97.02%

Table 1: Results on test set specified in terms of cross entropy error and detection accuracy.

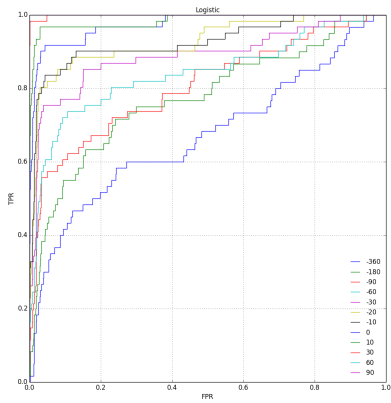


(a)

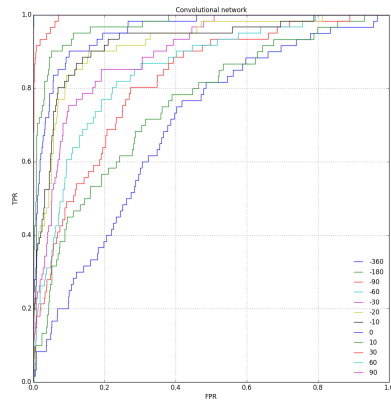


(b)

Figure 2: Examples of the predictions from the convolutional and logistic model for two of the turbines in the test set. The ERC and the component exchange date has been marked with red vertical lines. A clear indication of a fault can be observed in all cases, as well as the return to a no-fault condition when the component is exchanged.



(a)



(b)

Figure 3: Receiver Operating Characteristics for logistic regression (left) and deep convolutional network (right).

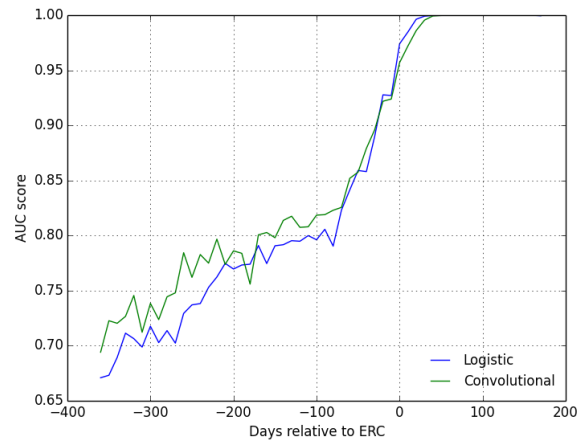


Figure 4: Area Under Curve plots based on detection lead time dependent ROCs.

5 Conclusion

The conclusions to be drawn from this study are as follows:

- A structured learning setting was established on which data driven fault detection methods can learn from the output of an experienced vibration analyst.
- Building on this framework, impressive results were demonstrated for a main bearing fault detection application using both shallow and deep learning architectures:
 - The fault detection models show clear fault/no-fault state separation on turbines with main bearing failures.
 - From the large failure data set used for this study, the performance of the models in large scale settings was evaluated using metrics such as Receiver Operating Characteristics. At detection lead times near those of a human expert, the models provides high true positive rates and low false positive rates.
 - The demonstrated methods are easily scalable to large turbine fleets. The good performance of the logistic model coupled with a simple training schedule makes it the first choice for any practical application.

6 Learning objectives

Although the handling of the large data streams collected from wind turbine fleets of today and tomorrow might seem challenging at first, this study shows that many opportunities can also be created from this data deluge using data driven methodologies. It was demonstrated how state-of-the-art results were obtained in a fault detection application without any detailed knowledge of the underlying mechanical system, by solely relying on the specification of some high level features and then adapting from the analytical process performed by a human expert.

References

- [1] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Proceedings of Power Electronics and Machines in Wind Applications 2009*, IEEE, June 2009.
- [2] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated scada data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009.
- [3] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaalias, "Condition monitoring of wind turbines: Techniques and methods," *Renewable Energy*, vol. 46, pp. 169–178, 2012.
- [4] R. B. Randall, *Vibration-based Condition Monitoring: Industrial, aerospace and automotive applications*. Wiley, 2011.
- [5] M. Schlechtingen and I. Ferreira Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1849–1875, 2011.
- [6] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renewable Energy*, vol. 48, pp. 110–116, 2012.
- [7] P. Guo, D. Infield, and X. Yang, "Wind turbine generator condition-monitoring using temperature trend analysis," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 1, pp. 124–133, 2012.
- [8] Z.-Y. Zhang and K.-S. Wang, "Wind turbine fault detection based on scada data analysis using ann," *Advances in Manufacturing*, vol. 2, no. 1, pp. 70–78, 2014.
- [9] P. Cross and X. Ma, "Model-based condition monitoring for wind turbines," in *Automation and Computing (ICAC), 2013 19th International Conference on*, pp. 1–7, IEEE, 2013.
- [10] N. Talebi, M. A. Sadrnia, and A. Darabi, "Robust fault detection of wind energy conversion systems based on dynamic neural networks," *Computational intelligence and neuroscience*, vol. 2014, 2014.
- [11] S. Butler, F. O'Connor, D. Farren, and J. V. Ringwood, "A feasibility study into prognostics for the main bearing of a wind turbine," in *Proceedings of IEEE International Conference on Control Applications 2012*, pp. 1092 – 1097, IEEE, October 2012.
- [12] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [13] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *CoRR*, vol. abs/1207.0580, 2012.
- [14] 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.

- [15] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for on-line learning and stochastic optimization,” *The Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.
- [16] T. Fawcett, “Roc graphs: Notes and practical considerations for researchers,” *Machine learning*, vol. 31, pp. 1–38, 2004.