Mechanical Systems and Signal Processing 99 (2018) 403–419

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Simulation-driven machine learning: Bearing fault classification

Cameron Sobie *, Carina Freitas, Mike Nicolai

Siemens Industry Software NV, Interleuvenlaan 68, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 16 January 2017
Received in revised form 16 June 2017
Accepted 19 June 2017

Keywords:
Machine learning
Condition monitoring
Roller bearing
Fault detection
Prognostic health monitoring

ABSTRACT

Increasing the accuracy of mechanical fault detection has the potential to improve system safety and economic performance by minimizing scheduled maintenance and the probability of unexpected system failure. Advances in computational performance have enabled the application of machine learning algorithms across numerous applications including condition monitoring and failure detection. Past applications of machine learning to physical failure have relied explicitly on historical data, which limits the feasibility of this approach to in-service components with extended service histories. Furthermore, recorded failure data is often only valid for the specific circumstances and components for which it was collected. This work directly addresses these challenges for roller bearings with race faults by generating training data using information gained from high resolution simulations of roller bearing dynamics, which is used to train machine learning algorithms that are then validated against four experimental datasets. Several different machine learning methodologies are compared starting from well-established statistical feature-based methods to convolutional neural networks, and a novel application of dynamic time warping (DTW) to bearing fault classification is proposed as a robust, parameter free method for race fault detection.

1. Introduction

Rotating machines are ubiquitous in power production and transportation, among countless other applications. Downtime caused by failures of components such as bearing or gear faults directly reflects on the economic viability of large systems such as wind turbines [1] and industrial machinery [2]. Early warning of faults leading to probable failures can mitigate such risks as well as improve safety. The fatigue lives of common components in rotating systems can be described using widely accepted methods e.g. ISO standards; however, the lifetime of an individual component is inherently stochastic because of the nature of material variability, manufacturing defects, and loading regimes unique to each component. Therefore, lifetime predictions of these components are limited to probability distributions [3,4] that evolve differently for each physical realization of a component. Novel methods such as digital twins [5] aim to reduce this variability by explicitly accounting for the specific loading history on a structure. While this may become effective for larger structures such as aircraft, whose lifetimes are limited to tens of thousands of cycles between maintenance intervals, these methods are impractical for rotating machines with lifetimes from millions to billions of cycles and an early detection strategy is more feasible than a prediction strategy.

* Corresponding author.
E-mail address: cameron.sobie@siemens.com (C. Sobie).

http://dx.doi.org/10.1016/j.ymssp.2017.06.025
0888-3270/© 2017 Published by Elsevier Ltd.

Roller bearing failures constitute the majority cause of failures for industrial electric motors [6] and wind turbine gearboxes [7]; therefore, the early detection of their associated faults has the potential to significantly decrease unplanned downtime. The spatially constrained contact between the rollers and races combined with the time-varying nature of bearing loading drives fatigue damage. Such faults arise on the inner/outer races or roller elements when imperfections on the material surface propagate to form cracks. When these cracks merge, material is lost and pits are formed. This damage process is inherently multi-scale in nature, highly dependent on local material properties and structure, and influenced by multiple stochastic processes. While the microscopic spalling process is relatively well-understood [8], predicting the remaining time until spall formation requires detailed knowledge of the material microstructure and loading conditions; therefore, effort may be better spent on defect detection (and remaining useful life prediction) rather than predicting the formation of such faults. Rather than relying on recorded acceleration signals from defective bearings, physics-based modeling can provide insight into the vibration response of a faulted bearing. The simplest model for the excitation of a damaged bearing is a Dirac delta pulse train with a period associated to that of the defect as proposed by McFadden and Smith [9]. The physical fidelity of the model was improved with pulse modulation by the direction relative to the radial load on the bearing, and the bearing vibration response was assumed to be an under-damped oscillation. Following models proposed improvements such as waviness in the bearing race surfaces [10], introducing random fluctuations about the expected defect period [11], elastohydrodynamic film effects [12], and bearing race flexibility at higher modes and centrifugal load effects [13]. Finite element simulations have also been used to investigate the interactions between a roller element and a race defect [14,15].

Bearing fault detection is most often performed using linear acceleration or acoustic emission signals, and force sensors have been also shown to be effective [16,17]. Motor current signatures are also a proposed method for bearing fault detection [18–20] as bearing faults are a common cause for motor failure. Time-series data can be used to produce features that can differentiate between nominal and faulted bearings with measures such as root mean square (RMS), kurtosis, or skewness. Preprocessing with methods such as bandpassing, calculation of the envelope signal [21], minimum entropy deconvolution [22,17,23], and cepstral methods [17,23] have also been proposed. Frequency domain analysis is a more natural approach because of the periodic nature of rotating machines, and the fact that each bearing fault type (inner/outer race, roller element) has a characteristic frequency (assuming no slipping between the components). The power in the bearing vibration signal around these frequencies can be used to create features, in addition to other typical features such as spectral kurtosis. However, Fourier transform methods are inherently limited to steady state operation and are not well-suited to non-stationary operating conditions without additional processing. Modern time-frequency domain methods have been developed to address this challenge. These methods include the Hilbert-Huang transform [24,25], and discrete [26,27] and continuous [28] wavelet transforms (DWT/CWT). These techniques decompose the signal into a set of signals based on the frequency content. These transforms cannot be performed in closed form in general and the precise manifestation of a bearing fault in a signal processed with these transforms is difficult to predict, and therefore these methods are favored for numerical rather than analytical studies. Apart from transform methods, angle synchronous averaging (ASA) is a commonly used method to emphasize bearing defect signals and reduce noise under stationary and non-stationary regimes [29–33]. From the shaft speed, one can segment a time-series vibration signal into a series of intervals during which the bearing rotates over the angular defect period. Averaging the envelope signals over several of these intervals will isolate a response occurring at the averaging period length while reducing noise. This analysis method is resistant to varying shaft speeds and it includes the bearing geometry such that angle synchronous averages between different bearing geometries can be compared. These analysis techniques enable one to extract quantitative measures of the bearing vibrational signal characteristics; however, diagnosing a bearing fault using these measures, or combinations of them, presents a significant challenge to which machine learning algorithms have proven effective.

Numerous machine learning algorithms have been applied successfully in mechanical engineering applications and more specifically in the field of fault detection, condition monitoring, and prognostics. Contemporary condition monitoring methods using machine learning have required either commissioning experimental setups running under accelerated failure conditions until a fault is observed and measured, or the collection of in-service data (e.g. SCADA) [34] with which to train. At the system level, wind turbine performance and faults (for several applications such as power production prediction [35], pitch [36] and bearings faults [37]) have been heavily analyzed because production wind turbine units have integrated SCADA capabilities and years of data from a large number of turbines has been recorded. Within the domain of bearing diagnostics, experimental data-driven machine learning uses features such as time-series RMS, standard deviation, crest factor [38], and wavelet variance [39] among numerous other statistical measures. Combining all of the aforementioned and additional analysis methods, Rauber et al. [40] created a “feature pool” including time-series as well as frequency domain statistics and the energies of CWT decompositions to successfully classify bearing faults with several different classifiers. More recently, one-dimensional convolutional neural networks applied to the bearing acceleration frequency spectrum have been proposed as a more effective method of bearing fault detections than statistical feature analysis with the additional benefit of removing the need for expert knowledge associated with feature extraction [41]. This method shows promise, but in its proposed form is limited to stationary operating regimes as the input to the convolutional neural network is the power spectrum, which is inherently dependent on the operating speed and bearing configuration. In general, data-driven approaches are limited by the requirement of in-service data and there is little reason to expect that a classifier trained in one configuration would be applicable to a different bearing in a different application. Rather than relying on in-service data to train algorithms, simulation-driven machine learning overcomes this barrier by creating training data that can capture a wide range of operating conditions that can be later combined with in-service fault data applicable to its unique operating conditions as it