Engineering Applications of Artificial Intelligence 64 (2017) 187–196

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Online feature learning for condition monitoring of rotating machinery

S. Martin-del-Campo [a,b,*], F. Sandin [a]

[a] EISLAB, Luleå University of Technology (LTU), 97187, Luleå, Sweden
[b] SKF-LTU University Technology Center, 97187, Luleå, Sweden

A R T I C L E I N F O

Keywords:
Condition monitoring
Dictionary learning
Sparse representation
Dictionary propagation
Anomaly scoring
Bearings

A B S T R A C T

Condition-based maintenance of rotating machinery requires efficient condition monitoring methods that enable early detection of abnormal operational conditions and faults. This is a challenging problem because machines are different and change characteristics over time due to wear and maintenance. The efficiency and scalability of conventional condition monitoring methods are limited by the need for manual analysis and re-configuration. The problem to extract relevant features from condition monitoring signals and thereby detect and analyze changes in such signals is a central issue, which in principle can be addressed using machine learning methods. Former work demonstrates that dictionary learning can be used to automatically derive signal features that characterize different operational conditions and faults of a rotating machine, but the use of such methods for online condition monitoring purposes is an open problem. Here we investigate online learning of features using dictionary learning. We describe dictionary distance and signal fidelity based heuristics for anomaly detection, and we study the time-propagated features and sparse approximation of vibration and acoustic emission signals in three different case studies. We present results of numerical experiments with different hyperparameters affecting the approximation accuracy, computational cost, and the adaptation rate of the learned features. We find that the learned features change slowly under normal variations of the operational conditions in comparison to the rapid adaptation observed when a fault appears (bearing defects, magnetite particles in the lubricant, or plastic deformation of steel). Furthermore, we find that a sparse signal approximation with 2.5% preserved coefficients based on a propagated dictionary is sufficient for bearing defect detection.

1. Introduction

Detection of abnormal operational conditions and prediction of faults are important for the overall efficiency of rotating machines. Condition-based maintenance (CBM) is a maintenance approach that recommends maintenance decisions based on condition monitoring information (Jardine et al., 2006), thereby reducing machine downtime and ensuring proper operation of machines. The CBM approach can be divided in three stages: data acquisition, data processing and decision making. In the data acquisition stage, measured parameters are collected, such as vibration, acoustic emission, current or temperature. Subsequently, in the data processing stage features are extracted from the raw data. Afterward, in the decision making stage, actions are taken based on an analysis of the features. In addition to fault detection and diagnosis, CBM requires methods for prediction of faults (Heng et al., 2009).

Machine learning can be used to automate the diagnosis and prediction tasks. A central issue is the extraction of relevant features from condition monitoring signals. Currently, practice is widely based on manual selection of standard features or features that are defined by experts. This implies that features are defined without explicit knowledge about the configuration, state and evolution of each particular machine.

In this work we investigate a method for online learning of features, which can enable automatic extraction of features that are optimized to each individual machine. The proposed method is based on sparse-coding theory and dictionary learning, which also enables substantial compression of the raw sensor data. The transformation of raw data to succinct information is key for further analysis and resource requirements, for example in systems based on wireless sensors and Internet-of-Things technology, which have limited data processing and communication capacities.

We focus on rolling element bearings and vibration monitoring. Rolling element bearings are essential machine elements used to carry loads and reduce friction between moving parts in rotating machines. Therefore, the condition monitoring of bearings is an important aspect

* Corresponding author at: EISLAB, Luleå University of Technology (LTU), 97187, Luleå, Sweden
E-mail addresses: sergio.martindelcampo@ltu.se (S. Martin-del-Campo), fredrik.sandin@ltu.se (F. Sandin).

http://dx.doi.org/10.1016/j.engappai.2017.06.012
Received 19 October 2016; Received in revised form 29 April 2017; Accepted 16 June 2017

for improving the overall efficiency and reliability of rotating machinery.

Vibration signals from bearings are commonly analyzed in terms of pre-defined frequency-domain or time-domain features. For example, pre-defined frequency bands and threshold values are used for the detection of faults and abnormal conditions (Patidar and Soni, 2013; Tandon and Choudhury, 1999; Yang et al., 2003). The detection, prediction and diagnosis of faults in bearings is a difficult task due to the high number of variables affecting their operation. Defect propagation rates in bearings are stochastic due to the probabilistic nature of the bearing operational conditions and structural integrity (Heng et al., 2009; Randall and Antoni, 2011). Therefore, machine-learning and pattern-recognition methods play a natural role for the further development of automated diagnosis and prognosis systems. In particular, unsupervised methods that enable anomaly ranking and support the analysis of abnormal operational conditions without reference to labeled training data are needed, because it is difficult and expensive to generate suitable datasets for supervised training (Randall and Antoni, 2011).

Sparse representation (Mallat, 2008; Elad, 2010, 2012) and analysis (Nam et al., 2013) of signals is an approach that has attracted wide interest in the last decade. Sparse signal models based on a learned dictionary of features require a minimum of information for modeling and analysis, thereby potentially simplifying the signal processing task. Former work demonstrates that such models can be used to characterize different operational conditions and faults of a rotating machine (Liu et al., 2011; Martin-del-Campo et al., 2013; Martin-del-Campo and Sandin, 2015) and for the detection of anomalies (Tang et al., 2014; Adler et al., 2015).

However, it remains to investigate how dictionary learning can be implemented in an online condition monitoring system for automated detection and scoring of emerging faults. That is the motivation of this paper, which presents an investigation of measures and heuristics for automatic scoring of abnormal operational conditions, which are based on learned features and sparse-coding. We consider vibration signals from a rotating machine with bearings in healthy and faulty states of operation, acoustic emission from elastic and plastic deformation of steel, and acoustic emission from a bearing where contaminant particles are introduced in the lubricant. We investigate different sets of hyperparameters governing the calculation of the sparse code and learning of features. Furthermore, we propose indicators based on the dictionary distance measure of Skretting and Engan (2011), which are used to monitor the propagated dictionary over time for anomaly detection and scoring purposes. In particular, we investigate the effects of varying sparsity and the computational cost on the methods proposed here for scoring of abnormal conditions.

2. Method for sparse coding and feature learning

Sparse signal approximation based on a linear combination of a small number of elementary waveforms selected from a large collection is applicable to a wide range of signals and tasks like compression, detection, separation and denoising (Bruckstein et al., 2009; Elad, 2010; Mallat, 2008; Starck et al., 2010). The general problem of finding the optimal n-term approximation is intractable in practice and several algorithms have been proposed that reduce the computational complexity and provide reasonable approximation results. The general approaches are l_1-norm minimization with iterative convex optimization techniques, and greedy algorithms that iteratively decrease the approximation error with a relaxed sparsity constraint (Donoho, 2006). The method described here belongs to the latter class, and in addition to optimizing the n-term approximation it also optimizes a set of shift-invariant elementary waveforms used for the sparse signal approximation. The resulting shift-invariant waveforms are features of the signal and here we are interested in the correlation between such automatically learned features and the conditions of rotating machines.

2.1. Inspiration from biological sensory systems

We adopt the feature learning approach proposed by Smith and Lewicki (2005, 2006), which is inspired by the work of Olshausen and Field (1996, 1997) on sparse visual coding. A fundamental task of a biological sensory system is to infer information about the environment under resource constraints. Therefore, efficient encoding mechanisms have evolved that reduce the influence of noise and the redundancy of the raw sensory signal. When optimizing sparse representations of speech in this manner the resulting features match cochlear response properties (Smith and Lewicki, 2006), and in the case of vision the resulting features resemble some receptive field properties of cells in the primary visual cortex (V1) (Olshausen and Field, 1996, 1997). We hypothesize that response functions that characterize other systems and source processes can be learned in a similar fashion, and that such automatically learned features are useful for signal analysis purposes. The resulting sparse signal representations are sometimes referred to as “succinct”, meaning that they are both compact and informative. Sparse-coding sensors can potentially be adopted in resource-constrained sensor systems to improve the quality of sensor information and reduce the cost of further processing and communication.

2.2. Sparse signal model

The signal, $x(t)$, is modeled as a linear superposition of waveforms, ϕ_m, with compact support and additive noise

$$x(t) = \sum_{m=1}^{M} \sum_{i=1}^{N_m} a_{m,i} \phi_m(t - \tau_{m,i}) + \epsilon(t). \tag{1}$$

The functions ϕ_m are *atoms* that represent shift-invariant elementary waveforms of the signal and M indicates the total number of such atoms. The term $\epsilon(t)$ represents the model residual, including noise. The variable N_m refers to the number of instances of atom ϕ_m. The temporal position and amplitude of the ith instance of atom ϕ_m are denoted by $\tau_{m,i}$ and $a_{m,i}$, respectively. The set of M atoms defines a dictionary

$$\Phi = \{\phi_1, \ldots, \phi_M\}. \tag{2}$$

Eq. (1) is an inverse problem that is solved with a two-step optimization process for each consecutive signal segment:

1A. *Atom Selection*—Find the atom, ϕ_m, that has the highest cross-correlation with the signal residual and identify the corresponding offset $\tau_{m,i}$.

1B. *Residual Update*—Given the selected atom(s), calculate the $a_{m,i}$ that minimizes the approximation error and update the signal residual accordingly. Repeat from 1A until a stopping condition is fulfilled.

2. *Dictionary Learning*—Given the set of selected atoms and corresponding $\tau_{m,i}$, $a_{m,i}$ and residual, $\epsilon(t)$, update the atoms, ϕ_m, so that the approximation error is reduced.

Steps 1A and 1B is a signal *encoding* process that is repeated until a stopping condition is reached, which typically is defined in terms of the approximation error and/or the number of terms in the approximation. The number of terms in Eq. (1) is related to the *sparsity* of the representation, which is directly related to the number of iterations of steps 1A and 1B in the optimization process. Thus, the model allows for a dynamic trade-off between computational cost and representation accuracy. The second step, dictionary-learning, is performed after the encoding process. Thereafter the optimization process restarts with the next segment of the signal. By processing partially overlapping segments in this fashion, a continuous signal can be encoded online in terms of continuously learned atoms (Martin-del-Campo et al., 2013). The mathematical details of this process are outlined below.