Mechatronics 24 (2014) 1032–1041

Contents lists available at ScienceDirect

Mechatronics

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

A data-driven approach to diagnostics of repetitive processes in the distribution domain – Applications to gearbox diagnostics in industrial robots and rotating machines

André Carvalho Bittencourt [a,*], Kari Saarinen [b], Shiva Sander-Tavallaey [b], Svante Gunnarsson [a], Mikael Norrlöf [a,c]

[a] Department of Electrical Engineering, Linköpings University, Linköping, Sweden
[b] ABB Corporate Research, Västerås, Sweden
[c] ABB Robotics, Västerås, Sweden

ARTICLE INFO

Article history:
Received 1 February 2013
Accepted 27 January 2014
Available online 28 February 2014

Keywords:
Repetitive processes
Diagnostics
Fault detection and isolation
Wear
Vibration analysis
Gearbox

ABSTRACT

This paper presents a data-driven approach to diagnostics of systems that operate in a repetitive manner. Considering that data batches collected from a repetitive operation will be similar unless in the presence of an abnormality, a condition change is inferred by comparing the monitored data against an available nominal batch. The method proposed considers the comparison of data in the distribution domain, which reveals information of the data amplitude. This is achieved with the use of kernel density estimates and the Kullback–Leibler distance. To decrease sensitivity to disturbances while increasing sensitivity to faults, the use of a weighting vector is suggested which is chosen based on a labeled dataset. The framework is simple to implement and can be used without process interruption, in a batch manner. The approach is demonstrated with successful experimental and simulation applications to wear diagnostics in an industrial robot gearbox and for diagnostics of gear faults in a rotating machine.

1. Introduction

In the manufacturing industry, preventive scheduled maintenance is a common approach used to improve equipment's safety, reliability, availability and maintainability. This setup delivers high availability, reducing operational costs (e.g. small downtimes) with the drawback of high maintenance costs since unnecessary maintenance actions might take place. Condition based maintenance (CBM), "maintenance when required", can deliver a good compromise between maintenance and operational costs, reducing the overall cost of maintenance. The extra challenge of CBM is to define methods to determine the condition of the equipment. This can be done by comparing the observed and expected (known) behaviors of the system through an algorithm. The output of such algorithm is a quantity sensitive to a fault,[1] i.e. a *fault indicator*, which can be monitored to determine the current state of the system (e.g. healthy/broken).

A common approach to generate fault indicators is based on the use of residuals, i.e. fault indicators that are achieved based on deviations between measurements and the output of a *system model*, see e.g. [25,27]. A system model is a map from input to output data. A system model provides important information about the behavior of the system, facilitating the generation of fault indicators. Different approaches for residual generation are based on, e.g., observers, parity-space and parameter identification. When a model of the system is not available or it is too costly to be developed, alternatives are still possible. These alternatives will typically require extra (redundant) sensory information or expert knowledge about the measured data, e.g., their nominal frequency content or the use of *labeled data*. Essentially, however, any method will attempt to generate quantities that can be used to infer the actual condition of the system given the available knowledge and observations, i.e. data.

* Corresponding author. Tel.: +46 13 282636.
E-mail addresses: andrecb@isy.liu.se (A.C. Bittencourt), kari.saarinen@se.abb.com (K. Saarinen), shiva.sander-tavallaey@se.abb.com (S. Sander-Tavallaey), svante@isy.liu.se (S. Gunnarsson), mikael.norrlof@se.abb.com (M. Norrlöf).
URL: http://www.control.isy.liu.se/andrecb (A.C. Bittencourt).

[1] A fault is defined as a deviation of at least one characteristic property of the system from the acceptable/usual/nominal condition.

http://dx.doi.org/10.1016/j.mechatronics.2014.01.013
0957-4158/

The use of model-based approaches is common for diagnostics of machines. For robotics, many approaches have been suggested based on the use of nonlinear observers, where the observer stability is typically guaranteed by analyses of the decay rate of a candidate Lyapunov function, see e.g. [9,10,12,15,20,22,30]. Observers can also be designed based only on data, without a description of the system based on first principles. Data-driven design of observers are typically based on subspace identification of linear models and have been suggested for fault detection in [14,16,42,44]. Parameter estimation is also a natural approach to model-based diagnostics because of the physical interpretation of the system parameters, see e.g. [5,21,29].

In cases where the data are ordered with time, signal-driven methods are common for machinery diagnostics. These are typically based on the use of integral transforms, e.g. Fourier, Radon, Karhunen–Loéve or Wavelet. Each transform will enhance different properties in the transformed domain and are suitable depending on the characteristics of the signal, e.g. periodic, stationary, etc. The analysis of data in the frequency domain or time–frequency has found particular success in the monitoring of rotating machines, see e.g. [13,19,23–25,38,40]. Some approaches have also been proposed for the diagnostics of industrial robots with the use of additional sensory information [18,32].

A common challenge to data-driven methods is that the data characteristics' will vary depending on the operating points, which may complicate determination of fault presence. This is particularly restricting for an industrial robot where the kinematic configuration of the robot may give varying load torques at the joints during motion. This shortcoming can be circumvented by considering data from a specific operation of the system, e.g. under repetition. A *repetitive operation* is found in various applications, e.g. in automated manufacturing. Repetition can also be forced with the execution of specific diagnostic routines but with the drawback of reduced availability. Much attention has been given recently to repetitive processes [36,37]. Study of repetitive processes have mainly focused on control [35,39] and estimation problems [1,2]. A few approaches have been also suggested for model-based diagnostics, e.g. [43].

In this paper, a data-driven method is proposed for the generation of fault indicators for systems that operate in a repetitive manner. It is considered that in case the condition of the system is nominal, data batches collected from repetitive executions of the system will be similar to each other and will differ if the condition changes. The comparison of a given data batch against a nominal one can thus be used to infer whether an abnormality is present. The fault indicator proposed here relates to changes in the distribution of these batches of data. This is made possible with the use of kernel density estimators and the Kullback–Leibler distance between distributions. A distribution domain approach does not consider the dynamics of the system generating the data as is the case in, e.g., observer-based approaches. As it will be presented, this leads to very simple diagnostics solutions that can perform well in practical setups.

The proposed framework was initially developed with the interest focused on the diagnostics of wear in industrial robots and a preliminary version of the work can be found in [7]. Here, more aspects are covered, including approaches to detection, isolation and reduction of sensitivity to disturbances. Also, more experimental and simulation results are presented for the robotics application. An additional application is also included for diagnostics of rotating machinery based on vibration data collected from an accelerometer. The paper is organized as follows; a general presentation of data-driven diagnostics and repetitive systems is given in Section 2, followed by the presentation of the proposed approach for diagnostics in the distribution domain in Section 3. The applications are presented in Sections 4 and 5. Conclusions and future work are given in Section 6.

2. Data-driven diagnostics and repetitive systems

Consider a general system from which it is possible to extract a sequence of data batches,

$$\mathbf{Y}_K = [\mathbf{y}_1, \cdots, \mathbf{y}_k, \cdots, \mathbf{y}_K], \tag{1}$$

where $\mathbf{y}_k = [y_{k,1}, \cdots, y_{k,n}, \cdots, y_{k,N}]^T$ denotes the data vector in $\mathbb{R}^N$ (e.g. measurements or known inputs) with batch index k and element index n. The sequence $\mathbf{y}_k$ could have been generated as the result of deterministic and stochastic inputs, $\mathbf{Z}_K$ and $\mathbf{V}_K$, where $\mathbf{V}_K$ is unknown, and $\mathbf{Z}_K$ may have known and unknown components. For example, the data generation mechanism could be modeled as

$$\mathbf{y}_k = h(\mathbf{z}_k, \mathbf{v}_k), \tag{2}$$

where $h(\cdot)$ is an unknown function. Let the set of deterministic inputs $\mathbf{Z}_K$ be categorized in three distinct groups, $\mathbf{R}_K$, $\mathbf{D}_K$ and $\mathbf{F}_K$. The sequence $\mathbf{f}_k$ is unknown and of interest (a fault), $\mathbf{r}_k$ is known (e.g. references or control inputs) and $\mathbf{d}_k$ is unknown (disturbances).

The objective is to define a data-driven framework for the generation of fault indicators to determine the presence of a fault $\mathbf{f}_k$. Because a data-driven approach depends on availability of data, it is assumed that data generated under no fault is available. Let $\mathcal{Y}^0 = \{\mathbf{y}_k : \mathbf{f}_k = \mathbf{0}\}$ denote the set of data batches that were generated under no fault, the following assumption is made:

A-1 (**Nominal data are available**) A sequence $\mathbf{y}^0 \in \mathcal{Y}^0$ is available.

The rationale is then to generate fault indicators from the comparison of the nominal data $\mathbf{y}^0$ (available from Assumption A-1) against the remaining sequences $\mathbf{y}_k$. In order to generate fault indicators for $\mathbf{y}_k$ using the nominal data $\mathbf{y}^0$, two basic questions arise:

Q-1 How to characterize a sequence $\mathbf{y}_k$?
Q-2 How to compare the sequences $\mathbf{y}^0$, $\mathbf{y}_k$?

The first question targets the issue of finding a data processing mechanism of $\mathbf{y}_k$, written in a general form as $g_k \triangleq g(\mathbf{y}_k) : \mathbb{R}^N \mapsto \mathbb{G}$ with domain $\mathbb{G}$, whose output enhances the ability to discriminate the presence of non-zero $\mathbf{f}_k$. Given the nominal data in the transformed domain $g^0 \triangleq g(\mathbf{y}^0)$, fault indicators can be achieved from the comparison between g^0 and g_k. This is typically, but not necessarily, done with the use of a distance function represented as $d(g^0, g_k) : \mathbb{G} \times \mathbb{G} \mapsto \mathbb{R}_0^+$. Different distances are possible depending on the domain $\mathbb{G}$. For example, for diagnostics of rotating machines g_k could be the spectra of $\mathbf{y}_k$ and $d(\cdot,\cdot)$ a spectral distance, see e.g. [3].

2.1. Detection, performance and isolation

Let $\mathcal{D}^{m,n} = \{d(g_i, g_j) : \mathbf{y}_i \in \mathcal{Y}^m, \mathbf{y}_j \in \mathcal{Y}^n\}$, then $\mathcal{D}^{0,0}$ describes the behavior of the fault indicator when no fault is present and $\mathcal{D}^{0,f}$, where $\mathcal{Y}^f = \{\mathbf{y}_k : \mathbf{f}_k \neq \mathbf{0}\}$, describes all possible faulty behaviors. A criterion for **detectability** of an abnormality is that $\mathcal{D}^{0,f}$ is not completely contained in $\mathcal{D}^{0,0}$, i.e. $\mathcal{D}^{0,f} \setminus \mathcal{D}^{0,0} \neq \emptyset$. Since the distance $d(g^0, g_k)$ measures how far g_k is from the nominal g^0, it is expected that it will remain close to zero if $d(g^0, g_k) \in \mathcal{D}^{0,0}$ and to deviate to positive values if $d(g^0, g_k) \in \mathcal{D}^{0,f}$. Suppose that it is possible to find a threshold $\hbar$ such that $d(g^0, g_k) \leqslant \hbar$ most of the times when $d(g^0, g_k) \in \mathcal{D}^{0,0}$, a simple criterion for **detection** is then to consider a threshold check. Let $\mathcal{H}^0$ denote the hypothesis that no fault is