Sensorless drive diagnosis using automated feature extraction, significance ranking and reduction

Conference Paper · September 2013

DOI: 10.1109/ETFA.2013.6648126

6 authors, including:

Olaf Enge-Rosenblatt
Fraunhofer Institute for Integrated Circuits IIS, Division Engineering of Adaptive Syst…

Martyna Bator
Technische Hochschule Ostwestfalen-Lippe University of Applied Sciences and Arts

Uwe Mönks
coverno GmbH

Volker Lohweg
Technische Hochschule Ostwestfalen-Lippe

Sensorless Drive Diagnosis Using
Automated Feature Extraction, Significance Ranking and Reduction

Christian Bayer, Olaf Enge-Rosenblatt
Fraunhofer Institute for Integrated Circuits IIS
Design Automation Division EAS
Dresden, Germany
christian.bayer@eas.iis.fraunhofer.de

Martyna Bator, Uwe Mönks,
Alexander Dicks, Volker Lohweg
inIT – Institute Industrial IT
Lemgo, Germany
martyna.bator@hs-owl.de

Work in Progress Contribution

Abstract

Systems for process automation become increasingly complex and also tend to be composed of autonomous sub-systems, which is strongly driven by the progress made in information technology. An active field of research is the implementation of monitoring and control at sub-system level using cognitive approaches. In this paper we present a method for autonomous and sensorless condition monitoring of an electric drive train. Based on experiment design we measured phase currents of a physical demonstrator device including mechanical defects and extracted signal features using proper orthogonal decomposition. In favor of classification of different defect states we performed a linear discriminant analysis, which yields appropriate data for a Fuzzy-Pattern-Classification algorithm. As a result we were able to identify different reference defect states as well as previously unknown states.

1 Introduction

A worldwide trend to increasingly complex systems for process automation, strongly driven by computer science and information technology (IT) can be observed. Due to the wide range of topics of the tasks to be faced it is not surprising that solutions are often treated separately, thus causing complexity problems. In many cases, cognitive approaches are therefore used to establish autonomous systems. However, holistic concepts and implementations of such systems for process automation and production engineering which secure, control and monitor complex installations are still in the initial stage of research and development activities [4]. The important reduction of complexity can be achieved by (partial) autonomy of the systems. In this context the results presented in this paper are related to and part of the strategies of the initiatives [2, 5].

In this paper we present a method to perform sensorless diagnosis of an autonomous electric drive train. The unit under investigation is composed of a synchronous motor and several attached components, e. g. bearings, axles, and a gear box. It, thus, represents a typical and crucial component within a plant or other machinery. Damage to the drive may cause severe disturbances and increases the risk of encountering breakdown costs. Condition monitoring for such applications usually requires additional sensors. Our approach instead is to directly use the phase currents of the motor for health state characterization of the entire unit. The measurement does not involve any sensors apart from the built-in current sensors of the motor control. In this respect the diagnosis performs autonomously.

Section 2 describes the experimental setup of the drive train. Since it has been built for the development of condition monitoring methods, we refer to it as a demonstrator. It consists of several plug-in modules to emulate certain defect states. To decrease the experimental effort we rely on design of experiment techniques (DoE) yielding suitable parameter sets to be applied for each measurement. Typically, the raw data is not particularly suitable for feature generation and needs to be pre-processed in consideration of the properties of a synchronous motor. In Sect. 3 we describe the steps of data processing and the application of a coordinate transform in order to densify signal information in a manageable number of so-called feature variables. We use subspace methods for this purpose.

An adequate feature set, described in Sect. 4, needs to be identified for classification of defect states. This process also performs feature reduction, as a smaller number of features leads to more efficient and reliable decisions. Increased interpretability as well as reduced information complexity lead to less learning effort for classification algorithms. Moreover, the response time of condition monitoring systems decrease—an important criterion for future implementations in industrial applications. The Linear Discriminant Analysis (LDA) [1] is used for the definition of an optimized feature space, which allows for a robust distinction between defect classes. Section 5 addresses the validation of the proposed method based on measurements of the demonstrator. The classification is done by a Fuzzy Pattern approach, which provides a membership value for each data set with respect to known classes.

2 Experimental Setup and Data Acquisition

The basis for data acquisition is a demonstrator developed within a publicly funded research project being part of the German Federal Ministry for Economics and Technology's funding programme *AUTONOMICS* [2].Various intact and defective components can be plugged into the demonstrator, which is then operated under different load conditions. The basis for feature extraction is given by two measured phase currents of the electric drive denoted by φ_i, with $i = \{1, 2\}$. For experimental convenience, the currents are measured with an external probe using an oscilloscope.

The purpose of the demonstrator is to cover a well defined range of typical defects in drive train applications. Within this project artificial pitting of ball bearings, axle displacement and inclination of gear-wheels served as reference damages. To reduce experimental effort, design of experiment techniques (DoE) were used to identify reasonable combinations of those defects. According to [10], an orthogonal array $OA\left(12,\ 4^1 3^2\right)$ was used representing 12 runs of the experiment in total with different configurations. The term 4^1 refers to four realized levels of pitting, whereas the other two reference damages had 3 levels each, leading to the term 3^2. The DoE approach ensures a high variation of influential parameters with respect to each other, while the neccessary number of experiments is kept to a minimum.

For characterization of the demonstrator a variation of operating conditions was applied to all 12 basic runs, each representing a certain combination of reference defects. We chose to vary the rotational frequency and load torque of the drive train as well as the lateral forces of the bearings with three levels each in terms of DoE. This leads to an experimental design using the orthogonal array $OA\left(9,\ 3^3\right)$. Hence, $12 \cdot 9 = 108$ states of the demonstrator were measured in total.

3 Feature Generation

3.1 Data Pre-processing

The classification algorithms presented in this paper rely on the nonambiguous correlation of continuously measured time signals and certain health states of the demonstrator drive train. It is therefore important to preprocess the signals in favor of comparability and reproducibility. In case of the sinusoidal phase currents φ_i a potential phase offset will disturb the classification process. To address this problem we use the analytic signal representation

$$\varphi_{a_i}(t) = \varphi_i(t) + j\hat{\varphi}_i(t) \tag{1}$$

instead of the original phase currents, where $\hat{\varphi}_i$ is the Hilbert transform of the corresponding signal :

$$\hat{\varphi}_i(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\varphi_i(t)}{t-\tau} d\tau. \tag{2}$$

The transformed signal $\varphi_{a_i}(t)$ is interpreted as a series of complex numbers with a certain amplitude and angle α, which allows for statistical analysis. We confine the analysis to the calculation of a limited number of normalized statistical values $\bar{\varphi}_a$ with respect to N values of $\alpha \in [0, 2\pi]$. The resulting vectorial signal

$$\mathbf{s} = \left[\bar{\varphi}_a(\alpha_1), \ldots, \bar{\varphi}_a(\alpha_N)\right]^\mathsf{T} \tag{3}$$

is the basis for feature extraction described in the following section.

3.2 Feature Extraction

In pattern recognition, machine learning and classification the processing is rarely done on signal level. Instead, features are used to derive a decision since this approach is usually less computational expensive. The challenge in machine diagnosis is to find features representing the most typical physical characteristics of a technical system. Apart from empirical methods a model based technique can be used as described in [3] for instance. Once a suitable feature set of n features is found, the classification process is applied to the so-called feature space of dimension n. Ideally, features generated from signals of a certain machine condition will form dense clusters in the feature space. The distance between clusters representing different operating states should be as large as possible to facilitate classification. For the drive train application we propose a feature generation method based on proper orthogonal decomposition (POD) as known from model order reduction theory [8]. We assume that (3) represents a single data point in a space of dimension N. POD is used to find a reduced orthogonal basis, which spans a smaller subspace containing approximately all of the measured and pre-processed data. For this purpose all measurement data is collected within the snapshot matrix

$$\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_K] \tag{4}$$

considering K measurements in total. The eigenvalue decomposition

$$\left(\mathbf{S}\mathbf{S}^\mathsf{T}\right)\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{5}$$

provides eigenvectors $\mathbf{v}_i$, which represent the new orthogonal basis. The eigenvalues λ_i contain information about the significance of the respective eigenvector. The reduced basis results from selecting only the first n significant eigenvectors. Any signal $\mathbf{s}$ incorporated in the POD process may then be expressed as

$$\mathbf{s}_k \approx \sum_{i=1}^{n} c_{k_i}\mathbf{v}_i = \mathbf{V}\mathbf{c}_k;\ 1 \leq n < N \tag{6}$$

with $\mathbf{v}_i$ being the reduced basis vectors, sorted according to their eigenvalues. This method captures significant characteristics of $\mathbf{s}$ within the first n coefficients c_{k_i}. The value of n depends on the desired projection error and is typically much smaller than the signal length N. Hence, the signal can be described by a comparatively small number of coefficients. This consolidation of information usually leads to a good clustering of data within the subspace. If there are noticeable differences in measured data of different drive train conditions, discriminable clusters will form as long as noise does not obscure the effects. The features of the measured data may be arranged within a feature matrix

$$\mathbf{F} = [\mathbf{c}_1, \ldots, \mathbf{c}_K]^\mathsf{T} \tag{7}$$

of size $K \times n$, with K representing the number of data sets and n being the number of considered features.

Apart from classes representing only one drive train condition, the user may also want to define separate classes comprising of more than one condition of the system. In the following each of the above mentionend 12 basic setups refers to a separate defect class containing 9 different machine conditions respectively. The next section describes how to generate a reduced feature set comprising much less than n features, which will then be used for identification of such user defined classes.

4 Feature Analysis and Reduction

The feature matrix $\mathbf{F}$ in (7) contains all measured data. To discriminate between user defined classes, i.e. defect states, separate feature matrices $\mathbf{F}^{(d)}$ are introduced with d as the class index. We use the LDA algorithm [1] to define a minimum set of new features derived from the original set because of its possibility of faster discriminant adaption compared to Support Vector Machines (SVM) [9]. The LDA is a method to find a direction $\mathbf{w}$ in the feature space such that the projection of data onto $\mathbf{w}$ leads to maximum separation between the projected classes. The values

$$y_k = \mathbf{w}^\mathsf{T}\mathbf{c}_k \tag{8}$$

may then be used for classification. The class separation D can be expressed as

$$D(\mathbf{w}) = \frac{\mathbf{w}^\mathsf{T}\mathbf{\Sigma}_b\mathbf{w}}{\mathbf{w}^\mathsf{T}\mathbf{\Sigma}\mathbf{w}} \longrightarrow \max_{\mathbf{w}} \tag{9}$$

where $\mathbf{\Sigma}_b$ is the interclass covariance and $\mathbf{\Sigma}$ is the sum of intraclass covariance matrices. $\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}$ are of size $n \times n$ and are generated using the class feature matrices $\mathbf{F}^{(d)}$. For a certain direction $\mathbf{w}$ the separation $D(\mathbf{w})$ will be maximized, more precisely the ratio of interclass and intraclass distances becomes maximal. Equation (9) leads to a generalized eigenvalue problem, where the solution for $\mathbf{w}$ equals the most significant eigenvector of $\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_b\right)$. For the given problem one direction $\mathbf{w}$ may not be sufficient for reliable classification. Instead we construct a subspace with the non-orthogonal basis $\left\{\mathbf{w}_1, \ldots, \mathbf{w}_{\hat{d}}\right\}$ with $\mathbf{w}_d$ representing the most significant eigenvectors and $\hat{d}$ is the number of user defined classes. We use $\hat{d}$ basis vectors instead of the sufficient number of $(\hat{d} - 1)$ for separation of known classes. This inreases the capability of detecting unknown class data, which was not part of the analysis in (4). Let the projection matrix be $\mathbf{W} = \left[\mathbf{w}_1, \ldots, \mathbf{w}_{\hat{d}}\right]$, then the reduced feature vectors $\mathbf{y}_k$ read

$$\mathbf{y}_k = \left(\mathbf{W}^\mathsf{T}\mathbf{W}\right)^{-1}\mathbf{W}^\mathsf{T}\mathbf{c}_k. \tag{10}$$

In (10) we account for the rectangular matrix $\mathbf{W}$ by using its pseudo-inverse, such that the elements of $\mathbf{y}_k$ are equivalent to the coordinate values of the projection of $\mathbf{c}_k$ with respect to the basis vectors. The reduced feature matrix

$$\mathbf{F}_r = [\mathbf{y}_1, \ldots, \mathbf{y}_K]^\mathsf{T} \tag{11}$$

denotes the input training data for the Fuzzy-Pattern-Classifier described in the next section.

5 Classification method

The features chosen by the aforementioned concepts are classified using a Fuzzy-Pattern-Classification approach, actually the *Modified-Fuzzy-Pattern-Classifier* (MFPC). As MFPC is well-suited for industrial implementations, it has been already applied in many applications where it proved its performance [7]. Another reason we chose MFPC is to avoid rigorous classification and to increase interpretability of the class membership values. Based on fuzzy membership functions $\mu_i^{(d)} \in [0, 1]$, MFPC is employed as a useful approach to modelling complex systems and classifying noisy data. MFPC's membership function for an observation x_i is a unimodal parameter-based potential function of the form

$$\mu_i^{(d)}\left(x_i, \mathbf{p}_i^{(d)}\right) = \exp\left\{-\left(\frac{x_i - \bar{y}_i^{(d)}}{\nu_i^{(d)}}\right)^{\gamma_i^{(d)}} \ln 2\right\}$$

with respect to class d and feature number i. The parameter vector $\mathbf{p}_i^{(d)} = (\bar{y}_i^{(d)}, \nu_i^{(d)}, \gamma_i^{(d)})$ defines the membership function's properties, namely mean value $\bar{y}_i^{(d)}$, width $\nu_i^{(d)}$, and steepness of its edges $\gamma_i^{(d)}$ with

$$\bar{y}_i^{(d)} = \Delta + \min_k(y_{k_i}^{(d)}),\ \nu_i^{(d)} = \left(1 + 2\,\nu_{i_\mathrm{E}}^{(d)}\right) \cdot \Delta,$$
$$\Delta = \frac{\max_k(y_{k_i}^{(d)}) - \min_k(y_{k_i}^{(d)})}{2}.$$

Parameter $\nu_{i_\mathrm{E}}^{(d)} \in \mathbb{R}^+$ is a tuning parameter which can be adjusted by an expert, if needed. The MFPC membership function's parameters are obtained automatically during a learning phase, where the features y_{k_i} are extracted from K typical observations of a class. All outputs of the functions are aggregated with a fuzzy averaging function network resulting in a single membership value $\mu^{(d)}$

$$\mu^{(d)}\left(\mathbf{x}, \mathbf{p}^{(d)}\right) = \left(\prod_{i=1}^{M} \mu_i^{(d)}\left(x_i, \mathbf{p}_i^{(d)}\right)\right)^{\frac{1}{M}} \tag{12}$$

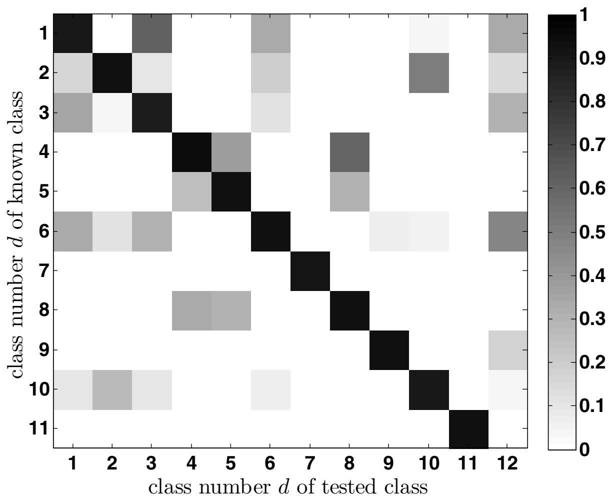

Figure 1. Membership values $\bar{\mu}^{(d)}$ for each tested class against known classes.

per class d, where M is the number of features. The most likely class membership of a single observation $\mathbf{x}$ is found by defuzzifying the corresponding membership values (12) using

$$d_{\mathbf{x}} = \arg\max_{d} \left(\mu^{(d)} \left(\mathbf{x}, \mathbf{p}^{(d)} \right) \right).$$

More detailed information can be found in [6].

6 Results and Outlook

The previously elaborated feature selection approach is validated with data from the demonstrator. Our evaluation is focused on measured data as explained in Sect. 2. Based on DoE, there were $\hat{d} = 12$ measurement runs of different defect states, which represent the defect classes. Class $d = 1$ does not comprise any defects and acts as a reference for the *good* state. For each class a reduced feature set (cf. Eq. (10)) was calculated. For training of the MFPC we used classes $d = 1 \dots 11$, whereas class 12 was considered as an unknown state for testing purposes. The MFPC parameters were set to $\gamma^{(d)} = 8$ and $\nu_{i_{\mathrm{E}}}^{(d)} = 0.01$, which lead to satisfactory results within the scope of the considered application. We performed a classification of the known training data as well as the unknown data from class 12. All membership values $\mu^{(d)}$ of a data set $\{k\}$ related to a certain class d were averaged. Hence, mean membership values $\bar{\mu}^{(d)}$, $d = 1 \dots 11$ can be assigned to each class, representing the mean membership with respect to all known classes. The re-classification procedure of known classes should lead to a 11×11 matrix with a dominant diagonal. Furthermore we added the classification results for class 12, such that the matrix becomes 11×12. Figure 1 shows the classification result. As can be seen, the first 11 classes were correctly classified. Since class 12 represents an unknown state, it cannot be assigned to any known class resulting in the comparatively small membership values. Thus, the classification algorithm is able to detect new, i. e. unknown defect states, which is important for practical applications.

We presented a method to perform condition monitoring for electric drive trains without using any additional sensors. The feature generation and classification algorithm relies on reference measurements at least of the *good* state, which leads to an adaptive method suitable for applications similar to the one presented in this paper. The identification of previously unknown states is possible and allows for the estimation of the general health state of a system. For this paper we used data from a demonstrator device. The integration of signal analysis into the drive control and application under field conditions is planned as the next step.

Acknowledgements

This work was partly supported and financed by the German Federal Ministry of Economics and Technology, Grant Number 01MA09006A.

References

[1] E. Alpaydın. *Introduction to Machine Learning*. MIT Press, Cambridge, 2010.

[2] AUTONOMICS – Autonomous, simulation-based systems for small and medium-sized enterprises: Technology programme of the Federal Ministry of Economics and Technology, http://www.autonomik.de, last visit: 2013-05-29.

[3] O. Enge-Rosenblatt, C. Bayer, and J. Schnüttgen. Modeling a drum motor for illustrating wearout phenomena. In *9th International Modelica Conference (Modelica'2012)*, pages 889–896, Munich, 2012.

[4] E. Geisberger and M. Broy, editors. *Cyber-physical Systems*. acatech POSITION PAPER. Springer, Berlin, 2011.

[5] Intelligent Technical Systems OstWestfalenLippe (it's OWL): Leading-Edge Cluster of the Federal Ministry of Education and Research, http://www.its-owl.de, last visit: 2013-05-29.

[6] V. Lohweg, C. Diederichs, and D. Müller. Algorithms for hardware-based pattern recognition. *EURASIP Journal on Applied Signal Processing*, 2004(12):1912–1920, 2004.

[7] V. Lohweg and U. Mönks. Fuzzy-Pattern-Classifier Based Sensor Fusion for Machine Conditioning. In C. Thomas, editor, *Sensor Fusion and its Applications*, pages 319–346. InTech, 2010.

[8] W. Schilders, H. van der Vorst, and J. Rommes. *Model Order Reduction*. Math. in Industry. Springer, Berlin, 2008.

[9] B. Schölkopf and A. J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, Cambridge, 2002.

[10] C. Wu and M. Hamada. *Experiments - Planning, Analysis and Optimization*. Probability and Statistics. Wiley & Sons, Hoboken, New Jersey, 2009.