Applied Soft Computing 63 (2018) 50–58

Contents lists available at ScienceDirect

Applied Soft Computing

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

The design of multiple linear regression models using a genetic algorithm to diagnose initial short-circuit faults in 3-phase induction motors

Arismar M.G. Júnior [a,*], Valceres V.R. Silva [b], Lane M.R. Baccarini [b], Lívia F.S. Mendes [a]

[a] Universidade Federal de São João del Rei, Electrical Engineering Postgraduate Program, Praça Frei Orlando 170, São João del Rei, 36307-352 Minas Gerais, Brazil
[b] Universidade Federal de São João del Rei, Electrical Engineering Department, Praça Frei Orlando 170, São João del Rei, 36307-352 Minas Gerais, Brazil

ARTICLE INFO

Article history:
Received 19 March 2016
Received in revised form 3 October 2017
Accepted 9 November 2017
Available online 16 November 2017

Keywords:
Induction machine
Incipient short-circuit
Non-invasive fault diagnosis
Multiple linear regression modelling
Analysis of variance
Genetic algorithms

ABSTRACT

Induction motors are robust machines that are often exposed to a variety of environmental and operating conditions that can result in a number of failures during their use. One such fault is a short-circuit that starts in a few turns, and quickly extends to other winding sections. Early detection and diagnosis of this type of failure is very important and can prevent the complete motor loss. In this work, a multiple linear regression modelling technique is used in synergy with the genetic algorithm optimization and the analysis of variance methods to obtain models to classify the motor operating in normal and in short-circuit conditions. The proposed method is suitable for application in real industrial plants due to three important features: (i) it uses RMS values of voltages and currents, (ii) only simulation data is required to obtain the MLR classification model and (iii) incipient faults can be identified with high accuracy. Experimental tests carried out over a wide range of machine operation conditions demonstrates the simplicity and effectiveness of the new diagnosis method.

1. Introduction

The majority of industrial loads are powered by three-phase induction motors, which are robust machines, of low maintenance cost and can work in a range of environments. Even though they are robust, various failures can occur in these machines since they can work under different operating conditions and in hostile environments [1]. Fast fault detection in dynamic systems, like induction machines, provides a high degree of reliability in industrial processes, because it avoids production downtime, material losses, reduction in quality and even accidents with human beings [2,3].

According to [4], faults like short-circuits in coils represent 21% of the total faults in electric motors. In this case, the stator coil degradation starts with a short-circuit between a few turns in the same phase. The faulty current can be nearly twice the blocked rotor current and generates local heat that spreads quickly to some other coil parts [5]. Thus, an early detection of this kind of failure is very important since it can prevent more serious failures that can lead to irreversible loss of the motor. However, initial fault detection is not a trivial task, since its impact over the motor characteristics is still low [6–8]. In addition, the supply voltage unbalance can be mistaken for stator winding failure [9].

Several techniques for fault diagnosis in an induction motor stator have been developed by researchers using different machine variables such as magnetic flux [10], supply voltage [11], vibration [12] and stator current signals [13–17]. Among these approaches, it is clear that motor current signal analysis (MCSA) is by far the most preferred technique for diagnosing faults, since it uses sensor signals normally available in real plants, being non-invasive and cheap [18]. However, under certain conditions its application is not sensitive enough because it has a low signal to-noise ratio, which is more evident in inverter-fed motors [19–21].

Recently, a new non-iterative algorithm called empirical demodulation has been proposed by [22], which determines the envelope of signals providing important information about the operating condition of an electric machine. To show the application of the technique in the diagnosis of induction motor failures, the authors present case studies involving the detection of broken rotor bar and short-circuit faults in the stator winding, using magnetic flux and vibration signals.

Some modelling and classification techniques have been used to categorize features of the motor operating condition [23,24]. In paper [25], the authors proposed a new approach for identify-

* Corresponding author.

https://doi.org/10.1016/j.asoc.2017.11.015
1568-4946/

ing broken rotor bar, short-circuits in the stator coils and bearing defects, by analyzing the amplitudes of the current signals in the time domain. The experimental results showed a high efficiency of the proposed classifier technique. For the diagnosis of failures in induction motors fed by frequency inverters, an alternative intelligent method is presented in [26].

The starting point for the development of a reliable fault diagnosis tool is to have a good understanding of the difference between the behavior of the machine in a healthy state and under fault conditions [27]. The models also allow the signature of the failure to be obtained and the diagnostic method for machines with different power ratings to be tested. A dynamic model to analyze electrical (rotor broken bar and stator winding short-circuit) and mechanical faults (unbalance, misalignment) in induction machines was proposed by [28]. The proposed model only needs the same parameters of the traditional model and also includes net asymmetries and load conditions.

A variety of theories and methods, in particular neural networks, fuzzy logic, neural fuzzy and genetic algorithms (GAs) provide the possibility of making the diagnostic systems more autonomous [29,30]. Although giving valuable results, some methods normally need large data sets to provide correct answers. In addition, the models are normally trained with real data, which makes the application of diagnostic techniques in a real plant more difficult. Consequently, in this work, regression models were obtained for the diagnosis of incipient faults in stator windings. The models become part of a specialist system created to test and classify, in real time, the machine operating conditions. The new diagnostic approach uses multiple linear regression (MLR) in synergy with the GA and analysis of variance methods to design polynomial models, being applicable in industry due to the following relevant advantages: (i) it needs only the 60 Hz component of the voltage and current that can be obtained through voltmeters and ammeters, (ii) few and only simulation data are necessary to create the classification MLR models and (iii) incipient faults can be diagnosed with a high accuracy. The condition of voltage imbalance inherent in a power system, does not affect the diagnosis of the failure.

2. Multiple linear regression modelling

Regression analysis is a statistical technique for investigating and modelling the relationship between variables, where the goal is to obtain an approximate function between the response of interest (output) and independent variables (inputs). When more than one independent variable is involved, regression analysis is called Multiple Linear Regression (MLR) Montgomery et al. (2006) [31].

In the majority of MLR problems it is not known how the independent variables relate to the dependent ones. Thus, the first step, and the main problem, is to find an approach for relating them. Generally, the first-degree and second-degree polynomial models are enough, as shown in Eqs. (1) and (2) respectively:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon \tag{1}$$

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j>1}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{2}$$

where: β's are the polynomial coefficients, x the independent variables, y the dependent variable and ε other possible variation forms, like errors or lack of numerical convergence. The index k represents the total number of variables and the indexes i and j represent a specific variable between 1 and k.

Eqs. (1) and (2) can be written in a matrix form as in Eq. (3):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}' + \boldsymbol{\varepsilon} \tag{3}$$

For the second order models and n being the total number of the systems observations, the matrices and vectors involved are:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_k & \beta_{11} & \cdots & \beta_{kk} & \beta_{12} & \cdots & \beta_{1k} & \beta_{23} & \cdots & \beta_{2k} & \cdots & \beta_{(k-1)k} \end{bmatrix}'$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} & x_{11}^2 & \dots & x_{1k}^2 & x_{11}x_{12} & \dots & x_{11}x_{1k} & x_{12}x_{13} & \dots & x_{12}x_{1k} & \dots & x_{1(k-1)}x_{1k} \\ 1 & x_{21} & \dots & x_{2k} & x_{21}^2 & \dots & x_{2k}^2 & x_{21}x_{22} & \dots & x_{21}x_{2k} & x_{22}x_{23} & \dots & x_{22}x_{2k} & \dots & x_{2(k-1)}x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} & x_{n1}^2 & \dots & x_{nk}^2 & x_{n1}x_{n2} & \dots & x_{n1}x_{nk} & x_{n2}x_{n3} & \dots & x_{n2}x_{nk} & \dots & x_{n(k-1)}x_{nk} \end{bmatrix}$$

According to [31], the method of least squares is typically used to estimate the coefficients of the MLR models, i.e., the vector $\boldsymbol{\beta}$, as in (4):

$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{4}$$

The signals of currents and voltages of the induction motor, in different operating conditions, can be used in the establishment of a set of independent variables. In this way, by means of Eq. (4) and an output of interest, MLR models can be developed for the diagnosis of presence or absence of faults.

3. Analysis of variance

Through the analysis of variance, a fast measure of a MLR model quality can be obtained and low significant terms can be identified [32]. Thus, the MLR models developed for the classification of the induction motor operating conditions, as well as the independent variables employed in the adjustment of these models, i.e., the regressive variables used, can have their applicability assessed.

To evaluate whether at least one of the regressive variables x_1, $x_2, \ldots, x_k$ is significant for the model, the statistic distribution measure F_0 given by Eq. (5) is computed:

$$F_0 = \frac{SS_R/k}{SS_E/(n-k-1)} \tag{5}$$