Intelligent Data Mining for Turbo-Generator Predictive Maintenance: An Approach in Real-World

Alexandre Pellicel[1], Gonçalo Cássio[1], Marco Aurélio A. Lopes[1],
Luiz Eduardo Borges da Silva[2], Erik Leandro Bonaldi[2],
Levy Ely de Lacerda de Oliveira[2], Jonas Guedes Borges da Silva[2],
Germano Lambert-Torres[2], and Pierre Rodrigues[3]

[1] TermoNorte Energy Thermal Power Plant Co., Porto Velho, Brazil
{alexandre.pellicel, goncalo.cassio, marco.lopes}@termonorte.com.br
[2] Institute Gnarus, Itajuba, Brazil
{levy.oliveira, erik.bonaldi, jonas.borges, germanoltorres, leborgess}@gmail.com
[3] Jordão Engineering Co., Rio de Janeiro, Brasil
pierre.rodrigues@jordaoengenharia.com.br

Abstract. This paper presents the development of a supervision system for predictive maintenance and diagnosis of turbo-generators. The aim of the developed system is to verify the degradation conditions of TermoNorte generators. Initially, a system for extracting features of the turbo-generator operational database has been developed to detect possible problems that cause premature fails. The system has been divided in two parts. The first one is a data acquisition system directly connected to the generator in order to sample some operational variables. The second part concerns an intelligent data mining, based on Rough Sets Theory, into the database involving the supervision system variables, to use the existing historic data to perform analysis of the problems and possible causes.

Keywords: Electrical measurements, signal processing, rough sets, data mining, intelligent systems, turbo-generators.

1 Introduction

The generators are the most important equipments in the energy generation process. The power system reliability, power system supply and power system stability are indexes directly affects for the generator operational conditions. For this reason, protection and monitoring equipment are increasingly employed in order to prevent fails [1].

One of the technologies that can be employed within the purpose of predicting failures is the electric signature analysis (ESA) [2], which consists of a set of methods and techniques that monitor the condition of electric machines by identifying patterns and deviations. It is detected by processing and analysis of voltage and current signals acquired machinery under monitoring.

These techniques based on electrical signatures can be applied from the generator and primary source until the motor and load coupled. They may be based on: (a) invasive methods, such as the electric circuit analysis (with static analysis and non-energized machine, also referred to as offline analysis and therefore invasive), or (b) non-invasive methods, such as ESA (dynamic analysis, i.e. with the machine in operation, also referred to as online analysis) [2].

For a more comprehensive monitoring of the generator, it is important to the application of invasive and non-invasive methods, based not only on the signature electric as well as other monitoring techniques such as vibration analysis. It is recommended the application of invasive techniques in shutdowns, while non-invasive techniques should be applied periodically during the operating cycle of the machine.

This project aims to develop a methodology for the detection and dynamic analyses of online monitoring of the condition of turbo-generators based on acquisition, processing and analysis of voltage and current signals. The main fails such as short-circuit in stator and rotor windings, fails in excitement system, misalignment and eccentricity of spinning field have been studied by electric signature analysis.

The paper presents the developed system and some practical results in a TermoNorte Thermal Power-Plant, located in Porto Velho, northwest part of Brazil.

2 Electric Signature Analysis

Electric Signature Analysis (ESA) is the term used for all evaluations of voltage and current signals of electric machines. The most common analysis transforms the voltage and current signal to the frequency domain where they are analyzed. The analysis is based on two fundamental assumptions: (a) the signature of a machine with failure is different from the signature of a machine in perfect state of operation and (b) the failures are repeated with regular patterns, causing failure patterns, which can be identified and related parts of the machine.

These techniques can be applied in electric motors and generators. It is important to note that the Voltage Signature Analysis (VSA) is related to an upstream analysis, i.e. toward the generator; and the Current Signature Analysis (CSA) is related to a downstream, i.e. toward the motor. In this project, CSA and the Extended Park Vector Approach (EPVA) are the methods used in this development because they have more features applicable to electric generators. Also these methods have been applied in electric motors, but not to generators yet [2].

2.1 Overview of Current and Voltage Signature Analysis

CSA or VSA techniques are used to generate analyses and trend of electric machines. They aim to detect predictive failures in a plant, such as: problems in the stator winding, rotor problems, problems on coupling load, efficiency and loading of the system; bearing problems, among others.

It might be of surprise, but electrical signals (voltage and\or current) can carry additional information about electrical and mechanical problems of generating equipment, but the machine works as a transducer for mechanical failures, allowing the electrical signals (voltage and/or current) can carry information of electrical and mechanical problems. The signals of current and/or voltage of one (or three) phases of the machine produce, after examination, the *signature of* machine, i.e., its operating pattern. This signature is composed of frequency magnitudes of each individual component extracted from their signals of current or voltage. This fact allows the monitoring of the evolution of the frequency magnitudes, which can denote some sort of evolution of the operational conditions of the machinery.

The response that the user wants is to know whether your machine is "healthy" or not, and what part of machine is in failure.

This analysis (diagnosis) is not easily done because it involves a set of comparisons with previously stored patterns and own "history" of the machine under analysis. At this moment, usually an expert is called to produce the final diagnosis, generating command when stopping the machine. Thus, the system developed in this project for automatic diagnosis combines the history of turbo-generator, expert knowledge and failures patterns and it can be very useful for a power company.

2.2 Extended Park Vector Approach

The EPVA technique should be used to verify electric stator imbalances. However, it can only be used if the signals of voltage and/or current have been demodulated [3]. The central idea of this technique is checking failures by the distortion of Park´s circle, i.e., more distortion in the Park´s circle more is the unbalance of the machine. The current components of the Park vector are described by i_D and i_Q:

$$i_D = \left(\frac{\sqrt{2}}{\sqrt{3}}\right)i_A - \left(\frac{1}{\sqrt{6}}\right)i_B - \left(\frac{1}{\sqrt{6}}\right)i_C \tag{1}$$

$$i_Q = \left(\frac{1}{\sqrt{2}}\right)i_B - \left(\frac{1}{\sqrt{2}}\right)i_C \tag{2}$$

Where the currents i_A, i_B and i_C are the three phases. In ideal conditions:

$$i_D = \left(\frac{\sqrt{6}}{2}\right)i_M \cos(\omega t - \alpha) \tag{3}$$

$$i_Q = \left(\frac{\sqrt{6}}{2}\right)i_M \sin(\omega t - \alpha) \tag{4}$$

For normal conditions, Park circle is centered at the origin of coordinates.

The Park circle has distortions when there are abnormal conditions of operation or when mechanical or electric failures occur. However, these distortions in the Park

circle are not easy to be seen or measured, hence the proposition of the Extended PVA (EPVA), observing the spectrum module of Park vector.

The EPVA technique combines the robustness analysis of Park circle and the flexibility of spectral analysis [4]. An important feature of the Park transformation process is the fundamental component of analyzed signals is erased [5]. This fact allows the component characteristics of failure to appear with greater prominence. And more, to be a method that covers the three-phase simultaneously electric stator imbalances are also covered by this method [6].

3 Electrical Signal Processing

For the characteristic extraction of digital signals, there is a pre-conditioning process and then some paths to compute the values of each variable. Different parameters are obtained in the time domain and frequency domain. In Fig. 1, a flowchart of the used techniques is shown.
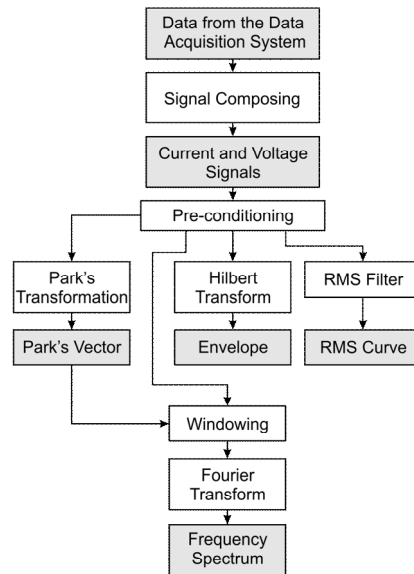

Fig. 1. Block diagram of the algorithms of signal conditioning and processing.

The grayish background blocks represent a processed signal that can be viewed or used for the characteristic extraction. The blank blocks represent an algorithm or a processing applied to digital signal. Below a brief commentary on each of the processing blocks is presented:

• Signal Composition: it converts the data from the data acquisition system to digital signals whose amplitudes represent actual values for current and voltage.

• Pre-conditioning: process that eliminates the initial part of the signal to avoid samples obtained during the transients of the filters. Then the average value of each signal whose nature is alternating is deleted.

• Park Transformation: when there is a three-phase electrical system composed by three currents (I_A, I_B and I_C) and three voltages (V_{AB}, V_{BC} and V_{CA}), the Park transformation is applied to obtain the Park vector, composed by components I_Q, I_D and I_0. In some cases, it is also used the spectrum of this vector module for electric system imbalance.

• Hilbert transformation: when applied to a signal, it returns the magnitude (envelope) and instantaneous phase.

• RMS Filter: knowing the fundamental frequency of the signal, the RMS filter returns the instantaneous RMS value of the signal during the sampling period, resulting in the so-called RMS Curve.

• Windowing: filter applied to a signal in time to reduce the effect of "leakage" in the frequency spectrum. There are several types of windowing (Blackman, Hamming, Hanning, etc.), the Blackmann windowing has been used. This window allows the identification of peaks as lobes slightly wider and less "leakage" on their side bands than other Windows.

• Fourier Transform: used to transform a time domain signal into the frequency domain, the discrete Fourier transform (DFT) returns a vector with the spectrum amplitudes and their phases. To accelerate the achievement of the DFT, we used an algorithm called FFT (Fast Fourier Transform).

For each acquired electrical signals, various parameters are computed. These parameters are used for the evaluation process and for the extraction of new features, and they are listed below.

• Average amplitude: the average value of the signal in the period under review;

• RMS amplitude: also called effective value or mean square;

• Minimum and maximum amplitude values: maximum and minimum values of amplitude in the period under review;

• Amplitude, phase and fundamental frequency: value of amplitude and frequency in Hz of the fundamental component of signal (electrical system to signal fundamental frequency is 60 Hz);

• Fundamental harmonics: multiples of the fundamental component.

• Harmonic distortion index (HDI): it indicates the significance of harmonic content when compared to the fundamental component of the signal.

4 Description of the Data Mining Algorithm

This section introduces expeditiously the data mining algorithm used to perform comparisons between the processed signals, the database signals and failures patterns. The used technique was based on the Rough Set Theory [7]. This technique aims to extract a set of rules (or conditions) from a database through two hyper-sets, called

upper approximation set and lower approximation set. The set of rules contains the lower approximation set and is contained by the upper approximation set. The central idea of the algorithm is to reduce the number of elements in the upper approximation set and to increase the number of elements in the lower approximation set. In an ideal condition, these two sets would become only one set that would be the required set. This set is represented by the set of production rules.

The used algorithm [8] has six main steps, they are:
1. initialization;
2. remove equal examples;
3. remove of dispensable attributes;
4. compute the core set;
5. compute the reduce set; and
6. merge rules.

In the first step (initialization), ranks (classes) of each attribute (input or output variables) are defined, and each interval receives an identification label. This division creates a cross-linked sample space and the next step may apply (remove equal examples). All examples within a same hyper-cube are grouped into only one.

The next step is to check equal attributes or unnecessary classification (dispensable attributes). This is done in the first case by mere inspection and, in the second case, by the removal of each of them and subsequent verification of the inclusion of issues of classification.

Then, with only the essential attributes, the core of each rule set is computed. The core is formed by those attributes indispensable for that rule, and those sets of examples. Next step of forming core set consists of computing reduce set, which contain only of core attributes augmented by attributes qualifying exactly a rule. Finally, the rules are similar and the set of production rules for the classification of input signals.

In the context of the developed software for turbo-generator predictive maintenance, the algorithm and mathematical structures described were implemented and serve to extract knowledge of diagnostic data. Thus, the system is able to diagnose new cases on the basis of the knowledge extracted from previous cases. It is important to note that the procedure is transparent to the user, that is, it occurs within the computational package developed, activated by a button command in the program window itself and providing the user with the proper classification.

5 Illustrative Example of the Data Mining Process for Turbo-Generator Feature Extractions

The current development has been applied in the TermoNorte Thermal Power Plant, located in the Brazilian north region close to Amazon jungle. This power plant is composed by two plant in the same area. The TermoNorte I has a total generation capacity equal to 64 MW, from 4 Diesel Wärtsilä motor-generators, each one with 16

MW. The TermoNorte II has a total generation capacity equal to 340 MW, from 3 GE gas turbines.

5.1 Some Features about the Data Acquisition

The installed data acquisition system is composed by current and voltage transducers, a pre-processing acquisition module and a data acquisition module, shown in Fig. 2. This system has been entirely developed to TermoNorte Co. and very-well adapted to the severe conditions of Amazon jungle humidity. The current signals are taken from the secondary of panel CTs and the voltage signal from the panel PTs.
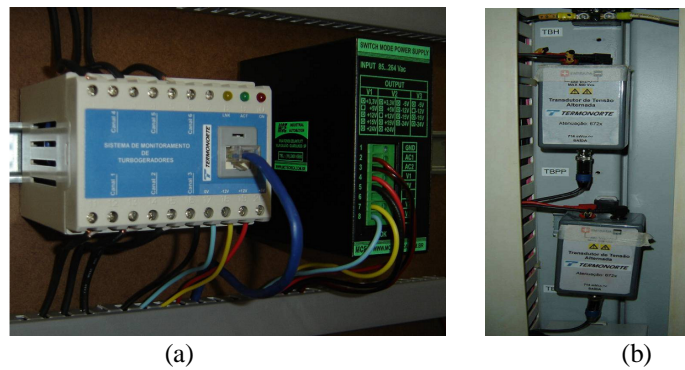

(a)   (b)

Fig. 2 Data acquisition system: (a) data acquisition pre-processing acquisition modules, and (b) voltage transducers.

5.2 Computational Package

The developed computational package is composed to two main parts: (a) data acquisition control and (b) feature extraction and signal processing. The first part contains parameters for data acquisition process. The main signal acquisitions are usually made through three acquisitions:

• Acquisition 1: it aims to collect both the signs of current and voltage (phase) of the turbine, so the EPVA techniques and energy quality are applied;

• Acquisition 2: acquisition of voltages, with the goal of applying the technique VSA (Voltage Signature Analysis);

• Acquisition 3: it does the acquisition of one of the stages for the application of the technique CSA (Current Signature Analysis).

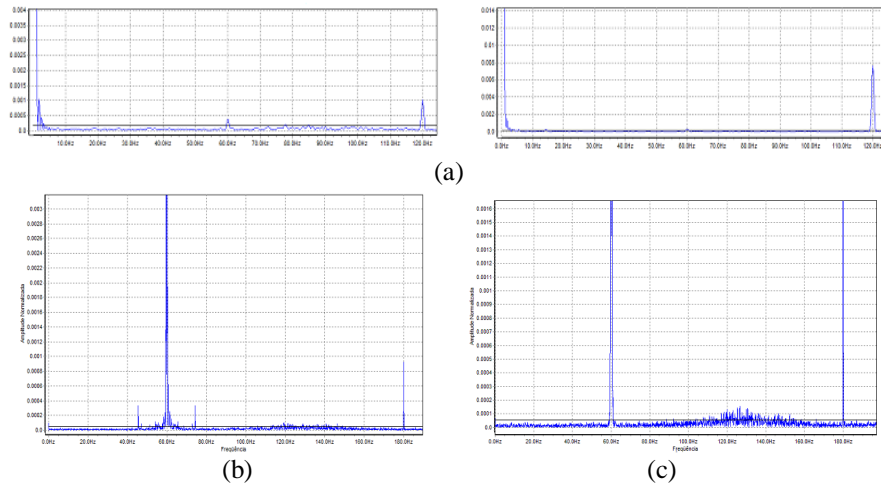Examples of signals acquire by the developed system are shown in Fig. 3.

Fig. 3 Acquired Signals by: (a) EPVA to current and voltage, respectively, (b) CSA signal, and (c) VSA signal.

Special window interfaces have been developed to transfer all system control to the operator. Examples of this interface are shown in Fig. 4. These interfaces are in Portuguese language. The first figure shows an example of the supervision interface with the data acquisition control information; and the second figure is one of the analysis procedures.
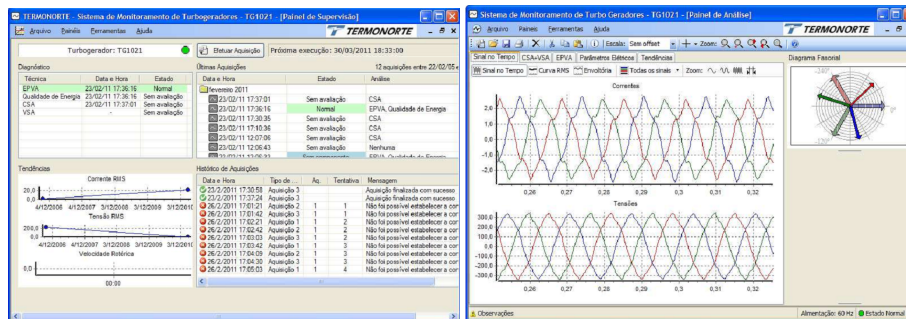
Fig. 4. Examples of user interface of the computational package.

5.3 Feature Extraction – Data Mining Process

The described algorithm in Section 4 has been also implemented in the computational package (second part of the package in our description – for users this division of the package doesn´t exist). The signals shown in Fig. 3 are expressed by their main features, such as frequencies, amplitudes, phases, and merge to turbo-generator parameters itself. This set of data is the input data, and must be related to a type of

previous operation condition: normal, abnormal, failure #1, failure #2, and so on. An example of the input signal database is shown in Table 1.

With the database the data mining process starts with the definition of labels (classes or ranks) for each attribute (input variable). The program contains a pre-set of labels for each attribute. This pre-set has been adjusted during the test phase of the prototype in the power plant. However, if the user would change the interval of these labels it is possible. However, in the daily operation, this pre-set of labels remains constant.

Table 1. Partial example of the signal acquisition database.

Acquisition	Sample Frequency (Hz)	Number of Samples	Time of Sample (s)	Spectral Definition (Hz)	Total Time (s)
1	8193	21845	2,7	0,3704	27
	1638		13,3	0,0752	37
2	8193	131072	16	0,0625	40
	1638		80	0,0125	104
3	8193	131072	16	0,0625	40
	1638		80	0,0125	104

Internally, the program merges the equal examples, verifies dispensable attributes, computes the core and the reduce sets, and finally produces the final set of rules. And academic example of this process is presented for a small database (part of the real database). Table 2 shows the data after the application of labels. Ten examples are shown with the following input attributes: frequency, amplitude, TDH (harmonic distortion level), and distortion (from Park Vector circle). The possible outputs are "normal", "warning", and "danger".

Table 2. Partial example of the signal acquisition database.

Example	Frequency	Amplitude	TDH	Distortion	Output
1	Low	Normal	Normal	Normal	Normal
2	Low	Medium	Medium	Normal	Normal
3	Low	Medium	Normal	High	Normal
4	Medium	Medium	Normal	Medium	Warning
5	Medium	Medium	Normal	High	Warning
6	Medium	High	Normal	High	Danger
7	Low	Medium	Medium	Normal	Normal
8	Medium	Medium	Normal	Medium	Warning
9	High	High	Medium	Medium	Danger
10	Medium	High	Normal	Medium	Danger

After the transformation from numbers in labels of the attribute values, the second step of the algorithm can be performed - to remove equal examples. In this case, examples 2 and 7 are equal, and 4 and 8 also. Then one of them can be removed without any type of information lack, resulting in Table 3.

Table 3. Original table of examples without repetitions.

Example	Frequency	Amplitude	TDH	Distortion	Output
1	Low	Normal	Normal	Normal	Normal
2	Low	Medium	Medium	Normal	Normal
3	Low	Medium	Normal	High	Normal
4	Medium	Medium	Normal	Medium	Warning
5	Medium	Medium	Normal	High	Warning
6	Medium	High	Normal	High	Danger
9	High	High	Medium	Medium	Danger
10	Medium	High	Normal	Medium	Danger

In order to verify possible dispensable attributes, each attribute is removed and a verification of possible mistake classification is performed. In this case, for instance, the attribute "frequency" is not dispensable because without it examples 2 and 5 present two different outputs for the same input. The same occurs with the attributes "amplitude" and "distortion". However, the attribute "TDH" is dispensable in this case because without this attribute the table remains consistent (Table 4).

Table 4. Original table of examples without repetitions and with the dispensable attributes.

Example	Frequency	Amplitude	Distortion	Output
1	Low	Normal	Normal	Normal
2	Low	Medium	Normal	Normal
3	Low	Medium	High	Normal
4	Medium	Medium	Medium	Warning
5	Medium	Medium	High	Warning
6	Medium	High	High	Danger
9	High	High	Medium	Danger
10	Medium	High	Medium	Danger

At this moment, the database is ready to compute the core set. Removing each value of each example and verifying the mistake in the classification, it is possible to computer each element of the core set. If the lack of the element creates a mistake, this element takes part of the core set, otherwise not. Table 5 presents the core set of the illustrative database example.

Table 5. Core set.

Example	Frequency	Amplitude	Distortion	Output
1	-	-	-	Normal
2	-	-	-	Normal
3	Low	-	-	Normal
4	-	Medium	Medium	Warning
5	Medium	Medium	-	Warning
6	-	High	-	Danger
9	-	-	-	Danger
10	-	High	-	Danger

Then the reduce set can be computed. It is made including the minimum number of attributes with the core to represent the example. In this case, it results in 11 examples (rules). Finally, the final set of rules is composed by 7 different rules, shown in Table 6.

Table 6. Reduce set and final set of rules.

Rule	Example	Frequency	Amplitude	Distortion	Output
1	1	Low	-	-	Normal
2	1	-	Normal	-	Normal
1	2	Low	-	-	Normal
3	2	-	Medium	Normal	Normal
1	3	Low	-	-	Normal
4	4	-	Medium	Medium	Warning
5	5	Medium	Medium	-	Warning
6	6	-	High	-	Danger
6	9	-	High	-	Danger
7	9	High	-	-	Danger
6	10	-	High	-	Danger

An example of the produced rule of the developed system is:

If $I_{PV} \geq -26db$ (0.05) then output = "warning" and failure = "stator current unbalance".

In English language:

If the current Park Vector component is equal to or bigger than -26 db (0.05) then the operational condition is "warning" and the possible failure is "stator current unbalance".

Rule-extraction algorithm is usually run once a quarter. The most important part of the process to the users is the analysis of the current signals, it means, the operational condition of the machine at this moment. For acquired current signals pass by the rule set and a condition of the generator is presented to the operator. The major part of the time the answer of the program is "Normal"; however, when a abnormal situation is
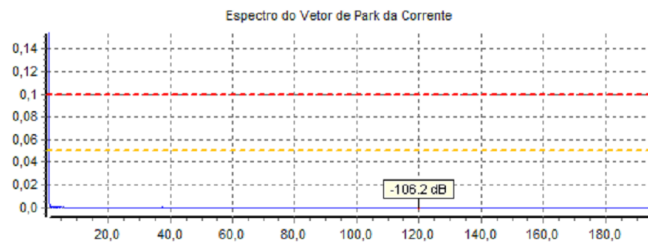

Fig. 5. Example of abnormal situation and warning and danger levels.

detected a failure pattern is shown to the operator. An example of this is presented in Fig. 5. This figure shows this abnormal situation with two pre-set lines. The yellow line expresses the warning level and the red line express a danger level. In this cases, -26 db and -20 db, respectively.

6 Conclusions

This paper shows a complete development of a supervision system with an intelligent data signal processing based on feature extraction using Rough Set Theory. The feature extraction relates processed current and voltage signals from the turbo-generator by VSA, CSA and EPVA techniques, turbo-generator electrical and mechanical parameters, and typical types of failures existing in this kind of machine.

Hardware and software have been developed to acquire and treat the electrical signals in a non-invasive process. It means, the operational condition of the generator is verified without any type of disturbance in the machine or in its control. The electrical signals are taken out of the machine, more specifically in the secondary of instrument transformers (CT and PT) in the panel control.

This system is currently in full operation at TermoNorte Thermal Power Plant, in Brazil.

References

1. Bonaldi, E.L., Oliveira, L.E.L., Lambert-Torres, G., Borges da Silva, L.E.: Proposing a Procedure for the Application of Motor Current Signature Analysis on Predictive Maintenance of Induction Motors. In: 20th Int. Cong. Exh. Condition Monitoring and Diagnosis Engineering Management, COMADEM 2007, Faro, Portugal (2007).
2. Bonaldi, E.L.: Failure Predictive Diagnostic in Three-Phase Induction Motors with MCSA and Rough Set Theory. Ph.D. Thesis, Itajuba Federal School of Engineering, Itajuba – Brazil (2006) - in Portuguese.
3. Cardoso, A.J.M.: Failure Diagnostic in Three-Phase Induction Motors. Coimbra Editora, Coimbra – Portugal (1991) - in Portuguese.
4. Benbouzid, M.H.: A Review of Induction Motors Signature Analysis as a Medium for Faults Detection. IEEE Trans. Industrial Eletronics 47, 984--993 (2000).
5. Cruz, S.M.A., Cardoso, A.J.M.: Diagnosis of the Multiple Induction Motor Faults Using Extended Park's vector Approach. Int. J. Comadem 1, 19--25 (2001).
6. Cruz, S.M.A., Cardoso, A.J.M.: Diagnosis of Stator Inter-Turn Short Circuits in DTC Induction Motor Drives. IEEE Trans. Industry Applications 40, 1349--1360 (2004).
7. Pawlak, Z.: Rough Sets. Int. J. Information and Computer Sciences .11, 341--356 (1982).
8. Rissino, S., Lambert-Torres, G.: Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications. In: Ponce, J., Karahoca, A. (eds.) Data Mining and Knowledge in Real Life Applications., pp. 35--58, ISBN 978-3-902613-53-0, In-Tech Press (2009).