CILAMCE 2017

Florianópolis - Santa Catarina - Brazil

CLASSIFICATION OF FAULTS IN A SWITCH MACHINE USING TYPE-1 AND NON-SINGLETON FUZZY LOGIC SYSTEM TRAINED BY HESTENES AND STIEFEL'S CONJUGATE GRADIENT METHOD

Alexandre Queiroz Zamagna Bouhid ⋆

alexandre.bouhid@engenharia.ufjf.br

Industrial and Mechanical Engineering Department

Renan Piazzaroli Finotti Amaral †

renan.finotti@engenharia.ufjf.br

Mechanical Engineering Department

Leonardo Goliatt da Fonseca ⋆

leonardo.goliatt@engenharia.ufjf.br

Graduate Program in Computational Modeling

Eduardo Pestana de Aguiar ⋆

eduardo.aguiar@engenharia.ufjf.br

Industrial and Mechanical Engineering Department

Electrical Engineering Post-Graduation Program

⋆ Federal University of Juiz de Fora, Rua José Lourenço Kelmer, Bairro São Pedro, CEP: 36036-900, Juiz de Fora, MG, Brazil

† Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente, 225 - Bairro Gávea, CEP: 22430-060, Rio de Janeiro, RJ, Brazil

Abstract. *This paper discusses and analyzes the performance of a technique for classifying possible faults (lack of lubrication, lack of adjustment and malfunction of a component) that can occur in an electromechanical switch machine, which is an equipment used for handling railroad switches. Aiming to better classify these faults, the type-1 and non-singleton fuzzy logic system trained by Hestenes and Stiefel's conjugate gradient method is presented. The performance of the proposal is compared with four classifiers reported in the literature (Bayes based,* *multilayer perceptron neural network, type-1 and singleton fuzzy logic system trained by steepest descent method and type-1 and singleton FLS trained by the Conjugate Gradient method), demonstrating higher accuracy and faster convergence speed. Based upon the attained results, the proposed model enables the railway company to adopt solutions to achieve operational excellence.*

Keywords: *Fuzzy Logic Systems, Non-singleton, Conjugate Gradient Method*

1 INTRODUCTION

A railroad switch is an electromechanical equipment that allows railway trains to be guided from one track to another, such as at a railway junction. A Switch Machine (SM) is an equipment used for handling railroad switches. Among all possible faults that can occur in an electromechanical switch machine, the three main ones are: lack of lubrication, lack of adjustment and malfunction of a component. This become a more complicated problem due to the increase of Brazilian railroad sector, which results in more use of SM's as well as its real time operation, monitoring, and diagnosis become of vital importance, especially with respect to predictive maintenance.

Researches on SMs have been in evidence in recent years, since they are important equipments that must be constantly monitored to avoid accidents. For example, [1] describes a strategy and a technical architecture for prognosis and health management (PHM) of a SM. Feature extraction techniques and Principal Component Analysis (PCA) have been used in the assessment of the machines' health. In turn, [2] makes use of a Support Vector Machine (SVM) in an online SM condition monitoring system based on current measurements trends to detect faults at their earliest stage or even before they occur. Wavelet Transforms and SVMs are used in [3], which also shows that electrical active power can be used as a parameter for condition monitoring systems for SMs. This technique has shown a high degree of accuracy. Also [4] applies expert system for fault identification in a SM, when the failure modes are not easily separable. A technique for the detection of gradual failure in a SM is described in [5]. Data were collected from tests on a common point mechanism and a Kalman Filter was used for pre-processing those data. Other works such as [6] and [7] present alternative techniques related to unsupervised and semi-supervised fault detection and identification in industrial plants based on the analysis of the control and error signals. A common aspect of all previous works is the focus on the detection of the existence of a fault, and not on classifying different types of faults, such as those contemplated in this work. Classification of specific types of faults (lack of lubrication, lack of adjustment and malfunction of a component) and the normal condition by monitoring the motor current of switch machine in the railroad was introduced in the literature by Aguiar *et al*. [8] [9]. The authors in [8], using the same data set, showed that the type-1 and singleton FLS trained by the Conjugate Gradient method can offer higher convergence speed and classification ratio than the Steepest Descent method.

This paper discusses and analyzes the performance of a technique for classifying typical kind of faults (lack of lubrication, lack of adjustment and malfunction of a component) that can occur in an electromechanical switch machine, which is a equipment used for handling railroad switches. To do so, we introduce the type-1 and non-singleton fuzzy logic system trained by the Hestenes and Stiefel's conjugate gradient method (i.e., 2nd-order information). Combinations of feature extraction techniques based on higher-order statistics, feature selection technique based on Fisher's discriminant ratio, and four classifiers (Bayes based [10], multilayer perceptron neural network [11], type-1 and singleton FLS [12] trained by steepest descent method and type-1 and singleton FLS [8] trained by the Conjugate Gradient method method) show the effectiveness of proposed technique when a dataset, which addresses the possible faults in SM, is considered. The dataset, which is constituted by measures of faults and normal conditions of SM and provided by company in the Brazilian railway sector, named MRS Logística S.A. (`https://www.mrs.com.br/`), is a part of a multi-years project that has the objective of introducing this functionality in 100% of 624 SMs belonging to this company. Additionally, the reported results show that the type-1 and non-singleton FLS trained by the Hestenes and Stiefel's conjugate gradient method can offer higher convergence rate and performance for a limited number of epochs than the type-1 and singleton FLS. Finally, but not the least, based upon the attained results, the proposed technique enables the railway company to adopt solutions to achieve operational excellence.

2 PROBLEM FORMULATION

Among all possible faults that can occur in an electromechanical switch machine, the three main ones are lack of lubrication, lack of adjustment and malfunction of a component. The system developed in this work is able to identify those faults by monitoring the current of the SM motor. The system makes it possible to reduce the impact on trains operation and on preventive maintenance, since interventions shall be carried out only when deviations are observed in the equipment.

The system responsible for identifying the type of failure shall:

- reduce the number of unproductive hours in maintenance, due to the knowledge of failures before moving the maintenance team to the location;
- reduce the number of recurrent preventive interventions;
- increase productivity of rail operations, given the reduction in frequency and time of operational maintenance;
- reduce the TST (Time Stopped Train) index through the increase of the Mean Time Between Failures (MTBF).

Let $\mathbf{x} \in \mathbb{R}^{N\times 1}$ a vector constituted by a sample signal of current of SM motor. Fig. 1 shows the paradigm used for the classification of events. In the block "Feature Extraction", $\mathbf{p}_l$, $\mathbf{p}_a$, $\mathbf{p}_c$ and $\mathbf{p}_n$ refer respectively to lubrication, adjustment, component and normal operation feature vectors. The block "Feature Selection" provides selected features $\mathbf{K}_{p_l}$, $\mathbf{K}_{p_a}$, $\mathbf{K}_{p_c}$ and $\mathbf{K}_{p_n}$ from $\mathbf{p}_l$, $\mathbf{p}_a$, $\mathbf{p}_c$ and $\mathbf{p}_n$, respectively. The block "Classification" applies one of the classification techniques to obtain the output vector $\mathbf{s}$, thereby identifying the type of fault.

The classification of events in the component $\mathbf{x}$ can be formulated as a simple decision between hypotheses related to the occurrence of the events covered in this work, as shown

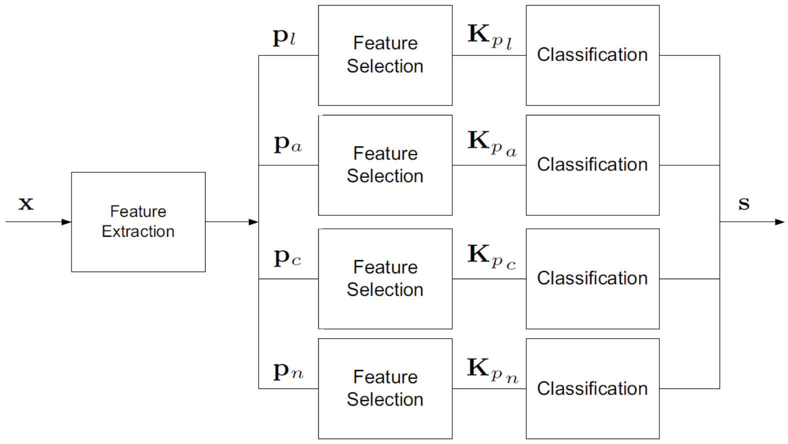

Figure 1: Block diagram of the scheme for classification of events.

below:

$$\begin{aligned}
&\mathcal{H}_{x,0} : \mathbf{x} = \mathbf{x}_{lub}, \\
&\mathcal{H}_{x,1} : \mathbf{x} = \mathbf{x}_{adj}, \\
&\mathcal{H}_{x,2} : \mathbf{x} = \mathbf{x}_{comp}, \\
&\mathcal{H}_{x,3} : \mathbf{x} = \mathbf{x}_{norm}.
\end{aligned} \tag{1}$$

The vectors $\mathbf{x}_{lub}$, $\mathbf{x}_{adj}$, $\mathbf{x}_{comp}$ and $\mathbf{x}_{norm}$ denotes the lubrication, adjustment, component and normal operation events, respectively. Four independent classifiers are developed, one for each possible event, since multiple events can occur simultaneously. The characteristics of each block in Fig. 1 will be presented in the following sections.

2.1 Feature extraction based on Higher-Order Statistics

Some contributions as [13], related to problems of detection, classification and identification of disturbances in electrical systems, have reported significant results obtained through the use of higher-order statistics (HOS). This is because HOS-based techniques are better suited to non-Gaussian processes and nonlinear systems, when compared to those that employ second-order statistics.

Given a stationary random process $\{z[n]\}$, such that $E\{z[n]\} = 0$, the cumulants of the second, third and fourth order can be computed as [14]:

$$c_{2,z}[i] = E\{z[n]z[n+i]\}, \tag{2}$$

$$c_{3,z}[i] = E\left\{z[n]z^2[n+i]\right\}, \tag{3}$$

and

$$c_{4,z}[i] = E\left\{z[n]z^3[n+i]\right\} - 3c_{2,z}[i]c_{2,z}[0], \tag{4}$$

where $E\{\cdot\}$ denotes the expected value operator and i is the ith lag.

Assuming that $\{z[n]\}$ is an L-length random process, then equations (2) - (4) can be stochastically approximated by the following expressions

$$\hat{c}_{2,z}[i] \cong \frac{2}{L}\sum_{n=0}^{L/2-1} z[n]z[n+i], \tag{5}$$

$$\hat{c}_{3,z}[i] \cong \frac{2}{L}\sum_{n=0}^{L/2-1} z[n]z^2[n+i], \tag{6}$$

and

$$\hat{c}_{4,z}[i] \cong \frac{2}{L}\sum_{n=0}^{L/2-1} z[n]z^3[n+i] - \frac{12}{L^2}\sum_{n=0}^{L/2-1} z[n]z[n+i]\sum_{n=0}^{L/2-1} z^2[n], \tag{7}$$

where $i = 0, 1, ..., L/2 - 1$.

An alternative approach would be [15]:

$$\tilde{c}_{2,z}[i] \cong \frac{1}{L}\sum_{n=0}^{L-1} z[n]z\left[\text{mod}(n+i, L)\right] \tag{8}$$

$$\tilde{c}_{3,z}[i] \cong \frac{1}{L}\sum_{n=0}^{L-1} z[n]z^2\left[\text{mod}(n+i, L)\right] \tag{9}$$

and

$$\tilde{c}_{4,z}[i] \cong \frac{1}{L}\sum_{n=0}^{L-1} z[n]z^3\left[\text{mod}(n+i, L)\right] - \frac{3}{L^2}\sum_{n=0}^{L-1} z[n]z\left[\text{mod}(n+i, L)\right]\sum_{n=0}^{L-1} z^2[n], \tag{10}$$

where $i = 0, 1, ..., L - 1$ and $\text{mod}(\cdot)$ is the modulus operator.

Thus, for each $\mathbf{s}$ it is possible to obtain a feature vector given by

$$\mathbf{p}_{\mathcal{H}} = [\hat{\mathbf{c}}_{2,z}^T\ \tilde{\mathbf{c}}_{2,z}^T\ \hat{\mathbf{c}}_{3,z}^T\ \tilde{\mathbf{c}}_{3,z}^T\ \hat{\mathbf{c}}_{4,z}^T\ \tilde{\mathbf{c}}_{4,z}^T]^T, \quad \mathcal{H} = 0, 1, \tag{11}$$

where $\mathcal{H} = 0$ is the class without disturbance, whereas $\mathcal{H} = 1$ is a class with disturbance, $\widehat{\mathbf{c}}_{2,z} = [\hat{c}_{2,z}(0), \cdots,$
$\hat{c}_{2,z}(L/2-1)]^T$, $\widetilde{\mathbf{c}}_{2,z} = [\tilde{c}_{2,z}(0), \cdots, \tilde{c}_{2,z}(L-1)]^T$, $\widehat{\mathbf{c}}_{3,z} = [\hat{c}_{3,z}(0), \cdots, \hat{c}_{3,z}(L/2-1)]^T$, $\widetilde{\mathbf{c}}_{3,z} = [\tilde{c}_{3,z}(0), \cdots, \tilde{c}_{3,z}(L-1)]^T$, $\widehat{\mathbf{c}}_{4,z} = [\hat{c}_{4,z}(0), \cdots, \hat{c}_{4,z}(L/2-1)]^T$ and $\widetilde{\mathbf{c}}_{4,z} = [\tilde{c}_{4,z}(0), \cdots, \tilde{c}_{4,z}(L-1)]^T$.

2.2 Feature selection based on Fisher's discriminant ratio

Recent works, such as [13], have used Fisher's discriminant ratio (FDR) for feature selection, since it is simple and provides satisfactory results. In this methodology, K_p features are selected in order to build the vector $\mathbf{p}_{\mathcal{H}}$ with length $\frac{9L}{2}$, due to $\hat{\mathbf{c}}_{2,z}$, $\hat{\mathbf{c}}_{3,z}$ and $\hat{\mathbf{c}}_{4,z}$ having length $L/2$ and $\tilde{\mathbf{c}}_{2,z}$, $\tilde{\mathbf{c}}_{3,z}$, $\tilde{\mathbf{c}}_{4,z}$ having length L. This makes the classifier complexity quite low. This pre-processing step is always executed before the classification stage. For a problem involving only two distinct classes associated with vectors with elements normally distributed and uncorrelated, it is defined as [10]:

$$\mathbf{F}_{FDR} = \Lambda_{\mu_0,\mu_1}\Lambda_{\sigma}^{-1}, \tag{12}$$

where $\Lambda_{\sigma} = diag\{\sigma_{0,0}^2 + \sigma_{1,0}^2, \sigma_{0,1}^2 + \sigma_{1,1}^2, \ldots, \sigma_{0,\frac{9L}{2}-1}^2 + \sigma_{1,\frac{9L}{2}-1}^2\}$ is a diagonal matrix formed by the vector covariance associated with each class, and $\Lambda_{\mu_0,\mu_1} = diag\{(\mu_{0,0} - \mu_{1,0})^2, (\mu_{0,1} - \mu_{1,1})^2, \ldots, (\mu_{0,\frac{9L}{2}-1} - \mu_{1,\frac{9L}{2}-1})^2\}$ is the diagonal matrix composed by their average vectors.

Note that $\mathbf{v}_{FDR} \in R^{\frac{9L}{2}\times 1}$ is a vector of elements from the main diagonal of $\mathbf{F}_{FDR}$, such that $v_{FDR}(0) \geq v_{FDR}(1) \geq \ldots \geq v_{FDR}(\frac{9L}{2} - 1)$. The K_p selected features, corresponding to the K_p first elements of the vector $\mathbf{v}_{FDR}$, will form vectors $\mathbf{K}_{p_l}$, $\mathbf{K}_{p_a}$, $\mathbf{K}_{p_c}$ or $\mathbf{K}_{p_n}$.

2.3 Classifiers

The classifiers used in this work are described as follows.

Bayes Classifier

Consider two possible classes $\mathcal{H}_0$ e $\mathcal{H}_1$ and a vector $\mathbf{x}_p \in \mathbb{R}^{K_p\times 1}$ formed by K_p parameters $\mathbf{p}_{\mathcal{H}}$, obtained through Equation (12). This vector **x** has *a priori* probability to be classified into one of two classes, given by $P(\mathcal{H}_0)$ and $P(\mathcal{H}_1)$. The conditional likelihood functions are denoted by $p(\mathbf{x}|\mathcal{H}_0)$ and $p(\mathbf{x}|\mathcal{H}_1)$. According to [10], Bayes rule gives:

$$P(\mathcal{H}_0|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_0)P(\mathcal{H}_0)}{p(\mathbf{x})} \tag{13}$$

and

$$P(\mathcal{H}_1|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_1)P(\mathcal{H}_1)}{p(\mathbf{x})}, \tag{14}$$

Assuming now that the *a priori* probability of both classes occurring is equal, i.e. $P(\mathcal{H}_0) = P(\mathcal{H}_1) = 1/2$, and that the likelihood function has a Gaussian distribution, the following equation applies:

$$\frac{|\sum_0|^{\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)}}{|\sum_1|^{\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\mathbf{x}-\boldsymbol{\mu}_0)}} \gtreqless 1 \tag{15}$$

The proposed classifier is based on the Maximum Likelihood criterion and, given a vector **x**, the determination of which class it corresponds to will depend on the result of this inequality.

MLP Neural Network

Given vectors $\mathbf{x}_{p,i}$, with $i = 1, 2, ..., N_x$ samples, constituted by K_p extracted features from vectors $\mathbf{p}_{x,i}$, then the equations of an MLP neural network with K_p inputs, one hidden layer and one output layer, are the following:

$$\mathbf{s}_i = \mathbf{A}^T \begin{bmatrix} \mathbf{x}_{p,i} \\ 1 \end{bmatrix} \tag{16}$$

$$\mathbf{q}_i = \varphi(\mathbf{s}_i), \tag{17}$$

and

$$y_{rn,i} = \varphi\left(\mathbf{B}^T \begin{bmatrix} \mathbf{q}_i \\ 1 \end{bmatrix}\right), \tag{18}$$

where $\mathbf{A} \in \mathbb{R}^{(K_p+1)\times N_q}$ and $\mathbf{B} \in \mathbb{R}^{(N_q+1)\times 1}$ represent the weights between the input and hidden layers and between the hidden and output layers, respectively; N_q is the number of processors in the hidden layer and $y_{rn,i}$ is the output of the MLP neural network associated to the vector $\mathbf{x}_{p,i}$. In addition, $\varphi(\cdot)$ is a monotonically increasing function, which is considered here as a hyperbolic tangent sigmoid.

By concatenating the column vectors of matrices $\mathbf{A}$ and $\mathbf{B}$, the following weight vector is obtained:

$$\mathbf{w} = [\mathbf{a}^T \; \mathbf{b}^T]^T, \tag{19}$$

where $\mathbf{a}^T$ and $\mathbf{b}^T$ are the results for the concatenation of column vectors of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. The optimal vector $\mathbf{w}$, given by ($\mathbf{w}_o$), can be obtained by solving:

$$\mathbf{w}_o = \min_{\mathbf{w}} J(\mathbf{w}), \tag{20}$$

where

$$J(\mathbf{w}) = \frac{1}{2N_x}\sum_{i=1}^{N_x}\left(y_{rn,i} - y_{d,i}\right)^2 \tag{21}$$

is the cost function, and $y_{d,i}$ is the ith desired output of the MLP neural network. Note that $y_{d,i} = 1$ means that the parameter vector $\mathbf{x}_{p,i}$ is associated with the occurrence of the event. On the other hand, $y_{d,i} = -1$ states that the parameter vector $\mathbf{x}_{p,i}$ is associated with the absence of the event.

The Levenberg-Marquardt (LM) training method [16] [17] has been adopted for MLP neural network.

3 TYPE-1 AND NON-SINGLETON FUZZY LOGIC SYSTEM CLASSIFIER

Considering non-singleton fuzzification, max-product composition, product implication, height defuzzifier and gaussian membership functions, the output of a type-1 and non-singleton FLS is [12]:

$$f_{ns}(\mathbf{x}) = \sum\nolimits_{l=1}^{M} \theta_l \phi_l(\mathbf{x}), \tag{22}$$

where $\phi_l(\mathbf{x})$ is called fuzzy basis function (FBF) [12], that is given by:

$$\begin{aligned}\phi_l(\mathbf{x}) &= \frac{\prod_{k=1}^{p}\mu_{F_k^l}(x_k)}{\sum_{l=1}^{M}\prod_{k=1}^{p}\mu_{F_k^l}(x_k)} \\ &= \frac{\prod_{k=1}^{p}\exp\left[-\frac{1}{2}\left(\frac{x_k^{(q)} - m_{F_k^l}}{\sigma_{F_k^l}^2 + \sigma_{x_{F_k^l}}^2}\right)^2\right]}{\sum_{l=1}^{M}\left\{\prod_{k=1}^{p}\exp\left[-\frac{1}{2}\left(\frac{x_k^{(q)} - m_{F_k^l}}{\sigma_{F_k^l}^2 + \sigma_{x_{F_k^l}}^2}\right)^2\right]\right\}},\end{aligned} \tag{23}$$

where $\mathbf{x} \in \mathbb{R}^{K_p}$ is the vector constituted by the features selected with FDR, $\prod$ denotes the product operator, $m_{F_k^l}$ and $\sigma_{F_k^l}$ are the mean and standard deviation associated to the k-th input feature of the l-th rule, and θ_l is the weight associated with the l-th rule, $l = 1, \ldots, M$. We refer to the standard deviation associated to each input membership function as σ_x. The subscript "ns" in $f_{ns}(\mathbf{x})$ makes it clear that this is a type-1 and non-singleton FLS.

Given a set of input-output pairs $(\mathbf{x}^{(q)} : y^{(q)})$, where q denotes the qth iteration and each iteration consists of the presentation of all samples, the problem consists, in fact, of minimizing the following cost function [12]:

$$J(\mathbf{w}^{(q)}) = \frac{1}{2}\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right]^2. \tag{24}$$

The result is

$$\begin{aligned} m_{F_k^l}(q+1) = m_{F_k^l}(q) - \alpha\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \times \\ \cdots\left[\theta_l\left(q\right) - f_{ns}(\mathbf{x}^{(q)})\right]\left(\frac{x_k^{(q)} - m_{F_k^l}}{\sigma^2_{F_k^l} + {\sigma_x}^2_{F_k^l}}\right)\phi_l(\mathbf{x}^{(q)}) \end{aligned}, \tag{25}$$

$$\begin{aligned} \sigma_{F_k^l}(q+1) = \sigma_{F_k^l}(q) - \alpha\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \times \\ \cdots\left[\theta_l\left(q\right) - f_{ns}(\mathbf{x}^{(q)})\right]\sigma_{F_k^l}(q)\left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma^2_{F_k^l}(q) + {\sigma_x}^2_{F_k^l}(q)}\right]^2\phi_l(\mathbf{x}^{(q)}), \end{aligned} \tag{26}$$

$$\theta_l(q+1) = \theta_l(q) - \alpha\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right]\phi_l(\mathbf{x}^{(q)}), \tag{27}$$

and

$$\begin{aligned} \sigma_{xF_k^l}(q+1) = \sigma_{xF_k^l}(q) - \alpha\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \times \\ \cdots\left[\theta_l\left(q\right) - f_{ns}(\mathbf{x}^{(q)})\right]\sigma_{xF_k^l}(q)\left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma^2_{F_k^l}(q) + {\sigma_x}^2_{F_k^l}(q)}\right]^2\phi_l(\mathbf{x}^{(q)}) \end{aligned}. \tag{28}$$

Note that α is the stepsize used to update parameters from type-1 and non-singleton FLS.

3.1 Hestenes and Stiefel's Conjugate Gradient method for the type-1 and non-singleton FLS

Optimization based on the conjugate gradient was originally proposed in [18]. The initial goal was to solve an unrestricted quadratic problem, but it was soon extended to more general cases. This method can be considered from two points of view: as a method of descent with exact line search or as a method of solving linear systems based on a process of orthogonalization.

From (24), it is known that $\mathbf{w}^{(q)}$ and $\nabla J(\mathbf{w}^{(q)})$ denote, respectively, the parameters and the gradient vector of a type-1 and non-singleton FLS (see (25) to (28)) given by:

$$\begin{aligned}&\mathbf{w}^{(q)} = \\ &\Big[m_{F_1^1}(q), \ldots, m_{F_P^1}(q), \ldots, m_{F_1^M}(q), \ldots, m_{F_P^M}(q), \\ &\sigma_{F_1^1}(q), \ldots, \sigma_{F_P^1}(q), \ldots, \sigma_{F_1^M}(q), \ldots, \sigma_{F_P^M}(q), \\ &\theta_1(q), \ldots, \theta_M(q), \\ &\sigma_{xF_1^1}(q), \ldots, \sigma_{xF_P^1}(q), \ldots, \sigma_{xF_1^M}(q), \ldots, \sigma_{xF_P^M}(q),\Big]^T\end{aligned} \tag{29}$$

and

$$\begin{aligned}&\nabla J(\mathbf{w}^{(q)}) = \\ &\Big[\nabla_{m_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{m_{F_P^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{m_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{m_{F_P^M}(q)} J(\mathbf{w}^{(q)}), \\ &\nabla_{\sigma_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\sigma_{F_P^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \ \nabla_{\sigma_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\sigma_{F_P^M}(q)} J(\mathbf{w}^{(q)}), \\ &\nabla_{\nabla_{\theta_1(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\theta_M(q)}(q)} J(\mathbf{w}^{(q)}), \ldots, \ \nabla_{\sigma_{x_{F_1^1}}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\sigma_{x_{F_P^1}}(q)} J(\mathbf{w}^{(q)}), \ldots, \\ &\nabla_{\sigma_{x_{F_1^M}}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\sigma_{x_{F_P^M}}(q)} J(\mathbf{w}^{(q)})\Big]^T.\end{aligned} \tag{30}$$

The first-order derivative of $J(\mathbf{w}^{(q)})$ with regard to parameters $m_{F_k^l}(q)$, $\sigma_{F_k^l}(q)$ and $\theta_l(q)$, are:

$$\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \times \left[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})\right] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q), \tag{31}$$

$$\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \times \left[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})\right] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q), \tag{32}$$

$$\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)}) = \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \phi_l(\mathbf{x}^{(q)}), \tag{33}$$

$$\nabla_{\sigma_{x_{F_k^l}}(q)} J(\mathbf{w}^{(q)}) = \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right] \times \left[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})\right] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q), \tag{34}$$

where $l = 1, \ldots, M$ e $k = 1, \ldots, p$. Note that

$$a_{F_k^l}(q) = \frac{x_k^{(q)} - m_{F_k^l}}{\sigma^2_{F_k^l} + {\sigma_x}^2_{F_k^l}}, \tag{35}$$

$$b_{F_k^l}(q) = \sigma_{F_k^l}(q) \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma^2_{F_k^l}(q) + {\sigma_x}^2_{F_k^l}(q)}\right]^2 \tag{36}$$

and

$$c_{F_k^l}(q) = \sigma_{xF_k^l}(q) \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma^2_{F_k^l}(q) + {\sigma_x}^2_{F_k^l}(q)}\right]^2 \tag{37}$$

The equations for training type-1 and non-singleton differ from those applied to type-1 and singleton due to the presence of σ_x, which model the uncertainty associated with input.

The algorithm that summarizes the use of a Conjugate Gradient method for training the type-1 and non-singleton FLS consists of the steps presented in Fig. 1, where $\alpha \in \mathbb{R}\,|0 < \alpha \leq 1$ is the stepsize. β^{HS} is Hestenes and Stiefel's scale factor [18],[8] and is given by:

$$\beta^{HS}(q) = \frac{\mathbf{g}^T(q+1)\zeta(q)}{\mathbf{d}^T(q)\zeta(q)}, \tag{38}$$

where $\zeta(q) = \mathbf{g}(q+1) - \mathbf{g}(q)$. The index HS refers to Hestenes and Stiefel;

Aiming at a better performance of the training algorithm, $\alpha(q)$ and $\beta^{HS}(q)$ have been restricted to previously tested values, being $\alpha = 10^{-4}$ and $\beta_{\max} = 5$.

Input
l: Number of rules of the type-1 and non-singleton FLS;
k: Number of parameters of the type-1 and non-singleton FLS;
η: Step-size;
$\beta_{\max}$: Restrictive value for β^{HS};
$\alpha_{\max}$: Restrictive value for α;
nn: Number of epochs;
$nn_{\max}$: Maximum number of epochs;
N: Number of samples;
p: Number of extracted parameters;
Output
$performance$: Performance of type-1 and non-singleton FLS;
begin
  while $nn_{\max} > nn$ **do**
    for q = *1* ***to*** N **do**
      for $l = 1$ ***to*** M **do**
        for $k = 1$ ***to*** p **do**
          To calculate equations (22), (31), (32), (33) and (34)
        end
      end
      if $y(\mathbf{x}^{(q)}) > 0$ **then**
        $y^{(q)} = 1$
      else
        $y^{(q)} = -1$
      end
      To calculate equations (29) and (30)
      if q = *0* **then**
        $g(0) = -\nabla J(\mathbf{w}^{(0)})$
        $d(0) = -g(0)$
      else
        $g(q) = -\nabla J(\mathbf{w}^{(q)})$
        $d(q) = -g(q)$
        $\beta^{HS}(q) = \frac{\mathbf{g}^T(q+1)\zeta(q)}{\mathbf{d}^T(q)\zeta(q)}$
        if $0 < \beta^{HS}(q) < \beta_{\max}(q)$ **then**
          $\mathbf{d}(q+1) = -\mathbf{g}(q+1) + \beta^{HS}(q)\,\mathbf{d}(q)$
        else
          $\mathbf{d}(q+1) = -\mathbf{g}(q+1)$
        end
      end
      To update $m_{F_k^l}(q)$, $\theta_l(q)$, $\sigma_{F_k^l}(q)$ and $\sigma_{x\,F_k^l}(q)$ from $\mathbf{w}^{(q+1)}$
    end
    $performance\,(\%) = (type, nn) = 100 \times (find\,(y_d - y) == 0)/N$
  end
end

Algorithm 1: Type-1 and non-singleton FLS trained by Hestenes and Stiefel's Conjugate Gradient algorithm that makes use of equation (38).

4 EXPERIMENTAL RESULTS

The performance analysis made use of measured data provided by MRS Logística, which the signal consists of current signal [A] against time sampled through four channels of an industrial data acquisition board. The sampling rate was of 100 Hz in a period of 2 seconds. The acquired signals fall within the following four classes: lack of lubrication, lack of adjustment, malfunction of a component and normal operation. As normal operation occurred mostly, the original data had to be balanced. Eventually, there were 1376 signals for each of the four classes. In addition, two outliers were removed for lubrication class, resulting in only 0.145% of incorrect signals. Due to the low number of outliers, no evaluation was performed without the outlier removal step. No tests were performed including outliers in the training stage of the method. A sample example of the signals can be seen in Fig. 2.

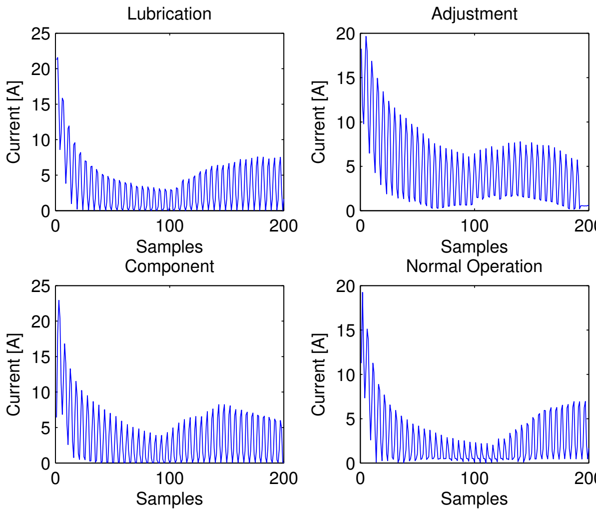

Figure 2: Signals referring to used classes.

The Fig. 3 shows an example of values obtained for the FDR for the event lubrication. In this case, five parameters with the highest values of FDR were selected. After evaluating the performance with $K_p = 1, 2, 3, 4$ and 5 features extracted from FDR, we chose $K_p = 3$, since it provides the best performance in terms of accuracy and convergence speed. Therefore, all other available features (12,384) were not considered in the classification block.

The final configuration of the type-1 and non-singleton FLS is composed of four rules ($M = 4$), two rules for the class that has the presence of the event and another two for the class that does not have the presence of the event. The membership functions parameters were defined heuristically from the calculation of means and variances of the feature vectors coefficients. The MLP neural network has $K_p = 3$ inputs, $N_q = 1$ hidden layer with two processors and an output layer with one output. Different configurations of the hidden layer were also evaluated, but the best performance was obtained with two processors.

The experiments with the type-1 and singleton FLS and with the MLP neural network considered 100 epochs for training. In the case of the FLS, the samples were divided equally between training and test sets. In the case of the MLP neural network, 50% of the samples were used for training, 25% for validation and 25% for testing.

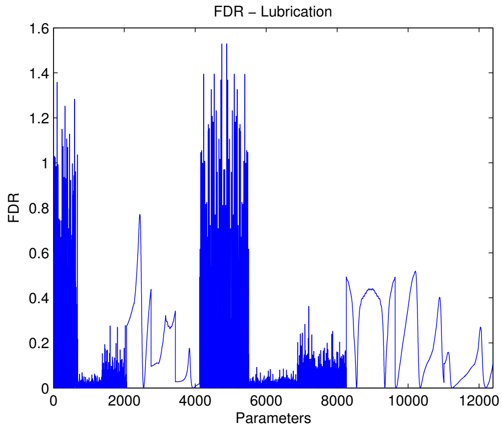

Figure 3: FDR obtained for lubrication.

4.1 Convergence Analysis

Simulations were performed with the algorithms for classification of faults assuming the maximum number of epochs equals to 100. Figures 4, 5, 6 and 7 illustrate the convergence of the FLS for each of the four classes. To facilitate the understanding acronyms are adopted. Thus, we have the following situations: **HS T1FLS NS** refers to type-1 and non-singleton FLS trained by the Hestenes and Stiefel's conjugate gradient method; **SD T1FLS NS** refers to type-1 and non-singleton FLS trained by the Steepest Descent method; **HS T1FLS S** refers to type-1 and singleton FLS trained by the Hestenes and Stiefel's conjugate gradient method [8] and **SD T1FLS S** refers to type-1 and singleton FLS trained by the Steepest Descent method [9]. It can be seen that the both HS T1FLS NS and SD T1FLS NS provide faster convergence ratio than the previous methods reported in the literature.

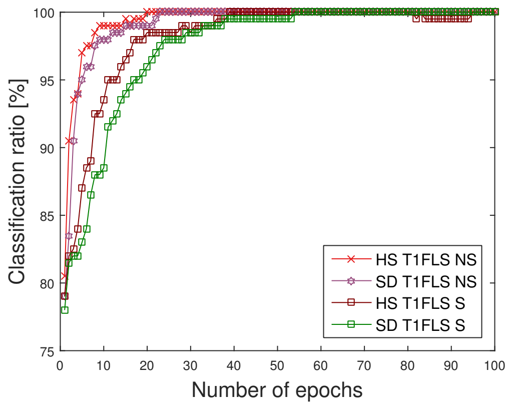

Figure 4: Convergence of type-1 and non-singleton FLS for *lubrication*.

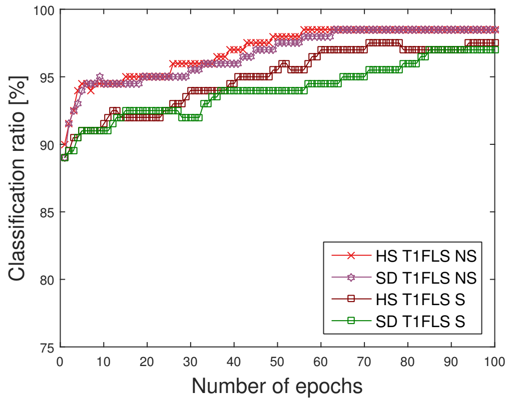

Figure 5: Convergence of type-1 and non-singleton FLS for *adjustment*.

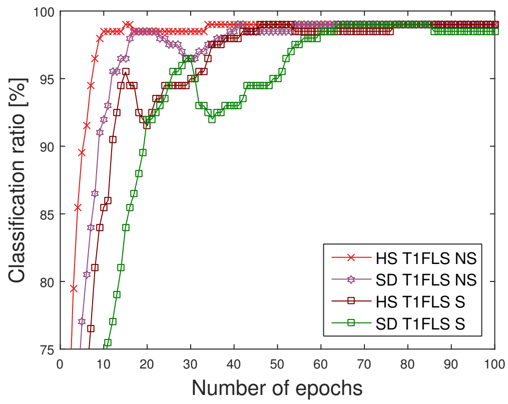

Figure 6: Convergence of type-1 and non-singleton FLS for *component*.

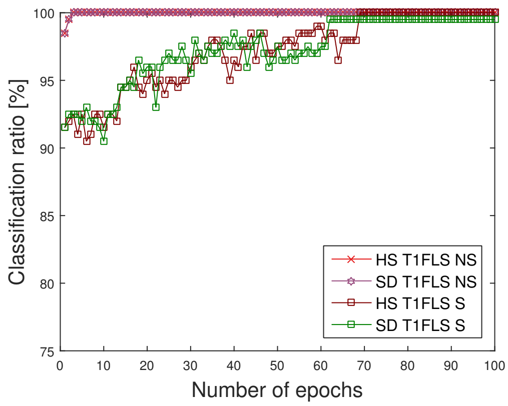

Figure 7: Convergence of type-1 and non-singleton FLS for *normal operation*.

4.2 Performance Analysis

The final results for all techniques are shown in Table 1. The efficiency ρ is the arithmetic mean of the classifiers performances for all events. We can note that the HS T1FLS NS is promising, since it outperformed the techniques based on Bayes theory [9], SD T1FLS S [9] and HS T1FLS S [2]. Moreover, it achieves a similar performance in [%] in comparison with multilayer perceptron neural network [9].

Table 1: Comparative performance in (%) of the classifiers adopted during the test phase.

Events	Bayes [9]	MLP [9]	SD T1FLS S [8]	HS T1FLS S [8]	SD T1FLS NS	HS T1FLS NS
Lack of lubrication	94.5	100	100	100	100	100
Lack of adjustment	98	100	97.5	97.5	98.5	99
Component malfunction	99	98.3	98.5	99.5	98.5	99.5
Normal Operation	98.5	100	99.5	100	100	100
Efficiency **(ρ)**	**97,5**	**99,6**	**98,9**	**99,3**	**99,3**	**99,6**

CONCLUSION

This work discussed the use of signal processing and computational intelligence techniques in the classification of faults in SM. The extraction of HOS-based features and the selection of reduced set of features by FDR have proved to be relevant and have contributed to a reduction in the dimensionality of the data to be presented to the classifiers. Additionally, Hestenes and Stiefel's Conjugate Gradient method for training type-1 and non-singleton FLS has been introduced.

Results obtained through the use of real measured data revealed that the type-1 and non-singleton FLS tuned by the Hestenes and Stiefel's Conjugate Gradient method converges faster and classification rates are better for a reduced number of epochs. All classifiers are attractive options due to their mathematical simplicity.

Future works will be directed to the design of a prototype to be integrated into the existing SM supervision system in the company MRS Logística S.A.. It will be considered the development of an approach to exclude outliers and the use of type-2 FLS in order to achieve better results.

ACKNOWLEDGEMENTS

The authors would like to thank FAPEMIG and MRS Logística S.A. for supporting this work.

REFERENCES

[1] Ardakani, H.D., Lucas, C., Siegel, D., Chang, S., Dersin, P., Bonnet, B., Lee, J.: PHM for railway system − A case study on the health assessment of the point machines. 2012 IEEE Conference on Prognostics and Health Management (PHM). 1−5, Jun.(2012)

[2] Vileiniskis, M., Remenyte-Prescott, R., Rama, D., Andrews, J.: Fault Diagnostics for Railway Point Machines. 20th Advances in Risk and Reliability Technology Symposium (AR2TS), At Loughborough, UK. 1−5, May (2013)

[3] Asada, T., Roberts, C.: Development of an effective condition monitoring system for AC point machines. IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 2009. SDEMPED 2009. 1−5, Aug. (2009)

[4] Atamuradov, V., Camci, F., Baskan, S., Sevkli, M.: Failure diagnostics for railway point machines using expert systems. 5th IET Conference on Railway Condition Monitoring and Non-Destructive Testing (RCM 2011). 1−6, Nov. (2011)

[5] Márquez, F. P. G., Schmid, F., Collado, J.C.: A reliability centered approach to remote condition monitoring. A railway points case study. Reliability Engineering & System Safety. 80(1), 33−40 (2003)

[6] Costa, B. S. J., Angelov, P. P., Guedes, L. A.: Real-Time Fault Detection Using Recursive Density Estimation. Journal of Control, Automation and Electrical Systems. 25(4), 428−437 (2014)

[7] Costa, B. S. J., Angelov, P. P., Guedes, L. A.: Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier. Neurocomputing. 150(A), 289−303 (2015)

[8] de Aguiar, Eduardo P, ”EANN 2014: a fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine.” Neural Computing and Applications 27.5 (2016): 1175-1189.

[9] E. Aguiar, ”Classification of events in switch machines using Bayes, fuzzy logic system and neural network,” in Proc. 15th Int. Conf. Eng. Appl. Neural Netw., Sep. 2014, pp. 81-91

[10] Theodoridis, S. and K. Koutroumbas, K.: Pattern Recognition. Academic Press, San Diego (1999)

[11] Haykin, S.: Neural Networks. A Comprehensive Foundation. Prentice Hall (1999)

[12] Mendel, J.M.: Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. Prentice Hall PTR (2001)

[13] de Aguiar, E.P., Marques, C.A.G., Duque, C.A., Ribeiro, M. V.: Signal decomposition with reduced complexity for classification of isolated and multiple disturbances in electric signals. IEEE Trans. on Power Delivery. 24(4), 2459-2460 (2009)

[14] Mendel, J.M.: Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. Proceedings of the IEEE. 79(3), 278-305 (1991)

[15] Ribeiro, M.V., Marques, C.G., Cerqueira, A.S., Duque, C.A., Pereira, J.L.R.: Detection of Disturbances in Voltage Signals for Power Quality Analysis Using HOS. EURASIP Journal on Advances in Signal Processing. 2007, 13-15 (2007)

[16] Hagan, M.T., Menhaj, M.-B.: Training feedforward networks with the Marquardt algorithm. IEEE Trans. on Neural Networks. 5(6), 989-993 (1994)

[17] Finschi, L.: An Implementation of the Levenberg-Marquardt Algorithm. ETH Zurich - Institute for Operations Research (1994)

[18] Hestenes, M. R., Stiefel, E.: Methods of Conjugate Gradients for Solving Linear Systems. Journal of Research of the National Bureau of Standards. 49(6), 409−436, Dec. (1952)