A Novel Fault Identifying Method with Supervised Classification and Unsupervised Clustering

Tao Xu
Department of Automation
Shenyang Aerospace University
China
xutao@sau.edu.cn

Journal of Digital
Information Management

ABSTRACT: *To satisfy the robust requirement when designing fault identifying method, this paper proposes a novel method to identify sensor fault. Conventional fault identifying method could only classify fault into explicit set. Yet, when a novel faulty pattern occurs, the conventional method can not identify this new pattern and will classify it into a set known ahead of time. For the purpose of robustness of fault identifying method, the supervised classification and the unsupervised clustering are integrated together. Faulty features are extracted with wavelet package decomposition to train the supervised classification algorithm and the unsupervised clustering algorithm. As the supervised classification method, Support Vector Machine (SVM) is utilized to classify those faulty patterns known ahead of time into explicit set for the purpose of identifying faulty pattern. As the unsupervised clustering method, subtractive clustering method is utilized to identify the novel fault pattern. Therefore, SVM and subtractive clustering are integrated to identify sensor fault pattern even when novel faulty pattern emerges. The applicability and effectiveness of the proposed method is illustrated by flow sensor data in an engine fuel providing system. The result shows that the method adopted provides better performance compared with conventional method while satisfying the robust requirement of fault identifying method.*

Categories and Subject Descriptors:
F.1 **[COMPUTATION BY ABSTRACT DEVICES]**: Self-modifying Machines; I.5.3 Clustering C.4 **[PERFORMANCE OF SYSTEMS]**: Fault Tolerance

General Terms:
Clustering, Support Vector Machines, Neural networks

Keywords: Fault Identify, Supervised Classification, Unsupervised Clustering, Support Vector Machine, Subtractive Clustering

Received: 19 December 2012, Revised 2 March 2013, Accepted 9 March 2013

1. Introduction

Many different methods have been proposed to identify fault in many areas. One of these methods, observer is more useful when the system mathematical model is known ahead of time. But in many applications, accurate system mathematic model is not available. Therefore, some methods without mathematical model have been proposed. Among these, pattern recognition is an effective method for fault identification [1] [2] [3]. Firstly, the method adopted with pattern recognition collects history data under normal condition and faulty condition. Then, the supervised decision making algorithm is trained with the feature samples that have been extracted from history data and belong to explicit set of each pattern [4]. Therefore, the algorithm possesses the capability to classify new samples into different sets to identify faulty pattern. Yet, fault identifying method based on pattern recognition should satisfy the robust requirement. Namely, it should identify the novel pattern which dislikes any faulty pattern known ahead of time when novel fault occurs [5]. [6] and [7] propose fault identifying methods with SVM and neural network respectively. But, they do not satisfy the robust requirement to identify the novel faulty pattern. The proceeding methods could not satisfy this competitive requirement that is focused in this paper.

Recently, SVM and Neural Network (NN) have become two important supervised classification methods [6] [7]. By the contrast with NN, SVM possesses better performance with higher accurate identifying rate especially when only few samples available [6]. So, SVM is chosen as the supervised classification method to identify fault patterns which are known ahead of time. To satisfy the robustness requirement of fault identifying method, subtractive clustering is integrated with SVM to propose a novel model. As a simple clustering analysis method, subtractive clustering is adopted to identify novel faulty pattern. To achieve feature samples, feature extraction method is proposed in this paper. Because wavelet package provides further decomposition for faulty signal in both coarse scale and fine scale, it extracts faulty feature accurately when fault occurs. Therefore, wavelet package is adopted to provide training samples for SVM and subtractive clustering.

This paper is organized as follows. Section II presents the theorem of wavelet package decomposition and proposes each step of feature extraction algorithm. Section III presents the theorem of SVM and proposes multi-classification for fault classification. In Section IV, we introduce the theorem of subtractive clustering. In Section V, we describe fault identifying model established and interpretation of its principle. Section VI shows how to apply this model into flow sensor fault identification in the engine fuel providing system. Finally, we present meaningful conclusion.

2. Feature Extraction with Wavelet Package Decomposition

Short-Time Fourier Transform (STFT) possesses linearity characteristic when decomposing the frequency domain. Though multi-resolution of wavelet transform could decompose signal into time-frequency domain, its scale varying in binary mode results into low frequency resolution in high frequency domain and low time resolution in low frequency domain. Wavelet package decomposition provides a more subtle analytical method that decomposes the high frequency domain even further. So, the time-frequency resolution is improved [8]. Figure 1 provides an illustration for three-layer wavelet package decomposition.

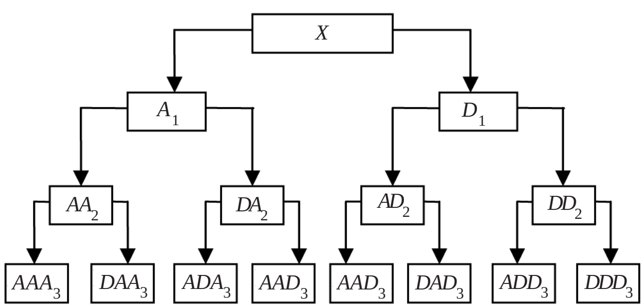


Figure 1. Architecture of three-lay Wavelet Package Decomposition

Here, A means low frequency, D means high frequency and the index number in the foot indicates the layer of wavelet package decomposition. The expression for decomposition is that $X = AAA_3 + DAA3 + ADA_3 + DDA_3 + AAD_3 + DAD_3 + ADD_3 + DDD_3$.

Wavelet package can decompose the signal into different scales. Then, different feature is achieved from each scale. Each step of the feature extraction algorithm is described below.

(1) Standardize sensor signal X

$$\overline{X} = D_{\sigma}^{-1}[X - E(X)] \quad (1)$$

where $E(X)$ is mean value of X and D_{σ}^{-1} is standard deviation of X. This step eliminates the influence of different sensor value.

(2) Decompose X into three layers with wavelet package with X_{30}, X_{31}, X_{32}, X_{33}, X_{34}, X_{35}, X_{36}, X_{37} to present each coefficient vector respectively.

(3) Process the vectors above with reduction algorithm as follow.

1) Calculate the reduction threshold of each vector with

$$Thr_{X_{3i}} = \sqrt{\frac{1}{k}\Sigma X_{3i}} \quad (2)$$

where k presents the length of X_{3i}.

2) Process the coefficients in X_{3i} with

$$cfs_{ij} = \begin{cases} cfs_{ij}, & |cfs_{ij}| \geq Thr_{X_{3i}} \\ 0, & |cfs_{ij}| < Thr_{X_{3i}} \end{cases} \quad (3)$$

Where cfs_{ij} presents j^{th} coefficient in i^{th} vector.

(4) Calculate total power of each vector with

$$E_{3i} = | S_{3i}|^2 dt = \sum | S_{3i}|^2 \quad (4)$$

(5) In order to facilitate classification, E_{3i} will be standardized as

$$\overline{E}_{3i} = \frac{E_{3i}}{\sum_{j=0}^{7} E_{3j}} \quad (5)$$

(6) Format the feature vector as $T = [\overline{E}_{30}\ \ \overline{E}_{31}\ \ \overline{E}_{32}\ \ \overline{E}_{33}\ \ \overline{E}_{34}\ \ \overline{E}_{35}\ \ \overline{E}_{36}\ \ \overline{E}_{37}]$.

3. Multi-classification with Support Vector Machine

SVM is regarded as the replacement of conventional classification methods, especially when small sample available and in nonlinear situation [9] [10]. It possesses better performance for generalization. SVM model is established based on Vapnik-Chervonenkis Dimension theory and structural risk minimization. It finds the best compromise between the model complexity and learning ability with finite samples information to achieve the best generation performance. As a new machine learning method, SVM has been successfully applied into the field of fault identification.

Suppose two class of training sample available.

$$(x_i, y_i) = 1,..., l \ x \in R^n, y \in \{+1, -1\} \quad (6)$$

Here, i is the number of samples, n is the number of dimensions and y is the identifier of classes. Nonlinear problem is converted into linear problem in higher-dimension space, in which the optimal classifier face is achieved. Base on theorem of functional, only when a sum-function satisfies Mercer condition, it corresponds to the inner product of a transformation space. So, an appropriate inner production function in the optimal classifier face implements linear classifier after nonlinear transformation. Radial Basis Function (RBF) is chosen in this paper.

$$K(x, x_i) = \exp\left\{\frac{\| x - x_i \|^2}{\sigma^2}\right\} \tag{7}$$

The general formation of the classifier face equation is below.

$$wx + b = 0 \tag{8}$$

Here, w is weight vector and b is classifier threshold. If all samples can be classified by this equation, it must satisfy the equation below.

$$y_i [(wx_i) + b] \geq 1 - \xi_i \qquad i = 1, ..., l \tag{9}$$

Here, ξ_i is relaxation factor.

Now, the distance between two classes equals $2 / \|w\|$. The classifier face which maximizes the distance is called optimal classifier class. To seek the optimal classifier hyper-face, quadratic programming problem with constraints should be solved.

$$\min \| w \|^2 / 2 + c \sum_{i=1}^{l} \xi_i \tag{10}$$

Large than 0, c is constant as specified and determines punishment to the samples that have been classified incorrectly. The problem above can be converted into dual formation as below.

$$\max \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j K(x, x_i) \tag{11}$$

Here, the constraint condition is that

$$\sum_{i=1}^{l} a_i y_i = 0, \quad 0 \leq a_i \leq C, \quad i = 1, 2, ..., l \tag{12}$$

Based on KKT condition, optimal parameter a_i satisfies

$$a_i \{y_i [(w, x_i) + b] - 1 + \xi_i\} = 0 \tag{13}$$

So, optimal parameters corresponding to most samples become 0. Only few of them are not be 0. Those corresponding samples are support vector. After the optimal parameter has been solved, discriminant function of SVM classifier is

$$y = sgn\left\{\sum y_i a_i K(x_i, x) + b\right\} \tag{14}$$

Finally, the classifier is determined by the signal of the discriminant function.

But, SVM is proposed for two group classification problem. Before applied into multi-fault identification, it must be appended with multi-classifiers. Generally, there are two solutions for this problem. One is one-time solution with multi-class objective function. Then seek the optimal solution of it. Many more variables will be used in the process when solving this optimization problem. So, it is often meaningless because of its complexity. The other one is to convert two-classification problem into multi-classification problem. This scheme includes one-against-rest and one-against-one which need N classifiers and $N(N-1)/2$ classifiers [11] [12]. When many more classifiers needed, the scheme need much more time in the process of learning. The most unacceptable issue is that there are unrecognized domains when utilizing the scheme. Based on binary classification tree, hierarchical SVM classifier only needs N–1 classifications without unrecognized domains [13]. It does not need to travel all of the SVM classifications with reasonable hierarchy SVM classifiers. Also, the learning speed is significantly improved. So, we utilize hierarchical SVM classifier to identify fault patterns.

To illustrate the SVM classification, four typical fault patterns of flow sensor in the fuel providing system of the engine are considered. Those fault patterns include bias, spike, drifting and cyclic failure. Besides of normal state, SVM will classify input feature samples into five patterns presented from SVM1 to SVM5 as described in figure 2.

Here, SVM1 classifies input feature vector into SVM2 and SVM3. SVM2 identifies bias failure and drifting failure. SVM3 identifies normal state and SVM4. Finally, SVM4 identifies spike failure and cyclic failure.

4. Subtractive Clustering

Subtractive clustering is a fast independent clustering method with similar characteristic [14] [15]. Each sample is the candidate clustering center, which is different from other clustering method. So, there is a linear relation between computation cost and number of samples.

Suppose n samples $(x_1, x_2,.., x_n)$ are in the M dimension space. Generally, all samples have been converted into unitary super cube space. Because each sample is the candidate clustering center, the density index of sample x_i is defined as follow.

$$D_i = \sum_{j=1}^{n} \exp\left[\frac{\| x_i - x_j \|^2}{(r_a / 2)^2}\right] \tag{15}$$

Here, r_a is a positive number. Obviously, if the sample considered has many neighbors, it possesses the high density index. Radius r_a defines the neighbor zone of this sample, and the samples outside of r_a contribute to the index much less than those inside of it.

After calculating the density index of each sample, the highest one will be chosen as the first clustering center. Suppose is the candidate sample, and is the density index. Then, the density index of each sample will be corrected as follow.

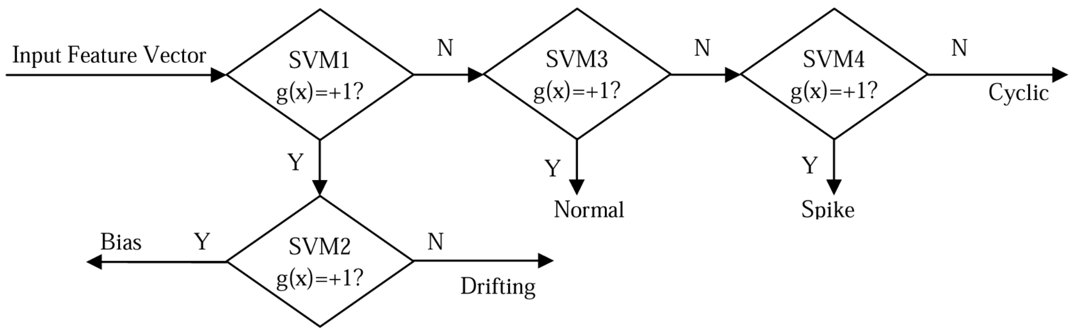


Figure 2. Faulty Patterns Classification with SVM

$$D_i = D_i - D_{c1} = \exp\left[\frac{\| x_i - x_{c1} \|^2}{(r_a/2)^2}\right] \qquad (16)$$

In the formula above, r_a is a positive number. Obviously, the density indexes of those samples nearby the first clustering centre will be subtracted dramatically. Hence, these samples would not become the next clustering center. Larger than r_a, r_b defines the zone where the density index function will be subtracted dramatically. To void clustering centers which have too short distance between each other, then, $r_b = 1.5 r_a$ generally. After the density index of each sample has been corrected, next clustering center is chosen. Then, correct the density function of each sample. With the repetition of operation above, this process ends after enough clustering centers have emerged.

First, analyze the training samples with subtractive clustering to determine the number of centers. In application, analyze the testing samples before classification with SVM to make sure whether new center emerges. If novel failure pattern occurs, the classification will identify false pattern without clustering analysis.

5. Fault Identifying Model with Supervised Classification and Unsupervised Clustering

Features extraction is implemented with wavelet package decomposition. After training SVM classification with these features samples, SVM classifies flow sensor fault into explicit pattern. Meanwhile, data clustering analysis is implemented by subtractive clustering. Firstly, analyze the training samples to determine the number of clustering centers. Before the classification with SVM, analyze the combination data set with training samples and testing samples to determine whether new center emerges. Novel fault identification is implemented by the contrast of the two clustering results. Then SVM which has been trained classifies testing data into different pattern to identify sensor fault. So, the model established satisfies the robust requirement when novel fault pattern emerges.

6. Illustration with Algorithm Simulation

The flow sensor in the fuel providing system of an engine is chosen to test the model proposed. Some typical sensor fault patterns are considered in this paper. Sensor fault can be divided into abrupt mode and incipient mode. Algorithm is simulated aiming at four typical senor fault patterns including drifting, bias, spike and cyclic failure. First, prepare 500 individuals of sensor data. Then extract feature from them with three layers of wavelet package decomposition. Figure 4(a) and figure 4(b) present the signal with cyclic failure both in time domain and in frequency domain respectively. Yet, fault pattern can not be identified by the figures because of the complexity of the cyclic failure.

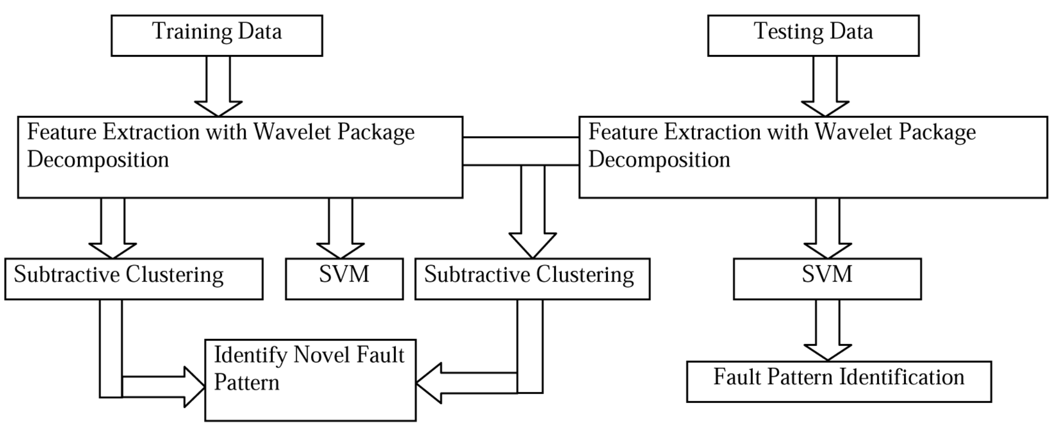
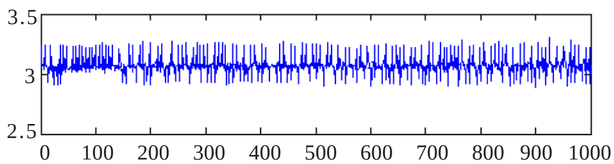


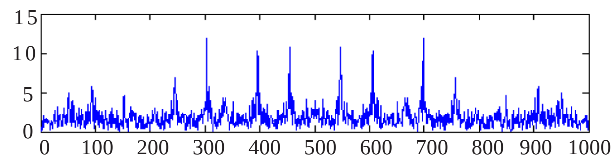
Figure 3. Model for Fault Identification with SVM and Subtractive Clustering

Because of the smoothness between abrupt points of faulty signal, it requires that wavelet function should possess more vanishing moments. Db10 has limited supported in time domain and its FFT transformation has N order zero-point where $\omega = 0$. So, db10 is chosen as wavelet base function in the process of extracting feature with wavelet package decomposition. Table 1 presents different patterns and the corresponding feature samples respectively.

20 groups are collected corresponding to each pattern. So, 100 groups are collected as training samples finally. As testing samples, 50 groups are collected to test the model proposed. From the SVM classification, different pattern of sensor fault can be identified as table 2.



(a) Signal in time domain



(b) Power spectrum of the signal

Figure 4. Cyclic Failures in Time Domain and Frequency Domain

Finally, the Multi-SVM classification successfully identifies 98% of sensor pattern which means the model proposed possesses the performance of identifying fault pattern.

To test the capability of the model for identifying novel fault pattern, another 20 groups of testing data with power failure have been simulated. Firstly, analyze the combination data set of training data and testing data with subtractive clustering method. To display the result conveniently, clustering result is represented with the first dimension and the last dimension of the clustering result as figure 5, in which there are 3 different clustering centers.

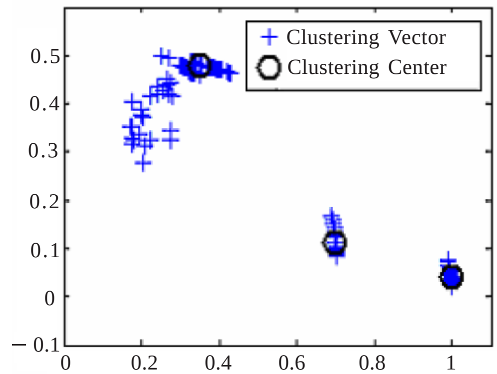


Figure 5. Result of the Original Testing Dataset after Subtractive Clustering

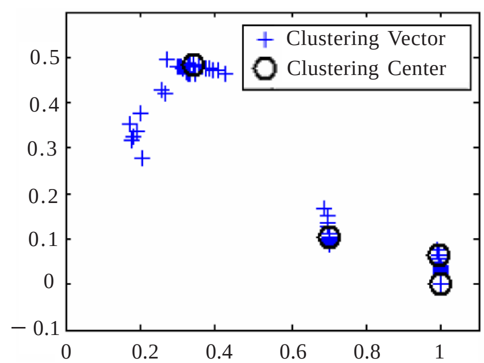


Figure 6 Result of the New Testing Dataset after Subtractive Clustering

Pattern	Feature
Normal state	[0.1768 0.2484 0.3192 0.1903 0.3647 0.5583 0.4725 0.3177]
Drifting failure	[0.4260 0.1496 0.2028 0.3679 0.5189 0.2851 0.2294 0.4652]
Bias failure	[0.9990 0.0074 0.0096 0.0181 0.0255 0.0140 0.0113 0.0229]
Spike failure	[0.7007 0.6817 0.0884 0.0683 0.0819 0.0929 0.0943 0.0875]
Cyclic failure	[0.3480 0.2354 0.2150 0.3908 0.5140 0.3099 0.2264 0.4616]

Table 1. Different Fault Patterns and Corresponding Feature

Testing Feature	Sensor Pattern
[0.1007 0.1415 0.1818 0.1084 0.2078 0.3180 0.2692 0.1810]	Normal
[0.21133 0.09875 0.1335 0.24299 0.34267 0.18836 0.15156 0.30719]	Drifting
[7.2433 0.099118 0.12947 0.24299 0.34267 0.1884 0.15163 0.30715]	Bias
[3.3308 3.1626 0.49087 0.40541 0.5532 0.51892 0.506 0.53814]	Spike
[0.2046 0.13001 0.13813 0.24933 0.34832 0.19992 0.15092 0.30736]	Cyclic

Table 2. SVM Classification Result of Testing Samples

7. Conclusion

A fault identifying method with the combination of supervised classification and unsupervised clustering has been proposed in this paper. It satisfies the robust requirement of fault identifying model for identifying novel fault pattern. Meanwhile, the effectiveness and accurateness of the model are illustrated by the flow sensor in the fuel providing system of an engine. The experimental result shows that the model has not only identified fault pattern known ahead of time, but also identified novel fault pattern that is not known ahead of time.

8. Acknowledgment

This work is supported by the Department of Education in Liaoning Province (Project NO. L2012045), China Postdoctoral Science Foudation (Project NO. 20100481013).

References

[1] Berners-Lee, T., Hendler, J. A., Lassila, O. (2001). The semantic web. *Scientific American*, 284 (5) 34–43, May.

[2] He Yong, Feng Lei. (2004). Diesel Fuel Injection System Faults Diagnosis based on Fuzzy Injection Pressure Pattern Recognition. Proceedings of the 5th World Congress on Intelligent Control and Automation, p. 1654-1657, Huangzhou, China, June.

[3] Diallo, Demba, Benbouzid, et al. (2005). Fault Detection and Diagnosis in an Induction Machine Drive: A Pattern Recognition Approach based on Concordia Stator Mean Current Vector. *IEEE Transactions on Energy Conversion*. 20 (3) 512-519.

[4] Dai Shuwu, Sun Huixian. (2005). Model based-Diagnosis and Fault Mode Identification for Spacecraft Power System. *Journal of Vibration and Shock*. 24 (3) 55-58.

[5] Tou, J. T., Gonzalez, R. C. (1974). *Pattern Recognition* Principles. MA: Addison-Welsley.

[6] Zhou Dong-hua, Ye Yin-chuan. (2000). Modern fault diagnosis and fault tolerant technology. p. 146-147, Beijing: Tringhua University Press.

[7] Sima Li-Ping, Huang Song-Bo, Dou Peng, et al. (2012). Probability estimation of interior fault position for power transformer based on SVM. *Power System Protection and Control*. 40 (14) 121-12.

[8] Zhang Song-He, Luo Yue-Gang, Wu Bin, et al. (2012). Fault diagnosis analysis of rotor system based on RBF neural network and dynamic systems. *Advanced Materials Research*. (460) 127-130.

[9] WICHEHAUSER, M. V. (1992). Lectures on wavelet package algorithims. Inormation Theory. 38, 713-718.

[10] Vapnik, V. (1995). The nature of statistics learning theory, p. 40-50, New York: Springer Verlag.

[11] Vapnik, V. (1998). Statistical learning theory, p. 102-110, New York: John Wiley and Sons.

[12] Bottou, L., Cortes, C., Denker, J. et al. (1994). Comp arison of Classifier Methods: A Case Study in Handwriting Digit Recognition. *In*: Proceeding of international conference on IAPR, p. 77-87.

[13] Krebel, U. (1999). Pairwise Classification and Support Vector Machines. Advances in Kernel Methods: Support Vector Learning. p. 255-268, MA: MIT Press.

[14] Friedhelm, Schwenker. (2000). Hierarchical Support Vector Machines for Multi-class Pattern Recognition. The 4th International Conference on Knowledge-based Intelligent Engineering System & Allied Technologies, p.561-565, Brighton, USA, August

[15] Chin, S. L. (1994). Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent and Fuzzy Systems*, 2 (3) 267-278.

[16] Ronald, R. Y., Dimitar, P. F. (1994). Essentials of Fuzzy Modelling and Control. John Wiley & Sons Inc., New York, USA.