15th IMEKO TC10 Workshop on Technical Diagnostics
Technical Diagnostics in Cyber-Physical Era
Budapest, Hungary, June 6-7, 2017

Optimization of the neural RBF classifier for the diagnostics of electronic circuit

Bartosz Połok, Piotr Bilski

[1] *Institute of Radioelectronics and Multimedia Technology, Warsaw University of Technology, ul. Nowowiejska 15/19 00-665 Warsaw, bpolok@ire.pw.edu.pl, pbilski@ire.pw.edu.pl*

***Abstract* – The paper presents the detailed analysis of the Radial Basis Function (RBF) neural network application for diagnostics of analog systems. In most cases RBF networks are used in the approximation tasks. In this work the network is used as the fault classifier. Because RBF networks are known to be dependent on the size of training data, the procedure to minimize the amount of the learning examples was proposed. The classifier is tested on the model of the electronic filter, which, was also diagnosed by alternative methods, such as Multilayered Perceptron and Support Vector Machine. Experimental results show what are advantages and drawbacks of the RBF classifier compared to other solutions.**

I. INTRODUCTION

Artificial Neural Networks (ANN) the most popular Artificial Intelligence (AI) tools used in the diagnostics of analog systems. Their numerous advantages include the ability to extract generalized knowledge from the available data, autonomous operation and (in some cases) the ability to accurately process data in uncertainty conditions. Memory and processing efficiency make them useful in embedded applications. Their disadvantage is the obscure form of stored knowledge, which is not crucial as long as the diagnostic system accurately detects and identifies faults.

The most widely used ANN are Multilayered Perceptrons (MLP), which were successfully applied to solve biomedical and financial problems [1]. Their implementations in the diagnostics cover power lines [2] or electrical machines [3]. They are currently widely exploited and work as the standard diagnostic tool. Using the MLP classifier requires selecting one of multiple available learning algorithms to the predefined architecture of the network, including selected types of neurons (computational units) and number of layers. These are often main design problems, as the optimal MLP architecture must be determined individually for each problem.

Similar problems are encountered during the application of Support Vector Machines (SVM), which are considered the optimal ANN-based classifiers in the uncertainty conditions [4]. Their disadvantage is the time consuming process of selecting the optimal kernel and its parameters [5]. Despite such challenges, SVM are also popular in diagnostics, used to identify faults in electrical machinery [6], electronic circuits [7] or power plants [8].

On the other hand, RBS networks are considered simpler in design and training. They are faster trained and have only one hidden layer, therefore the only parameter to determine is the number of computational units, which depends on the size of the training data. In most cases the RBF network is used for the approximation task, having linear neurons in the output layer. Its application to classification requires substituting the linear output neurons with their sigmoidal counterparts. Such approaches are rare and not well described in the literature [9]. This calls for the thorough investigation of the ability to apply the RBF classification module in the diagnostics of analog systems.

The aim of the paper is to present analysis of various RBF network configurations for the fault detection and identification. Monitored parameters include relations between the size of the network or the output neurons coding schemes and the diagnostic accuracy. The RBF-based classifier is compared to its well-established counterparts, i.e. MLP and SVM to see in which situations it should be selected for the task. All classifiers are tested on the model of the 5^{th} order filter.

The structure of the paper is as follows. Section II presents the applied diagnostic architecture. Details of the implemented network are also introduced here. Training and testing data sets' structure is presented in Section III. The analyzed circuit is introduced in Section IV, while experimental results are in Section V. Conclusions about the conducted analysis are in Section VI.

II. ANN-BASED DIAGNOSTIC ARCHITECTURE

The typical diagnostic scheme using any type of the ANN-based classifier is presented in Fig. 1. Here the System Under Test (SUT) is monitored by the hardware/software module, which performs data acquisition on the accessible nodes, extracting from measured signals the feature vector $\boldsymbol{e}$ (further called example), subsequently processed to make the diagnostic decision (hypothesis) h. It is assumed that knowledge stored in the module allows for the automatic fault detection and identification. This requires implementation of the machine learning algorithm. During the operation

the ANN-based classifier produces the binary output, in which every fault category is represented by the unique sequence of values *{0,1}* or *{-1,1}*, depending on the used type of sigmoidal neurons.

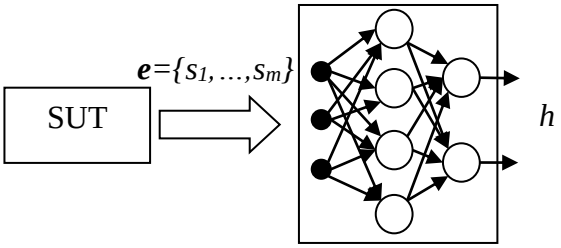

Fig. 1. ANN-based diagnostic system architecture.

Application of the RBF network for the fault classification task requires adjusting its structure to the specific problem. This includes selection of the number of neurons in the hidden layer and selecting coding scheme for binary units in the output layer. As opposed to MLP and similarly to SVM, the RBF network contains hidden layer units with the Gaussian activation function $\varphi(s_i)$ of the width (defined as the standard deviation σ, where c is the center of the function (1), usually set to *0* for all neurons) being the design parameter, which influences the capability to distinguish between different fault categories. These aspects of the RBF classifier must be optimized during the design stage, as described below.

$$\varphi(s_i) = \exp\left(-\frac{\|s_i - c\|^2}{2\cdot\sigma^2}\right) \quad (1)$$

A. Minimization of the number of hidden neurons

The number of neurons k in the hidden layer (usually greater than for MLP) is automatically adjusted during the training. The designer's task is to set the maximum acceptable value. Maximally k is equal to the number of examples in the set, which ensures the high classification accuracy, but makes the network large and complex. To avoid this, the number of hidden neurons should be minimized. In this work clustering of examples was applied to find the most similar ones. It is assumed that training data is redundant, caused by one of two factors:

- low sensitivity of the SUT on the changes of the specific parameter, which results in multiple examples describing the same category being close to each other. Such a group can easily be substituted by the single example representing all original members. It becomes their centroid.
- existence of ambiguity groups (AG) [10], resulting in multiple examples belonging to different categories being close to each other. In this case examples can't be replaced and must all remain in the data set.

Data processing requires determining all groups of similar examples by the cluster analysis, which results in the set of groups G_l easily distinguishable from each other, based on the similarity measurements. The detailed scheme is presented in Fig. 2.

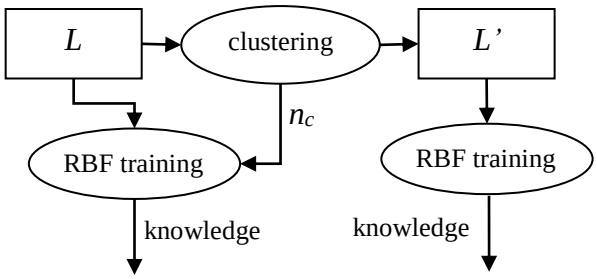

Fig. 2. Learning data set preprocessing for the RBF classifier training.

Redundancy in the set L may be exploited in two ways. The first one consists in providing the original data set for the RBF network training, but limiting the maximum number of neurons in the hidden layer to the number of clusters n_c. Alternatively, the reduced data set *L'* may be provided for training.

The distance-based similarity method was used in the presented case. Its parameter is the threshold θ, below which two examples $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are considered close to each other (2). The measure applied to create clusters exploits Euclidean d_E (3) and cosine d_c (4) distances, treating every example as the point in the m-dimensional space. This way groups located close to each other are easily identified. Thresholds θ_1 and θ_2 for both distances should be selected adaptively to minimize the number of clusters containing examples belonging to different classes (i.e. forming AG).

$$e_i, e_j \in G_l \Leftrightarrow d_E(e_i, e_j) < \theta_1 \wedge d_E(e_i, e_j) < \theta_2 \quad (2)$$

$$d_E(e_i, e_j) = \sqrt{\sum_{k=1}^{m}(s_{ik} - s_{jk})^2} \quad (3)$$

$$d_c = \frac{\sum_{k=1}^{m} s_{ik}\cdot s_{jk}}{\sqrt{\sum_{k=1}^{m} s_{ik}^2}\cdot\sqrt{\sum_{k=1}^{m} s_{jk}^2}} \quad (4)$$

The illustration of the clustering method for *m=2* is presented in Fig. 3.

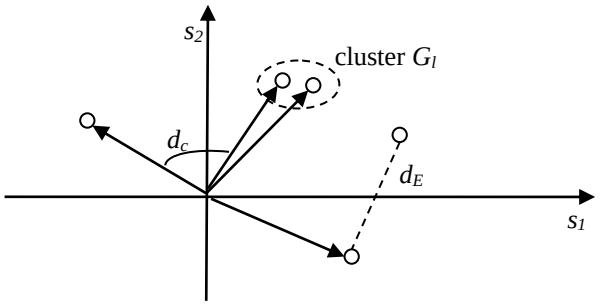

Fig. 3. Illustration of the clustering method.

B. Output layer coding schemes

The important parameter of the ANN-based classifier is representation of multiple categories by neurons in the output layer. Among multiple coding schemes, three were selected for experiments [11]. It is assumed that only single faults are considered (the most probable case). Schemes differ in the number of output neurons *o*:

- One-vs-All (OvA) – here each output unit is responsible for the separate category. The number of fault codes *l* determines the number of neurons *o* in the output layer (Table 1). During the identification only one neuron is active, determining the particular fault. If multiple neurons become active, they can be interpreted as fault candidates in the uncertainty conditions.

Table 1. Applied OvA coding scheme.

Category	Coding scheme
c_1	*1,0,0, ...,0*
c_2	*0,1,0, ...,0*
...	...
c_l	*0, ...,0,0,1*

- One-vs-One (OvO) – each neuron is responsible for distinguishing between the pair of categories. This time *o* classifiers with output neuron are created and simultaneously trained Their number *o* is driven by the number of distinct pairs: $l \cdot (l-1)/2$. In this scheme both active and inactive neuron determines the fault. The diagnostic decision is made based on the majority voting – the category pointed by the greatest number of network supporting it. This way classifiers are created. For instance, three fault codes require the following units: *1:* c_1 *vs* c_2, *2:* c_1 *vs* c_3, *3:* c_2 *vs* c_3.
- Minimum Output Coding (MOC) – the minimal set of neurons, which represents each fault code by the unique combination of active units. The subsequent codes are represented by the binary representation of integers (Table 2).

The more sophisticated coding schemes, such as Error Correcting Output Coding (ECOC) [12] were excluded from experiments, as their implementation would complicate the output layer even more, not necessarily increasing diagnostic accuracy. Computations were performed using the Matlab environment.

Table 2. Applied MOC coding scheme.

Category	Coding scheme
c_1	*0,0,0, ... ,0,1*
c_2	*0,0,0, ... 1,0*
c_3	*0,0,0, ... 1,1*
...	...

III. DATA SET DESCRIPTION

Learning *L* and testing *T* datasets are required to train the RBF network classifier and verify its accuracy. Both have the same structure, containing *n* feature vectors ***e*** with *m* attributes (symptoms *s*), supplemented by the fault code, describing the state of the SUT. The generic form of the set (4) is the table, identical for *L* and *T*.

$$L \equiv T = \begin{bmatrix} s_{11} & \cdots & s_{1m} & c_1 \\ \vdots & \ddots & \vdots & \vdots \\ s_{n1} & \cdots & s_{nm} & c_l \end{bmatrix} \tag{5}$$

The example is generated after simulating the SUT model. It is then possible to introduce the desired fault to the structure of the examined object and then observe the effect of the fault on the accessible nodes. The key issue is generation of the fault code for each example, based on the actual value of the faulty parameter. As before [13], it is the integer number, being the combination of the parameter identifier and its deviation from the nominal value. The code considers the discrete degree of change, both negative and positive (for values greater and smaller than the nominal one). The assignment of these degrees depends on the assumed thresholds for deviations. For instance, if the parameter's value is above ten percent of nominal, the degree is labeled with "1", indicating it being "too large" or "too small" (depending on the direction of deviation). Additional intervals depend on the need to increase the resolution of the diagnostic module. For instance, if the value of the parameter is above fifty percent of the nominal value, the degree "2" may be assigned. This way the second parameter with the value larger than the nominal one is represented by the code "21", while the fist parameter with the value much smaller than the nominal one obtains the code "-12". The exception is the nominal state, encoded with the value "0". The number of simulated parameter values depends on the expected data set size and the accuracy of the SUT modeling. In the presented work both sets are of the same size, containing mutually exclusive examples.

Diagnostic accuracy *acc* of the ANN-based classifier is measured using the set *T*, for which the number of incorrectly classified examples (i.e. the ones for which the hypothesis $h(\boldsymbol{e})$ is not equal to the fault code $c(\boldsymbol{e})$)

determines the sample error e_s:

$$acc = 1 - e_s = 1 - \frac{|\boldsymbol{e} \in T : c(\boldsymbol{e}) \neq h(\boldsymbol{e})|}{|T|} \quad (6)$$

Each SUT simulation was performed after changing the single parameter's value beyond the tolerance margins (here 10% of the nominal value) with all other parameters nominal.

IV. ANALYZED SUT

The fifth-order analog filter is a good example to implement the data processing methods. It is complex (because of the large number of nodes and SUT elements) and difficult to diagnose. Therefore multiple stamps are needed to allow for identification of all elements (resistances and capacitances). The circuit in Fig. 1 contains ten analysed elements of the following nominal values: $R_1=R_2=R_3=R_4=R_5=1k\Omega$, $C_1=16nF$, $C_2=19nF$, $C_3=13nF$, $C_4=51nF$ and $C_5=49nF$. Subsequent resistances were labeled with numbers from 1 to 5 respectively, while capacitances were referred to as parameters No. 6 to 10. The cutoff frequency of the filter for such values is 10kHz. The model of the circuit was implemented in the Simulink environment. Simulations were performed to obtain examples of the SUT behaviour for different values of elements (up to 90 percent of the nominal value). The excitation signal provided at node No. 1 was a sinusoid with 9kHz frequency (i.e. close to the cutoff frequency). The filter was analysed in the time domain, where at the accessible nodes sinusoidal responses were recorded. Measurements were taken at nodes 2, 3, 5, 6, 8 and 9. From each response the first three maximal and minimal values of the signal with their time instants and time instants of zero crossings. This gives the total number of 54 features for each example.

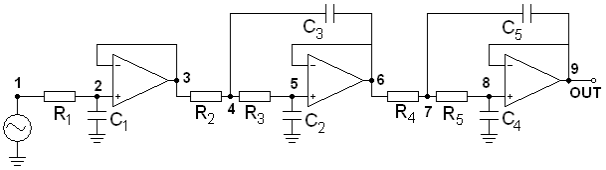

Fig. 4. Scheme of the 5th order lowpass filter.

To evaluate the dependency between the size of the set and the accuracy of RBF-based classifier, various sizes of the sets were prepared, strating from 70 examples (7 simulations for each parameter, including the nominal state) to 180 (18 simulations for each example). To consider tolerances, additional sets were created with results of simulations affected by the random value added to the actual parameter. This allows for verifying if random deviations of parameter values influence the ability of the diagnostic module to distinguish faults. The number of different fault codes was 41, as not all 5 were used for each parameter.

V. EXPERIMENTAL RESULTS

Conducted experiments consisted of three stages. First, coding schemes were compared. Next, relation between the size of the data set and the diagnostic accuracy was determined. Finally, comparison between RBF, SVM and MLP (also implemented in Matlab) was performed to find advantages of the proposed approach over other well established approaches. In all cases experiments were repeated ten times to obtain the average values, as each time knowledge represented by the network is different, so diagnostic results may vary.

A. Coding schemes verification

Presented coding schemes were used to design the RBF network, train it on the set L and test using the set T. Optimal results for all configurations (regarding the width of the RBF σ, number of output neurons o, number of hidden neurons k, the maximum achieved training error e and obtained accuracy on the set T) for the full data set (180 examples) are in Table 3.

Table 3. Comparison of performance of RBF network coding schemes.

Coding scheme	σ	k	o	e	*acc [%]*
OvA	0.5	82	41	0.0	81.11
OvO	0.6	96	820	0.0	80.56
MOC	0.3	95	6	0.02	80.56

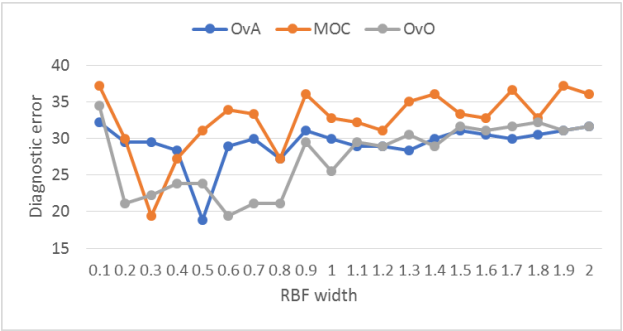

Fig. 5. Influence of the RBF width on the diagnostic accuracy.

In general, the OvA coding is the best scheme, producing the minimum error for most network configurations. Although requiring greater number of neurons than MOC, it ensures the greatest accuracy (the smallest error). The least practical is OvO because of the largest number of neurons representing categories. The full sweep optimization procedure was applied to determine the value of σ (although more sophisticated approaches, such as simulated annealing [5] or genetic algorithm can be used) The optimal RBF width and the number of hidden neurons are in all cases in the middle of

the verified range (between *0.1* and *2.0*). In Fig. 5 the relation between σ and the obtained accuracy (with all other parameters set as in Table 3) is presented. For all coding schemes, there is the best σ value, for which the minimum error is obtained. Results show that narrow Gaussian functions are preferred (like σ=*0.5* for OvA).

Similar effect is obtained for the number of hidden neurons. They strongly affect the accuracy, as is presented in Fig. 6. Disregarding the coding scheme, in all cases adding more hidden neurons leads to the decrease of the diagnostic error to some extent. Above the optimal point, further increase does not improve the network's efficiency (or even makes it worse).

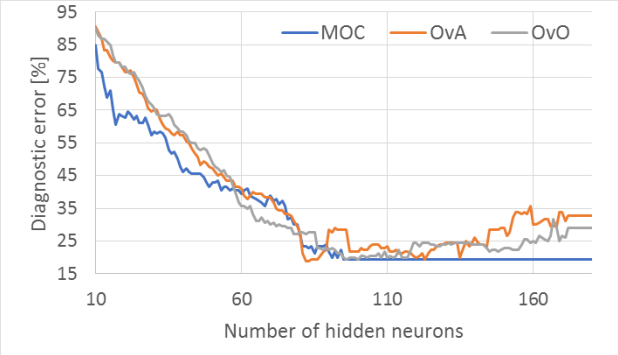

Fig. 6. Influence of the number of hidden neurons on the diagnostic accuracy.

B. Data set size

Experiments regarding the size of the training set *L* were divided into two steps. In the first one, various numbers of experiments were inserted to the set and then used to train the RBF classifier. During the second stage, the original set was processed to group the most similar ones (see Section II.A) and this way compress the size of examples. Additionally, dependence between the training duration and the size of processed data was established.

Table 4 compares the diagnostic accuracy for data sets of different sizes. As in other applications, the amount of training data influences the accuracy of the classifier, which may be a problem if the they are difficult to collect. On the other hand, the original set may be compressed to the smaller version based on the clustering scheme. This way it is possible to determine redundant examples, which should be represented by the single vector of features. Results of the diagnostic procedure for the RBF trained on the minimized version of the set are presented in Table 5. Increasing the number of examples improves fault detection and identification performance, but to obtain the accuracy equal to the RBF trained on the largest set with 180 examples, the density of clustering must not be large. Comparable diagnostic results are obtained on data sets with cardinality close to the original set. Again, the most effective is the OvA scheme.

Table 4. Influence of the data set size on the RBF classifier performance.

Data set size	acc_{OvA} [%]	acc_{MOC} [%]	acc_{OvO} [%]
70	31.42	32.85	32.48
120	24.17	27.12	28.04
180	81.12	80.56	80.56

Table 5. Influence of the data clustering on the RBF classifier performance.

Data set size	acc_{OvA} [%]	acc_{MOC} [%]	acc_{OvO} [%]
70	63.89	61.12	61.12
88	66.67	65.56	67.23
103	72.23	68.34	71.12
109	70.00	70.56	71.12
136	70.56	69.45	72.78
153	76.67	72.23	76.12
162	81.12	72.78	78.89
180	81.12	80.56	80.56

Relation between the RBF network structure and training duration for the OvA coding is in Fig. 7. The processing time directly depends on the number of hidden neurons in the network structure. The relation between the size of the data set and the number of neurons is more complex, as even larger data sets may be optimally processed by smaller networks. The relation is in general linear. For contemporary computers durations required to perform single training are negligible. Thorough optimization process for σ and *k* requires more time in the off-line mode.

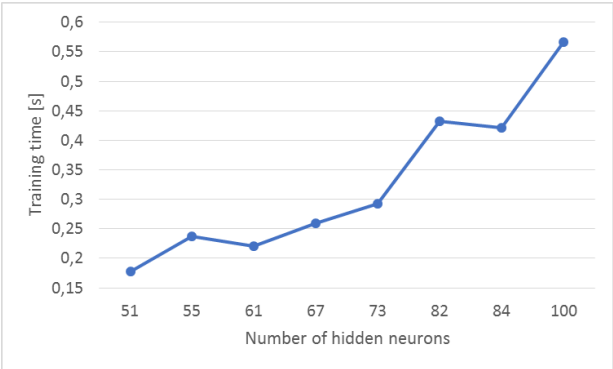

Fig. 7. Training duration of the OvA RBF

C. Comparison against other classifiers

Besides the RBF-based network classifier, also SVM and MLP were tested on the same data sets. In Tab. 6 diagnostic outcomes of the optimal network configurations for the largest set *T* are presented. Performance of all classifiers is comparable, the main difference lies in the training time, which is the longest for the MLP and comparable for RBF and SVM. In the last two cases, additional parameters must be determined,

which increases the training duration. This shows advantages of the presented classifier compared to its counterparts: comparable accuracy with the faster training process and relatively simple structure.

Table 6. Comparison of the ANN-based diagnostic modules.

Data set size	acc_{OvA} [%]	acc_{MOC} [%]	acc_{OvO} [%]
RBF	81.12	80.56	80.56
MLP	80.75	80.56	79.80
SVM	81.12	81.12	79.80

VI. CONCLUSIONS

The paper presented analysis and optimization of various RBF network-based classifier in the diagnostics of analog circuits. Multiple parameters, including width of the standard deviation, number of neurons in the hidden layer and coding scheme in the output layer were checked. The OvA scheme was the best in most of cases. The MOC scheme, although requiring the minimum number of output neurons is more susceptible to random classification errors. The OvO, provides similar accuracy, but requires the greatest number of networks with the single output neuron. Such structures are impractical in the on-line diagnostics.

The optimization process of the RBF training is similar to all other ANN-based classifiers. The supervised learning is implemented as the standard parameterized algorithm. The designer's task is to select parameters to ensure the highest diagnostic accuracy. This process is time consuming, but results in the improvement in the accuracy. Multiple continuous or discrete optimization methods may be used for this purpose, as should be verified in the future research.

REFERENCES

[1] T.M. Kwon and E.H. Feroz, "A multilayered perceptron approach to prediction of the SEC's investigation targets," IEEE Trans. Neural Networks, Vol. 7, No. 5, Sep 1996, pp. 1286-1290.

[2] M.N. Mahmud, M.N. Ibrahim, M.K. Osman, and Z. Hussain, "Selection of suitable features for fault classification in transmission line," IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 27-29 Nov. 2015, DOI: 10.1109/ICCSCE.2015.7482253.

[3] J. Penman, C.M. Yin, "Feasibility of using unsupervised learning, artificial neural networks for the condition monitoring of electrical machines," : IEE Proceedings - Electric Power Applications, Vol. 141, No. 6, Nov 1994, pp. 317-322.

[4] H. Ma, T.K. Saha, and C. Ekanayake, "Power transformer insulation diagnosis under measurement originated uncertainties," IEEE Power and Energy Society General Meeting, 25-29 July 2010, DOI: 10.1109/PES.2010.5589395.

[5] P. Bilski, "Automated selection of kernel parameters in diagnostics of analog systems," Przegląd Elektrotechniczny, No. 5, 2011, pp. 9-13.

[6] R. Fang and H. Ma, "Application of MCSA and SVM to Induction Machine Rotor Fault Diagnosis," Sixth World Congress on Intelligent Control and Automation, 21-23 June 2006, DOI: 10.1109/WCICA.2006.1714134.

[7] F. Ye, Z. Zhang, K. Chakrabarty, and X. Gu, "Board-Level Functional Fault Diagnosis Using Multikernel Support Vector Machines and Incremental Learning," IEEE Tran Computer-Aided Design of Int. Circuits and Systems, Vol. 33, No. 2, Feb. 2014, pp. 279-290.

[8] M. Berahman, A.A. Safavi and M.R. Shahrbabaki, "Fault detection in Kerman combined cycle power plant boilers by means of support vector machine classifier algorithms and PCA," 3rd International Conference on Control, Instrumentation, and Automation (ICCIA), 28-30 Dec. 2013, DOI: 10.1109/ICCIAutom.2013.6912851.

[9] I. Samy, I.-S. Fan and S. Perinpanayagam, "Fault diagnosis of rolling element bearings using an EMRAN RBF neural network- demonstrated using real experimental data," Sixth International Conference on Natural Computation (ICNC), 10-12 Aug. 2010, DOI: 10.1109/ICNC.2010.5583833.

[10] G.N. Stenbakken, T.M. Souders, and G.W. Stewart, 1989, "Ambiguity groups and testability," IEEE Trans. Instr. and Meas., Vol. 38, Issue 5, 1989, pp. 941-947.

[11] N. Garcia-Pedrajas and C. Fyfe, "Evolving Output Codes for Multiclass Problems," IEEE Trans. Evol. Comp., Vol. 12, No. 1, Feb. 2008, pp. 93-106.

[12] S. Escalera, O. Pujol and P. Radeva, "On the Decoding Process in Ternary Error-Correcting Output Codes," IEEE Trans. Pattern Anal. and Machine Intel., Vol. 32, No. 1, Jan. 2010, pp. 120-134.

[13] P. Bilski, J. Wojciechowski, "Automated Diagnostics of Analog Systems Using Fuzzy Logic Approach", IEEE Trans. Instr. and Meas., Dec. 2007, Vol. 56, Issue 6, pp. 2175-2185.