TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.5, May 2014, pp. 3616 ~ 3624
DOI: http://dx.doi.org/10.11591/telkomnika.v12i5.4054

A Novel Transformer Fault Diagnosis Approach Based on Information Fusion Method

Huang Xin-bo*, Song Tong, Lin Xiao-huan, WangYa-na
College of Electronics and Information, Xi'an Polytechnic University,
Xi'an 710048, Shaanxi Province, China
*Corresponding author, e-mail: huangxb1975@163.com

Abstract

To improve the power transformer fault diagnosis accuracy, this paper proposes a fault diagnosis method of information fusion which is based on fuzzy coding boundary and Bias regularization Levenberg-Marquardt (L-M) network. The algorithm uses a Bias approach to determine the hyper parameters, making the neural network adaptively adjust the parameter in the training process and then gets the optimization parameters of the objective function. On the other hand, the using of fuzzy coding boundary can reduce the variations and improve the accuracy of fault diagnosis. The contrast of the two fusion diagnosis results draws a conclusion. That is, the performance of Bias Regularization Fuzzy L-M Network is superior to the no feature reduction fusion model which is Bias Regularization L-M Neural Network, and the accuracy rate of the former is 89.83%.

Keywords: *power transformer, information fusion, bias regularization; L-M neural network; fuzzy coding boundary*

1. Introduction

Multi-Source Information Fusion (MSIF) technology is a theory and a method which researches the comprehensive treatment and application of uncertainty information, namely, getting more accurate and credible conclusion by processing the information from multiple information sources in multi-level recognition [1]. In the fault diagnosis aspect, MSIF technology uses the extracted feature information of system failure and deduce the fault type of the object according to the fault diagnosis methods, then the fusion center processes the complementary and redundant information based on certain criteria in the space and time, and ultimately get the fault decision of object types. Therefore, the information fusion method which is applied to fault diagnosis can greatly improve the completeness of fault character information [2]. At present, the diversity, uncertainty and complexity bring more difficulties in fault diagnosis technology of the transformer. The literature [3] is about the application of neural network in fault diagnosis of the transformer, but the rate of convergence is slow. The literature on [4] uses genetic algorithms (GA) to improve the neural network's weights and threshold, but the GA method is complex, so the network is easy to fall into local optimal. The literature [5] adopts fuzzy membership function which could overcome the absolute situation of ratio of coding boundary, but the parallel processing ability is poor. The literature [6] employ particle swarm optimization algorithm (PSO) for transformer fault diagnosis, improving PSO algorithm by linear decreasing strategy. However, the network is easy to fall into local optimum. Particles are easy to reach premature convergence too.

Relying on a single diagnosis method of transformer fault characteristics can reflect the transformer condition from only one aspect, and it can not make a comprehensive evaluation on the overall health status of transformer. Therefore, based on the idea of intelligent complementary fusion, this paper established the information fusion fault diagnosis model of fuzzy coding boundary and Bias Regularization L-M Neural Network. Simplify the input unit number of the L-M neural network by using the fuzzy coding boundary, and improve the generalization ability of L-M neural network by using Bias regularization algorithm. In this way, the two advantages are complementary, not only identifying the fault types, but also improving the rate of correct diagnosis. Finally, compared with several forecasting methods, the experimental results show that the model has strong ability of simulation and forecast.

Received July 30, 2013; Revised December 2, 2013; Accepted January 8, 2014

2. The Principle of the Proposed Algorithm

2.1. Regularization Theory of Bias

The basic idea of Bias regularization algorithm is as follows:

Given a set of training samples $K = \{(p_1, t_1), (p_2, t_2), ..., (p_n, t_n)\}$, the neural network learning objective is that look for a function to approach the samples, what's more, minimize the error function. Mean square error function is usually used in neural network training given by equation [1]:

$$E = \frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2 \tag{1}$$

In the formula, n as the total sample; t_i as the expected output values of network; y_i as the actual output value. However, in order to improve the generalization ability of the network, the Bias regularization method increases the arithmetic average value of the weights' square in the objective function. The objective function is given by equation [2]:

$$F = \phi E + \varphi E_W \qquad E_W = \frac{1}{m}\sum_{i=1}^{m}\omega_i^2 \tag{2}$$

In the formula, ω_i as connection weight of the neural network, m as the number of connection weights in the neural network, ϕ, φ as the parameters of the objective function. If $\varphi \le \phi$, then the training algorithm aims to minimize the network training error; If $\varphi \ge \phi$, then the training algorithm aims to enable the network to produce a smoother response; It means as far as possible to reduce the network parameters effectively, then make up for the network error. The conventional method of regularization is usually difficult to determine the size of regularization parameters, while the theory of Bias can adaptively adjust the size of regularization parameters and make them optimal in the network training process [7].

2.2. Fuzzy Three Ratio Coding Criteria

Fuzzy three ratio method is based on fuzzy theory to fault diagnosis of the power transformer, which makes the interval on three ratio boundary. According to the traditional coding rules of three ratio method , the boundary of characteristic gas ratio is "0.1", "1", "3" about $C_2H_2/C_2H_4, CH_4/H_2, C_2H_4/C_2H_6$ .On the basis of the empirical knowledge, the boundary of the "0.1" is fuzzy that "0.08 ~ 0.12", "1" border is fuzzy that "0.85 ~ 1.15" and "0.9 ~ 1.1", "3" is fuzzy that "2.9 ~ 3.1" and "2.85 ~ 3.15" [8]. The membership function of each gas ratio is fuzzy distributed by the method of assigned. The membership function of 0, 1 and 2 is respective partial small Γ, middle type ridge type and partial large Γ. The formulas are given as follows. So, when the code is 0, 1, 2, the corresponding membership function is $u_0(xb_i), u_1(xb_i), u_2(xb_i)$. Then, relying on the principle of maximum degree of membership to determine the final code. Therefore, the characteristics of gas sample data used in this paper change into the coding sequence of "0, 1, 2" as the input of the network.

$$u_0(xb_i) = \begin{cases} 1, xb_i \le 0.08 \\ e^{-50(xb_i - 0.08)}, xb_i \ge 0.08 \end{cases} \tag{4}$$

$$u_1(xb_i) = \begin{cases} 0, xb_i \le 0.08 \\ 0.5 + 0.5\sin[25\pi(xb_i - 0.1)], 0.08 < xb_i \le 0.12 \\ 1, 0.12 < xb_i \le 2.9 \\ 0.5 - 0.5\sin[5\pi(xb_i - 3)], 2.9 < xb_i \le 3.1 \\ 0, xb_i > 3.1 \end{cases} \tag{5}$$

$$u_2(xb_i)=\begin{cases}0, xb_i \le 2.85\\ 1-e^{-12(xb_i-2.85)}, xb_i > 2.85\end{cases} \quad (6)$$

2.3. The L-M Network Training Principle

The essence of neural network modeling is to find out the essential connection between the input and output data in the finite sample, namely the mapping relationship, thus the input without training can also give the appropriate output and has the generalization function. Refer to the documents [9, 10], the standard BP algorithms use the steepest descent method to modify weights, and the training process from a point along with the surface of error function, then gradually reach the minimum point to make the error zero. When the network is complex, the training process may be trapped in a local minimum, and the convergence speed is slow. In order to overcome these shortcomings in the algorithm, using L-M algorithm, also known as the damped least square method is used. It is better than the traditional BP and other improved algorithm in the number of iterations, having higher convergence speed and accuracy.

The weights adjustments are given by equation [7].

$$\Delta\omega=(J^T J+\mu I)^{-1} J^T e \quad (7)$$

In the formula, e as error vector; J as the error of the weight differential Jacobi matrix; μ as a scalar, when μ increases, it is close to the steepest descent method of smaller learning rate; When μ dropped to 0, the algorithm becomes the Gauss - Newton method. Therefore, the L-M algorithm is a smooth transition between the steepest descent method and Gauss Newton method [11].

The specifically iterative steps of L-M algorithm as follows:

Step one: give all inputs to the network and compute the output of the network, then adopt error function to calculate the training target's sum of square error;

Step two: calculate the error of the weights' differential Jacobi matrix J ;

(1) The definition of Marquardt's sensitivity: $S_i^m = \frac{\partial E}{\partial n_i^m}$, n as weighted sum of each layer of the network.

(2) The sensitivity of the recursion relations is $S_q^m = E(n_q^m)(\omega^{m+1})S_q^{m+1}$, the sensitivity can through the last layer of the network back to the first layer, and then calculate the Jacobi matrix.

Step three: using the formula [7] find out $\Delta\omega$;

Step four: calculate the sum of square error repeatedly. If the new sum smaller than the calculation in step one, the use of $\theta(\theta > 1)$ divided by μ and there are $\omega = \omega + \Delta\omega$ .Then go to step one; Otherwise, μ multiplied by θ, then go to step three. When the sum of square error decreases to a target error, the algorithm is considered convergence [12].

3. Multi-feature Fusion Fault Diagnosis

3.1. Fusion Principle

Due to the characteristics of large quantity, high dimension, correlation and repeat with each other, when using the Bias regularization L-M neural network to recognize the fault type of the extracted feature, there will be large network computation, slow training speed, and the classification effect is not good because of the existence of redundant information's interference. While, using the fuzzy boundary coding method not only to reduce feature, remove redundant information, but also to maintain the same classification ability by essential characteristics. Input the Bias regularization L-M network train again, reduce the amount of calculation, retain the key attributes, improve the rate of correct diagnosis ultimately. The steps of information fusion of gas feature are shown as Figure 1.

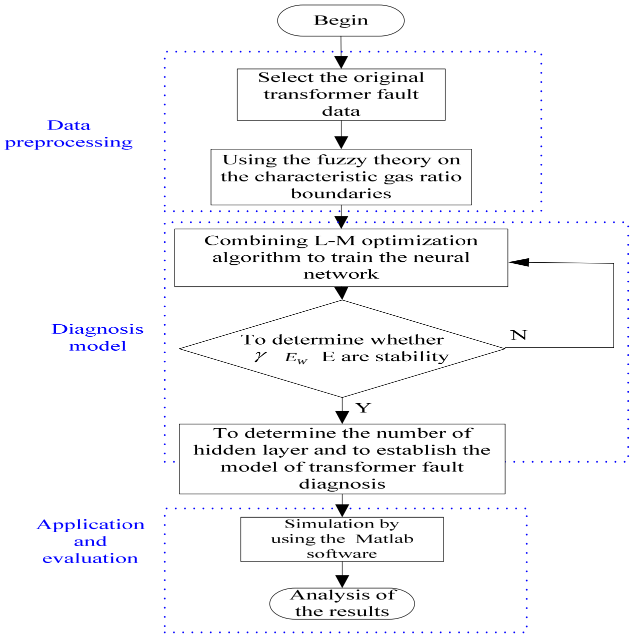


Figure 1. The Process Diagram of Multiple Information Fusion

3.2. The Establishment of Bias Regularization Parameters

A key of the transformer fault diagnosis model is to establish the parameters ϕ and φ. That is how to establish the size of the parameter φ,ϕ to make γ, E_W, E stable. What's more, ensuring the network to achieve the optimal. By the Bayesian formula calculating φ,ϕ given as follows.

The posterior distribution of φ and ϕ based on Bias theorem are given by equation [8]:

$$P(\varphi,\phi \mid D,M) = \frac{P(D \mid \varphi,\phi,S)P(\varphi,\phi \mid S)}{P(D \mid S)} \tag{8}$$

If the prior distribution $P(\varphi,\phi \mid S)$ is a wide distribution, and φ,ϕ two variables of the posterior probability is independent with the normalization factor $P(D \mid S)$.Therefore, it is only necessary to make the likelihood function $P(D \mid \varphi,\phi,S)$ maximum, which will make the posterior distribution of φ,ϕ maximum.

Bias methods focus on the probability distribution of weights in the space. S on behalf of the network structure. Firstly, the network structure has been determined and no sample data. If a priori distribution of weights $P(\omega \mid \varphi,S)$ is given, the posterior distribution of weight is $P(\omega \mid D,\varphi,\phi,S)$ when a sample data D has been set. According to the Bias theorem, the formula is given by equation [9]:

$$P(\omega \mid D,\varphi,\phi,S) = \frac{P(D \mid \omega,\phi,S)P(\omega \mid \varphi,S)}{P(D \mid \varphi,\phi,S)} \tag{9}$$

From the formula [9] shows, in order to obtain the posterior distribution $P(\omega \mid D,\varphi,\phi,S)$, the prior distribution $P(\omega \mid \varphi,S)$ and the likelihood function $P(D \mid \omega,\phi,S)$ should be known firstly. The following is the concrete solving process of the two functions.

$$P(\omega \mid \varphi,\phi) = \frac{1}{Z_W(\varphi)}\exp(\varphi E_W) \qquad Z_\omega(\varphi) = \left(\frac{2\pi}{\varphi}\right)^{\frac{\omega}{2}} \tag{10}$$

$$P(D\,|\,\omega,\phi,S)=\frac{1}{Z_D(\phi)}\exp(-\phi E) \qquad Z_D(\varphi)=(\frac{2\pi}{\phi})^{\frac{N}{2}} \tag{11}$$

Note that $P(D\,|\,\varphi,\phi,S)$ has nothing to do with the weight vector ω, thus substituting the prior distribution $P(\omega\,|\,\varphi,S)$ and the likelihood function $P(D\,|\,\omega,\phi,S)$ can derive the weights of the posterior distribution which is given by Equation (12).

$$\begin{cases} P(\omega\,|\,D,\varphi,\phi,S)=\dfrac{1}{Z_F(\varphi,\phi)}\exp[-F(\omega)] \\ Z_F(\varphi,\phi)=\int_{-\infty}^{+\infty}\exp(-\phi E-\varphi E_W)d\omega \end{cases} \tag{12}$$

Due to the independent between $Z_F(\varphi,\phi)$ and ω, the maximum of posterior distribution can be obtained by minimizing $F(\omega)$ .And the corresponding weight is required at this time. By the formula (9) and (12) get the formula [13]:

$$P(D\,|\,\varphi,\phi,S)=\frac{Z_F(\varphi,\phi)}{Z_W(\varphi)Z_D(\phi)} \tag{13}$$

In order to determine the $Z_F(\varphi,\phi)$, making the $F(\omega)$ expand at the minimum point ω^*. Because the gradient is 0, the approximation of $F(\omega)$ is given by Equation (14).

$$F(\omega)=F(\omega^*)+\frac{1}{2}(\omega-\omega^*)H(\omega-\omega^*) \tag{14}$$

The Hessian matrix is symmetric positive semi definite, so $H=(H^{\frac{1}{2}})^T H^{\frac{1}{2}}$. Let $u=H^{\frac{1}{2}}(\omega-\omega^*)$, making the formula (14) subsume into (10), then the integral of the both sides of the equation to get formula (15).

$$Z_F(\varphi,\phi)=(2\pi)^{\frac{N}{2}}e^{-F(\omega^*)}[\det(H(\omega^*)^{-1})]^{\frac{1}{2}} \tag{15}$$

Making the formula (15) subsume into (13), then the logarithm is used for the new equation. And the use of the first-order condition of optimal worth can obtain the optimal regularization parameter:

$$\varphi^*=\frac{\gamma}{2E_W(\omega^*)} \qquad \phi^*=\frac{N-\gamma}{2E(\omega^*)} \tag{16}$$

In the formula, $\gamma=N-2\phi\times tr(H)^{-1}$, N is the total number of the network weights; $\gamma\in(0,N)$, γ as the real effective parameters in general parameters N ,which reflects the actual size of the network.

Through the establishment of parameters φ and ϕ, making the L-M algorithm train the neural network by using the error objective function with weights. In this way, ensure the sum of square error minimum about the network, and effectively control the network complexity, which will help to improve the generalization ability [13].

4. Results and Analysis

4.1. Data Sources

In the paper, the simulation test data come from the main transformer oil chromatography monitoring Intelligent Electric Device (IED) normal operation data in ±660kV convertor station in the east Yin Chuan, Ning Xia province and ±220kV substation in An Shan, Liao Ning province, China. The main transformer oil chromatography monitoring IED is installed on the cabinet of main transformer intelligent components. IED not only implements the function of transformer dissolved gas monitoring and data remote transmission, but its condition monitoring master station and substation through the optical fiber communication system, and follows the IEC61850 communication protocols. The monitoring data is stored and displayed through the station level software of the monitoring canter in ± 660kV convertor station in the east Yin Chuan, Ning Xia province, China. An installation picture of DGA monitoring is shown as Figure 2. Software interface is shown as Figure 3. Real-time data fault diagnosis is shown as Figure 4.

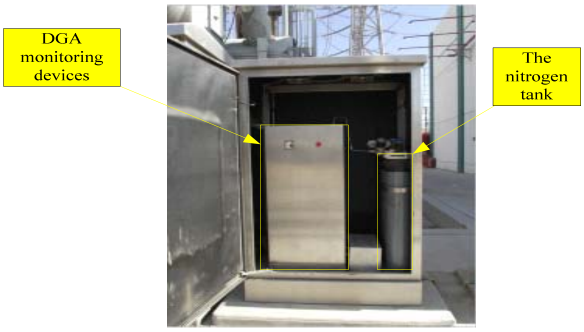


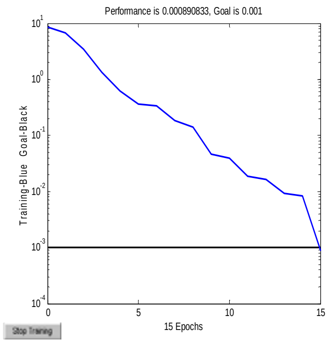
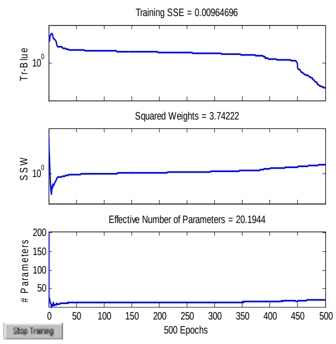
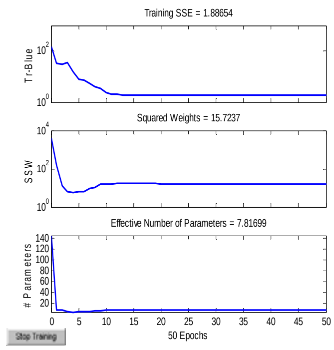
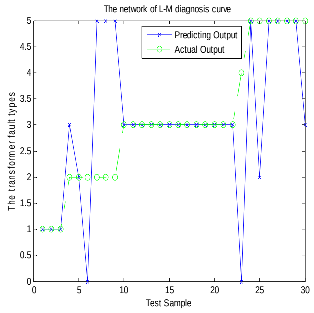
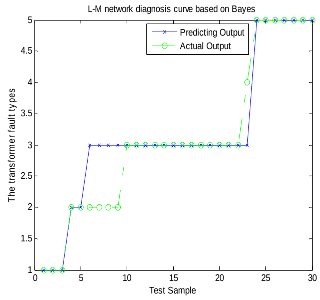
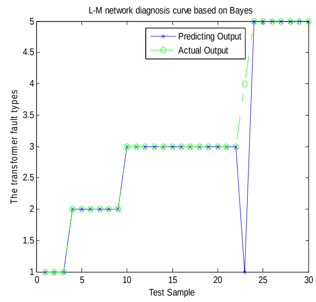
Figure 2. Installation Pictures of DGA Monitoring IED

Figure 3. Real-time Data Diagram

4.2. The Results of Simulation Testing Data

Based on the operation data of main transformer IED monitoring and the typical characteristics of the transformer fault data, there are 120 sets of data [14]. The selected 120 groups of data includes 5 kinds of characteristics gas and the corresponding transformer running state,in which 90 groups' data are as training samples of diagnostic model, while the remaining 30 groups' data are as test samples. Each group data corresponding the characteristics gas such as C_2H_2 , C_2H_4 , CH_4 , H_2 and C_2H_6 of the selected data are as the input signal of diagnostic model. The codes “1”, ”2”, “3”, ”4”, ”5” represent for ”normal”, “overheating in low-temperature”, “overheating in high-temperature”, “ spark discharge” and “arc discharge” of five kinds transformer running status, and these five kinds of operation state as output. Predicting and actual fault type curve are shown as follows.

Table 1. The Two Models Simulation Results of Power Transformer Fault Diagnosis

The model simulation results of L-M network	The model simulation results of L-M based on Bias regularization	The model simulation results of fuzzy L-M based on Bias regularization
		
		

The simulation comparison results of the three models on the number of iterations and the correct diagnosis number of sample are shown as follows:

Table 2. The Simulation Comparison Table

model	the model name	number of iterations	the correct diagnosis number of samples
a	L-M neural network	15	21
b	The fusion results of no feature reduced	23	25
c	The fusion results of feature reduced	15	29

4.3. The Analysis of Comparing with other Methods

In order to further explain the advantages of the Bias regularization of the fuzzy L-M neural network in fault diagnosis, put it in the following several prediction methods and make comparative analysis based on the 30 testing samples.

Table 3. The Simulation Comparison Table

Method	The model name	Training steps	Accurate rate
1	General gradient descent algorithm	175	72.88%
2	L-M neural network algorithm	15	76.27%
3	Bias regularization fuzzy L-M algorithm	15	89.83%

The transformer fault simulation results of the above 3 methods are shown as follows:

By comparing the training steps of methods 1 and 2: the BP network achieves the target error in the 175th step. In contrast, the L-M algorithm which is based on the adaptive adjustment to optimize network weights by the steepest gradient method and Gauss Newton method just needs 15 steps; What's more, comparing the simulation results of methods 2 and 3, the L-M neural network model has a large gap between the actual output and the expected output in the

6~9 test samples, what’s more, there are fault judgments of transformer type.

In addition, compared to other intelligent algorithm, firstly, the Bias regularization fuzzy L-M fusion model not only overcome the absolute ratio of coding boundary, but also improve the parallel processing ability. As for the 23th point, its fault type belongs to the spark discharge which is not an effective diagnosis of the fault type, because of the less fault types of training samples. The actual output and expected output of transformer are consistent with the residual test points. Secondly, by increasing the input dimension of sample datas, the L-M network overcomes the slow convergence rate and low convergence precision of the standard BP neural network. Finally, the GA method is so complex that the network is easy to fall into local optimal. Moreover, the PSO method is easy to fall into local optimum and particles are easy to reach premature convergence. Therefore, using the Bias regularization method is superior to the GA algorithm and PSO algorithm in optimizing the weight and threshold.

Thus, Bias regularization fuzzy L-M fusion model is a optimization algorithm in training rate and accuracy.

5. Conclusion

Through the contrast of the above two fusion diagnosis results, it can be concluded that the performance of Bias Regularization Fuzzy L-M Network is superior to the no feature reduction fusion model which is Bias Regularization L-M Neural Network. The former diagnosis model removes the redundant feature information, not only retaining key attributes but also fully reflecting the characteristics of inputs after the combination and optimization of the feature information. What’s more, the former diagnosis model achieves better classification results and greatly increases the accurate rate in the aspect of diagnosis results. For the former optimization algorithm which make the network error reach the expected value only after 15 iterations of training, the accuracy rate of fault diagnosis is 89.83%. And the prediction effect was far superior to the general gradient descent method and L-M algorithm. Through the analysis of examples, the information fusion fault diagnosis method based on Fuzzy coding boundary and Bias regularization L-M neural network is effective and feasible.

Acknowledgements

This paper is supported by National Basic Research Program of China(973 Program)(2009CB724507-3), Science and Technology Research and Development programs of Shaanxi Provincial Committee(2011KJXX09) and the Ministry of education about "Program for New Century Excellent talents" (NCET-11-1043).

References

[1] Li Bi-cheng, Huang Jie, Gao Shi-hai. *Information fusion technology and its application.* First Edition. Beijing: National Defence Industry Press. 2010: 13-18.
[2] Han Chong-zhao, Zhu Hong-yan, Duan Zhan-sheng. *Multi-source information fusion.* Second Edition. Beijing: Tsinghua University Press. 2010: 15-20.
[3] Peng Ningyun, Wen Xishan, Chen Jiangbo. Research on Power Transformer Fault Diagnosis with BPNN Method. *High Voltage Apparatus.* 2004; 40(3): 173-176.
[4] YAN Xue-song, WU Qing-Hua, ZHANG Can. An Improved Genetic Algorithm and Its Application. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1081-1086.
[5] Li Lin, Wan Zhicong. The research on powert ransformer insulation fault diagnosis based on the fuzzy three ratio method. *Zhe jiang Electric Power.* 2011; (2): 12-14.
[6] HE Jie, GUO Hui. A Modifified Particle Swarm Optimization Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(10): 5842-5850.
[7] Wang Yong-qiang, Lv Fang-cheng, Li He-ming. *Synthetic Fault Diagnosis Method of Power Transformer Based on Rough Set Theory and Bayesian Network.* Proceedings of the CSEE. 2006; 26(8): 138-141.
[8] Yang Li, Shang Yong, Zhou Yuefeng. *Probability reasoning andfuzzy technique applied for identifying power transformermalfunction*. Proceedings of the CSEE. 2000; 20(7): 19-23.
[9] CHENG Jia-tang, XIONG Wei, XU Shao-kun. Power Transformer Fault Diagnosis Based on Neural Networks with Improved Particle Swarm Optimization. *High Voltage Apparatus*. 2012; 48(2): 42-45.
[10] JIA Rong, XU Qi-hui, LI Hu. Power Transformer Fault Diagnosis via Neural Network Based on Particle Swarm Optimization with Neighborhood Operator. *High Voltage Apparatus.* 2008; 44(1): 8-10.

[11] ZHANG De-feng. *MATLAB neural network application design.* First Edition. Beijing: China Machine Press. 2009: 107-113.
[12] SONG Gong-yi, Guo Qing-tao, Tu Fu-rongl. *Novel Method forTransformer Faults Diagnosis Based on Theory of Fuzzy Bayesian Networks*. Proceedings of the CSU-EPSA. 2012; 24(2): 103-105.
[13] Cao Dun-kui. Gas in transformer oil analysis fault diagnosis and inspection. First Edition. Beijing: China Electric Power Press. 2005: 117-124.
[14] WANG Xiao-xia, WANG Tao. Power Transformer Fault Diagnosis Based on Neural Network Evolved by Particle Swarm Optimization. *High Voltage Engineering.* 2008; 34(11): 2362-2367.