ICIC Express Letters
Part B: Applications
Volume 7, Number 8, August 2016

ICIC International ©2016 ISSN 2185-2766
pp. 1843–1849

FAULT DIAGNOSIS OF POWER TRANSFORMER BASED ON MODIFIED DECISION DIRECTED ACYCLIC GRAPH SVM

Zhicheng Zhou[1], Le Ge[2,3], Fengbo Tao[1], Jiansheng Li[1]
and Chengshun Yang[2]

[1]Jiangsu Electric Power Company Research Institute
No. 1, Paweier Road, Jiangning Dist., Nanjing 211103, P. R. China
zhouzhicheng1977@126.com

[2]School of Electric Power Engineering
Nanjing Institute of Technology
No. 1, Hongjing Ave., Nanjing 211167, P. R. China

[3]College of Automation Engineering
Nanjing University of Aeronautics and Astronautics
No. 29, Yudao Ave., Qinhuai Dist., Nanjing 210016, P. R. China

Received January 2016; accepted April 2016

Abstract. *The method for fault diagnosis of power transformer based on the modified decision directed acyclic graph (DDAG) support vector machine (SVM) is proposed to enhance the accurate rate of the fault diagnosis of power transformer. DDAG is an extending scheme with superior performance, but its decision mainly depends on the sequence of nodes which is arbitrarily selected. So, in this paper, a modified DDAG based on node optimum algorithm is developed to improve the the accuracy of diagnosis, firstly. Then, multi-class SVM extended by the proposed method is applied to transformer fault diagnosis, and the fault diagnostic process based on modified DDAG-SVM is developed in detail. The validity and correctness of the proposed method are verified through experiments.*

Keywords: Power transformer, Fault diagnosis, Support vector machine (SVM), Decision directed acyclic graph (DDAG), Node optimum

1. Introduction.

Power transformer is one of the most crucial elements of an Electrical Power Transmission System (EPTS). However, it is inevitable to lead to the problems of aging of Oil-Paper Insulating System, material deterioration and other issues under the influence of heat, electricity and external damage during the course of long-running process. And then, those may lead to system impairments or system failures. Hence, there is an urgent need of a prefailure analysis that can protect the transformers from any kind of liabilities [1].

Analysis of the transformers dielectric oil is the classical and reliable method used for checking the irregularities presented in the transformers by using the Dissolve Gas-in-oil Analysis (DGA) method. Several gases are generated during the normal operation of a transformer. The ratio and concentration of certain gases facilitate the operator in the detection and prediction of the indiscretion and problems that exist in the transformers. The main gases responsible for the faults are hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) [2]. The International Electro technical Commission (IEC) ratio method follows the IEEE standard based on DGA, and this method has made a great contribution to discovering the potential failure of the transformers. However, in the long-term practice, the method has that a considerable part of the results are not covered in the encoding of the DGA, which resulted in certain failure cannot be diagnosed. With the persistent development of Artificial Intelligent technology, Neural Network [3], Fuzzy technology [4], and Grey System theory [5], Fuzzy Clustering Algorithm [6] are applied to transformer fault diagnosis, and achieved good recognition results. However, there are some defects in the above methods. For instance, the Neural Network and other knowledge-based methods are all based on the traditional statistics, according to the law of large numbers. So only the training sample is close to infinity, and those statistical laws can be expressed accurately. Therefore, the fault diagnosis accuracy of transformer based on the above methods is limited.

The support vector machine (SVM) is a limited sample learning method which was developed from statistical theory and it is a high performance learning algorithm that constructs a hyperplane to separate two-class data by maximizing the margin between them, and originally designed for binary. However, the fault diagnosis of power transformers is actually a problem of multi-classification. So, if the SVM is used for the fault diagnosis of the transformers, it is necessary to extend binary classification to multi-classification. And, how to effectively extend binary classification to multi-classification is an ongoing research issue. Several methods have been proposed for solving multi-class problems with SVMs, such as one-against-one (1-a-1) [7], and one-against-the-rest (1-a-r) [8]. The one-against-one method trains each binary classifier on only two out of N classes and builds $N(N-1)/2$ possible classifiers. The one-against-the-rest method constructs a set of N binary classifiers, where each ith classifier is learned from examples in the ith class and examples in the remaining classes, which are labeled with positive and negative examples, respectively. The class that corresponds to the classifier with the highest output value is used to make the final output. Based on this method, a scheme for fault diagnosis of power transformer based on multi-layer SVM classifier is proposed, and good result is got [9]. However, this method has large computation, the existence of errors and the rejection region, and the asymmetry of the positive and negative samples may lead to the over fitting problem. To reduce the calculating amount, Takahashi and Abe [10] introduce decision tree to multi-class classifier. In this method, the tree structure is used to control the decision process, and the calculation amount is reduced effectively and the blind area is also avoided. However, this method still has the problem of asymmetry of positive and negative sample in the training stage, and the method for different test data uses the same evaluation path in the decision stage, which will seriously affect the reliability of decision results.

Based on aforementioned works, in this paper, a scheme of fault diagnosis of power transformer based on modified decision directed acyclic graph (DDAG) SVM is proposed. The DDAG is introduced and its advantages and disadvantages are summarized, firstly. Then, a modified decision directed acyclic graph based on node optimum algorithm is developed. In Section 3, fault diagnosis of power transformer based on modified DDAG-SVM is presented and two experiments and some discussion are developed to show the effectiveness of the proposed method. Finally, some conclusions are made in Section 4.

2. Modified Decision Directed Acyclic Graph.

2.1. Decision directed acyclic graph.

Inspired by the directed acyclic graph (DAG) in the graph theory, Platt et al. [11] proposed a learning algorithm using the DAG to represent the classification task, which is called the DDAG. The architecture of DDAG represents a set of nodes that are concerned by edges with no cycles. Each edge has an orientation, and each node has either 0 or 2 edges. Among these nodes, there is a root node, which is the unique node with no edge pointing into it. In a DDAG, the nodes are arranged with a triangular shape where each node is labeled with an element of a boolean function. There is a single root node at the top, two nodes in the second layer, and so

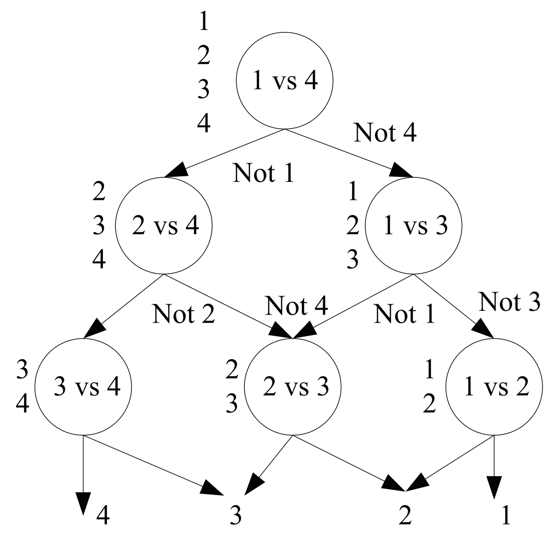


Figure 1. Decision process of the DDAG

on until the final layer of N leaves for an N-class problem. The decision process of the DDAG is shown in Figure 1.

This decision method can select the different decision paths for different dates without additional computation, and thus can improve the classification accuracy effectively. Consequently, this method provides a promising solution to extend SVM to multi-class classification issue. However, one disadvantage of the DDAG is that its classification result is affected by the sequence of binary classifier that is randomly selected in the evaluation process. To solve this problem, the method of repeated experiments can be used to select a better node arrangement. However, repeated experiments will lead to increased computational complexity. Based on above analysis, a modified decision directed acyclic graph based on node optimum algorithm is proposed in this paper.

2.2. Modified decision directed acyclic graph based on node optimum algorithm.

Theorem 2.1. *Assuming that there is a DDAG with k layer structure, and the division of risk probability of the root node is ϵ_1, and the division risk probability of nodes of the i layer is ϵ_i (Assuming that the same layer nodes have a similar division risk probability). The true classification corresponding to the test data appears for the first time in the m layer node. $E = \{\epsilon_1, \cdots, \epsilon_m, \cdots, \epsilon_{k-1}\}$. Then when the risk probability of the decision system is the minimum, there must be*

$$\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_{\tau-1} \geq \max\left(\epsilon_\tau, \epsilon_{\tau+1}, \geq \epsilon_{k-1}\right), \quad \tau = \left\lfloor \sqrt{2(k-1)+0.25} - 0.5 \right\rfloor \tag{1}$$

Proof: It is assumed that the risk probability of each layer node is $E = \{\epsilon_1, \cdots, \epsilon_m, \cdots, \epsilon_{k-1}\}$ when the risk probability of the decision system is the minimum. Then the risk probability of the decision system is

$$p = 1 - (1-\epsilon_m)(1-\epsilon_{m+1} \times \cdots \times (1-\epsilon_{k-1})) \tag{2}$$

If $\exists\ i \in [1, m-1]$, $j \in [m, k-1]$, makes $\epsilon_i < \epsilon_j$. Then, we must now conclude that

$$p = 1 - (1-\epsilon_m)\left(1-\epsilon_{m+1} \times \cdots \times (1-\epsilon_j) \times \cdots \times (1-\epsilon_{k-1})\right) \tag{3}$$

$$p^{'} = 1 - (1-\epsilon_m)\left(1-\epsilon_{m+1} \times \cdots \times (1-\epsilon_i) \times \cdots \times (1-\epsilon_{k-1})\right) \tag{4}$$

$$p^{'} < p \tag{5}$$

This contradicts the assumption that p is the minimum. So when p selects the minimum value, there must be

$$\min(E_i) \geq \max(E_i),\ E_i = \{\epsilon_1, \epsilon_2, \cdots, \epsilon_{m-1}\},\ E_j = \{\epsilon_m, \epsilon_{m+1}, \cdots, \epsilon_{k-1}\} \tag{6}$$

Because the test data are multi-class, m should be selected within an integer range. Moreover, through the observation of the structure of DDAG, we can find that all data categories are bound to occur in the first $k-1$ nodes. The number of nodes in the first m layers is $m(m+2)/2$, we have

$$m(m+1)/2 \leq (k-1) \tag{7}$$

According to (7), we obtain $m \leq \sqrt{2(k-1)+0.25} - 0.5$.

When $m = 2$, according to (6), we obtain $\epsilon_1 \geq \max(\epsilon_2, \epsilon_3, \cdots, \epsilon_{k-1})$. When $m = 3$, we obtain $\min(\epsilon_1, \epsilon_2) \geq \max(\epsilon_3, \epsilon_4, \cdots, \epsilon_{k-1})$. So, we have

$$\epsilon_1 \geq \epsilon_2 \geq \max(\epsilon_3, \epsilon_4, \cdots, \epsilon_{k-1}) \tag{8}$$

And so on, when selecting the minimum risk probability p, there will be

$$\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_{\tau-1} \geq \max(\epsilon_\tau, \epsilon_{\tau+1}, \geq \epsilon_{k-1}), \quad \tau = \lfloor\sqrt{2(k-1)+0.25} - 0.5\rfloor$$

□

However, the risk probability of the split at each node is closely related to the interval of training samples. Assuming that there is a real linear function F with unit weight vectors which is in inner product spaces Φ and $\chi \in R^+$. D is an arbitrary probability distribution on $\Phi \times \{-1, 1\}$. There is an error $E_D(f)$ such that, for all assumption of the interval $m_s(f) \geq \chi$ in ϱ random sample set S in a sphere whose center is at the coordinate origin and radius is r, with probability at least $1-\delta$, the error $E_D(f)$ has the following equation [12]:

$$E_D(f) \leq \epsilon(\varrho, F, \delta, \chi) = \frac{2}{\varrho}\left(\frac{64r^2}{\chi^2}\ln\frac{\varrho\chi}{4r}\ln\frac{128\varrho^2}{\chi} + \ln\frac{4}{\delta}\right), \quad \varrho > \frac{2}{\epsilon}, \quad \frac{64r^2}{\chi^2} < \varrho \tag{9}$$

According to (9), we can see that the the separation distance is greater and the risk probability is the smaller for the training samples at nodes. Assuming that all samples can be linearly separable by support vector machine, and then the sample interval is $d = 2\chi$. If the risk probability of the decision system is minimal, the sample interval of different layer nodes should be satisfied:

$$d_1 \leq d_2 \leq \cdots \leq d_{\tau-1} \leq \min(d_\tau, d_{\tau+1}, \cdots d_{k-1}), \quad \tau = \left\lfloor\sqrt{2(k-1)+0.25} - 0.5\right\rfloor \tag{10}$$

From Equation (10), we can know, if we reduce the probability of decision system, the sample classification intervals of the first τ layer nodes need to be arranged in ascending order. Based on the above analysis, modified decision directed acyclic graph-SVM (Modified DDAG-SVM) based on node optimum is proposed in this paper. And next, this method is applied to the fault diagnosis of the power transformer.

3. Fault Diagnosis of Power Transformer Based on Modified DDAG-SVM.

3.1. Fault classification and fault character parameter of transformer.

Usually, in the transformer fault diagnosis, the fault type of the transformer can be divided into low energy discharge (L.E.D), high energy discharge (H.E.D), partial discharge (P.D), low temperature overheating (L.O, $t < 300$), mediate temperature overheating (M.O, $300 < t < 700$) and high temperature overheating (H.O, $t > 700$). The key gases utilized to predict a specific problem are H_2 (Hydrogen) for corona in oil, C_2H_2 (Acetylene) for arcing, C_2H_4 (Ethylene) for severe overheating, CH_4 (Methane) for sparking, CO (Carbon Monoxide) for overheated cellulose and C_2H_6 (Ethane) for local overheating [13].

This paper mainly studies the fault diagnosis of overheating and discharge of the transformer, so we choose CH_4, C_2H_6, C_2H_4, H_2, C_2H_2 as the character parameter for diagnosing transformer faults.

3.2. **Fault diagnostic process based on modified DDAG-SVM.** The training sample of transformer fault is $S = \{S_1, S_2, S_3, S_4, S_5, S_6\}$. S_1, S_2, S_3, S_4, S_5, and S_6 are the fault samples of L.E.D, H.E.D, P.D, L.O, M.O, and H.O, respectively. The proposed modified DDAG-SVM method is used to classify the data in this paper and the implementation process based on MATLAB is as follows.

Step 1: 22 combinations can be combined into 15 kinds of binary classification SVM. $SVM_{i,j}$ represents the classifier obtained by the training samples S_i and S_j.

Step 2: Calculate the sample interval $d_{i,j}$ between S_i and S_j according to the support vector of $SVM_{i,j}$ and its corresponding lagrange multiplier.

Step 3: Find a class with the minimum sample interval (S_a, S_b) by $\arg\min\limits_{i,j} d_{i,j}$, $i, j \in (1, 2, \cdots, 6)$. And, (S_a, S_b) is the root node of the new process to train sample. That is $(S_a, S_b) \Longrightarrow (H_1, H_6)$ (Training samples are reordered as $H = H_1, H_2, \cdots, H_6$).

Step 4: Find the nearest sample data with (S_a, S_b) to train the nodes of next layer by the following algorithms:

$$S_c : \arg\min_i d_{a,i}(i = 1, 2, \cdots, 6, i \neq a, b),\ S_c \Longrightarrow H_2;$$
$$S_d : \arg\min_i d_{b,i}(i = 1, 2, \cdots, 6, i \neq a, b, c),\ S_d \Longrightarrow H_5.$$

Step 5: Reorder all the training samples as: $S = \{S_1, S_2, S_3, S_4, S_5, S_6\} \Longrightarrow H = \{H_1, H_2, H_3, H_4, H_5, H_6\}$ by the process of Step 4 to carry out 3 operations.

Step 6: According to Figure 1, construct the procedure for a six-type DDAG decision of the sample H. And, classification of the test data X gets the class of H_j.

Step 7: Obtain the classification results H_j of the test data based on H. According to relations table $S \Longrightarrow H$ by Step 5, get the category S_i of the test data X.

3.3. **Experiments and discussion.** In this section, we design the experimental setting to evaluate the performance of the proposed method. The experimental data are provided by Jiangsu Electric Power Company Research Institute. At the same time, we compare our method with the 1-a-1 method [7], the 1-a-r method [8] and decision tree [10], DDAG [11] to verify the effectiveness and advantage of the proposed scheme in this paper. The number of training samples and test samples for the experiment are as follows. The numbers of training samples of L.E.D, H.E.D, P.D, L.O, M.O and H.O are 25, 30, 22, 31, 28 and 25, respectively. The numbers of test samples of L.E.D, H.E.D, P.D, L.O, M.O and H.O are 20, 18, 19, 18, 20 and 27, respectively.

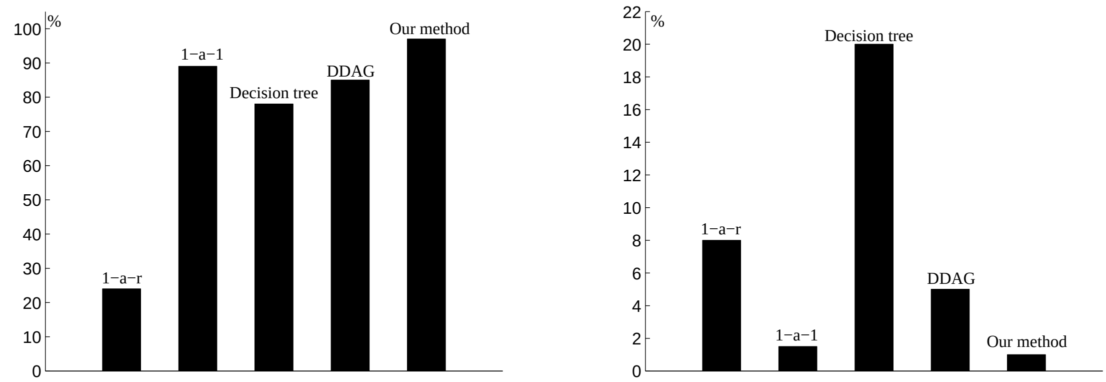
Experiment 1: Testing for the capability of diagnosis.

The methods of 1-a-r, 1-a-1, decision tree, DDAG and our method are applied to diagnose and classify the transformer fault data, in Experiment 1. Because the former two kinds of methods have classified blind area, in this experiment, the ratios of correct classification and error classification of each algorithm are statistically to test the capability of diagnosis for transformer fault.

The testing results for the capability of diagnosis are shown in Figure 2. As you can see from Figure 2, the proposed method and the 1-a-1 method have the smallest error rate. However, the result of our method has no blind area, so the correct classification rate is greater than 1-a-1 method. So, the advantages of the proposed method are verified by this experiment.

Experiment 2: Examples of fault diagnosis of power transformer.

In order to further verify the performance of the proposed method, the real fault diagnosis of power transformers is developed based on the method presented in this paper and 1-a-r method, 1-a-1 method, decision tree method, DDAG and the diagnosis results



(a) Ratios of correct classification (b) Ratios of incorrect classification

FIGURE 2. Testing result for the capability of diagnosis

TABLE 1. The fault diagnosis results based on five methods

gas concentration ($/\mu L \cdot L^{-1}$)					fault type					
H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	a.f.	1-a-1	1-a-r	d.t.	DDAG	our method
402	80.6	26.7	39.1	24.7	L.E.D	L.E.D	L.O	L.E.D	L.E.D	L.E.D
31.6	5.7	1.6	14.7	13.6	H.E.D	H.E.D	L.E.D	H.E.D	H.E.D	H.E.D
260	8	2.7	2	0	P.D	P.D	P.D	P.D	P.D	P.D
120	118	31	87	0.53	L.O	L.O	L.O	L.O	L.O	L.O
77.5	257	158	332	1.15	M.O	M.O	M.O	L.O	M.O	M.O
71.4	50.8	70.4	59.6	1.85	H.O	M.O	M.O	H.O	M.O	H.O

are shown in Table 1. In Table 1, the underlines represent the incorrect diagnosis results, and a.f., d.t. are the abbreviations for actual fault and decision tree.

The results of Table 1 show that the proposed method in this paper can make a correct diagnosis for 6 types of fault of transformer, and there are an incorrect diagnosis based on the method of 1-a-1, decision tree and DDAG, three incorrect diagnosis based on the method of 1-a-r, respectively. This further proves the validity and correctness of the proposed method in this paper.

4. **Conclusions.** DDAG is an extending strategy with outstanding performance. However, its decision largely depends on the sequence of nodes which is arbitrarily selected. So, a modified DDAG based on node optimum algorithm is to improve the accuracy of SVM-based diagnosis. Multi-class SVM extended by our method is employed as a transformer fault diagnosis, and the validity and correctness of the proposed method are verified through experiments. The important follow-ups of this work are the hardware and software implementation, leading to an electronic system that could be actually embedded in power transformers to achieve real time fault diagnosis of transformer.

Acknowledgement. This work was partially supported by Technology Project of Jiangsu Electric Power Company (J2015026), Prospective Project of University-Industry Cooperation of Jiangsu (BY201500905) and High-level Scientific Research Foundation for the Introduction of Talent of Nanjing Institute of Technology (YKJ201412).

REFERENCES

[1] F. C. Sica, F. G. Guimaraes and R. O. Duarte, A cognitive system for fault prognosis in power transformers, *Electric Power System Research*, vol.127, pp.109-117, 2015.

[2] S. Al-Janabi, S. Rawat, A. Patel et al., Design and evaluation of a hybrid system for detection and prediction of faults in electrical transformers, *Electrical Power and Energy System*, vol.67, pp.324-335, 2015.

[3] S. Souahlia, K. Bacha and A. Chaari, MLP neural network-based decision for power transformers fault diagnosis using an improved combination of Rogers and Doernenburg ratios DGA, *Electrical Power and Energy Systems*, vol.43, pp.1346-1353, 2012.

[4] V. Duraisamy, N. Devarajan and D. Somasundareswari, Neuro fuzzy schemes for fault detection in power transformer, *Applied Soft Computing*, vol.7, pp.534-539, 2007.

[5] J. Li, C. Sun, W. Chen et al., A method of synthesis based on the grey cluster and fuzzy cluster about internal fault diagnosis of transformer, *Proc. of the CSEE*, vol.23, no.2, pp.112-115, 2003 (in Chinese).

[6] Z. Song and J. Wang, Transformer DGA analysis based on fuzzy kernel clustering algorithm optimized by artificial immune, *Proc. of the CSU-EPSA*, vol.24, no.4, pp.39-43, 2014 (in Chinese).

[7] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[8] R. Anand, K. Mehrotra and C. K. Mohan, Efficient classification for multiclass problems using modular neural networks, *IEEE Trans. Neural Netw.*, vol.6, pp.117-124, 1995.

[9] A. K. Mehta, R. Naresh and S. Chauhan, Transformer diagnostics under dissolved gas analysis using support vector machine, *International Conference on Power, Energy and Control (ICPEC)*, pp.181-186, 2013.

[10] F. Takahashi and S. Abe, Decision-tree-based multiclass support vector machines, *Proc. of the 9th International Conference on Neural Information*, Washington D.C., USA, pp.1418-1422, 2002.

[11] J. C. Platt, N. Cristianini and J. Shawe-Taylor, Large margin DAG's for multiclass classification, *Proc. of Neural Information Processing Systems*, MIT Press, Massachusetts, pp.547-553, 2000.

[12] N. Cristianini, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, 1st Edition, Cambridge University Press, Cambridge, 2000.

[13] J. J. Kelly, Transformer fault diagnosis by gas-gas analysis, *IEEE Trans. Industry Applications*, vol.16, no.4, pp.777-782, 1980.