

PAPER • OPEN ACCESS

Fault Diagnosis for Transformers Based on FRVM and DBN

To cite this article: Renbo Chen *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **237** 062030

View the [article online](#) for updates and enhancements.

You may also like

- [BOF steelmaking endpoint carbon content and temperature soft sensor based on supervised dual-branch DBN](#)
Zongxu Lu, Hui Liu, Fugang Chen et al.
- [An improved fault diagnosis method for rolling bearings based on wavelet packet decomposition and network parameter optimization](#)
Fangyuan Zhao, Yulian Jiang, Chao Cheng et al.
- [An acoustic emission identification model for train axle fatigue cracks based on deep belief network](#)
Li Lin, Xiaowen Tang, Xiaoxiao Zhu et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Fault Diagnosis for Transformers Based on FRVM and DBN

Renbo Chen¹, Yue Yuan^{2*}, Zhiqiang Zhang², Xin Chen² and Feiyu He²

¹Mianyang power supply company of State Grid, Mianyang, Sichuan, 621000, China

² Sichuan University School of Electrical Engineering and Information, Chengdu, Sichuan, 610065, China

*Yue Yuan:1427288652@qq.com

Abstract. Dissolved gas analysis (DGA) of insulation oil is widely used in potential fault analysis for transformers. In order to improve the accuracy of fault diagnosis, a hybrid model which combines the FRVM with the depth belief network (DBN) is proposed to establish the mapping relationship between gas and fault types. Considering that DBN needs to extract a huge amount of feature information, this paper uses FRVM to separate the discharge and overheating faults, and then uses DBN to realize further fault diagnosis. The diagnosis accuracy is studied when IEC ratio, Rogers ratio, Dornenburg ratio and non-cod ratios are used as input parameters, and the results show that the correct rate of diagnosis is highest when the non-cod ratios are used as characteristic parameter. In addition, the method has better performance compared with single DBN, support vector machine and artificial neural network, and it has the ability to diagnose multiple faults.

1. Introduction

Power transformer is one of the key parts of power grid, Its reliability highly depends on the integrity of the insulation system [1]. In the long-term operation, transformers are often affected by heat, electricity, and mechanical forces, resulting in the deterioration of insulation structure, and then lead to faults. These faults will cause the decomposition of transformer oil, resulting in the production of various hydrocarbon gases [2].

Dissolved Gas Analysis approach (DGA) is a common fault diagnosis method for power transformers. On this basis, IEC three-ratio method and Rogers ratio method are formed, but there are some shortcomings such as coding missing and too absolute judgment standard[3]. Because of the fuzziness, uncertainty and non-linearity of DGA data, it can not be described accurately by a specified mathematical model. In the past few decades, with the development of artificial intelligence, some algorithms including artificial neural network (ANN), support vector machine (SVM) and Bayesian network have been applied in this field [4]. Artificial neural network has the shortcomings of slow convergence speed and over-fitting; Bayesian network needs a lot of sample training to get good diagnosis results; Support Vector Machine (SVM) has great advantages in solving small sample, nonlinear and high-dimensional pattern recognition, and has strong generalization ability, but the number will increase rapidly with the increase of sample number. Moreover, its kernel function should satisfy the Mercer condition and cross validation is required to verify the penalty coefficient C[5]. Relevance vector machine (RVM) is a learning method based on Bayesian framework. Kernel function does not satisfy Mercer condition and only needs a few free parameters [6]. Improved Fast relevance vector machine (FRVM) is widely used in binary classification problem [7]. Depth belief network (DBN) can extract features from a large number of samples quickly, but the features that



DBN needs to extract increase with the increase of training samples, and the structure of neural network becomes more complex. The ability to extract sample information is also getting lower and lower, which leads to lower diagnostic accuracy and greatly increases training time.

To solve this problem, it is obvious that reducing the number of features to be extracted by DBN and reducing the complexity of data are an effective way to simplify the DBN structure and improve the diagnostic accuracy. Through the research of DGA data, the article find that thermal fault and discharge fault are linearly separable. If there is a tool that can effectively classify two kinds of faults, it can greatly reduce the number of features that DBN needs to extract.

This paper presents a model which combines Fast relevance vector machine (FRVM) with DBN to solve the above problems. Considering the characteristics of DGA data, the fault samples are first classified by FRVM, and then analyzed by DBN. The research shows that the proposed method has high accuracy. Compared with single DBN, SVM and ANN, this method can extract feature information better and overcome the limitation that FRVM, SVM and ANN can only diagnose single fault.

2. Model building

2.1. Fast relevance vector machine

Power transformer RVM is a Bayesian extension of SVM. It has strong generalization ability and requires very little training data. Compared with SVM, it has the advantages that the relation vectors are thinner, the Mercer condition for kernel functions does not need to satisfied penalty term C. It has obvious advantages in dealing with binary classification problems [8]. The fast algorithm of relevance vector machine FRVM is more efficient and can deal with two classification problems well. Because the characteristics of thermal and discharge faults of transformers change obviously, they belong to two kinds of classification problems. In this paper, the trained FRVM is used to deal with the Binary classification. Suppose the training sample is $\{x_i, t_i\}_{i=1}^N$, the output of RVM is described as equation (1).

$$t_n = \sum_{i=0}^N w_i K(x_n, x_i) + \varepsilon_n \quad (1)$$

Where $K(x_n, x_i)$ is the kernel function; w_i is the weight of model, the Weight vector $w = [w_0, w_1 \dots w_N]$, N is the number of samples. The noise ε_n obeys $N(0, \delta^2)$, It can be obtained from all the above and analyses as equation (2)

$$p(t | \sigma^2, w) = \prod_{i=1}^N N(t_i | y(x_i, w), \sigma^2) \quad (2)$$

Where ϕ is $N \times (N+1)$ -matrix, and $\phi = [\varphi(x_1) \dots \varphi(x_N)]^T$. In order to simplify the learning process, let w obey Gaussian distribution $N(0, \alpha^{-1})$, the α is hyper parameter. The maximum likelihood method is used to solve w and σ^2 . The equation (3) can be obtained.

$$\begin{aligned} L(\alpha) &= \ln p(t | \alpha, \sigma^2) \\ &= -\frac{1}{2} [N \ln 2\pi + \ln |C| + t^T C^{-1} t] \end{aligned} \quad (3)$$

Where $C = \sigma^2 I + \phi A^{-1} \phi^T$, Because ϕ is calculated every iteration, the computation time is longer. FRVM has made reasonable sparsity of ϕ in the iterative process, thus improving the computational efficiency [9]

2.2. Deep belief network

Deep belief network is a generation model composed of several Restricted Boltzmann Machines (RBMs). Each layer of RBM is composed of visual layer v and hidden layer h. It can be trained in

different layers, which solves the defect that the traditional multi-layer neural network is difficult to complete the training [10].

Boltzmann machine is a feedback neural network which is completely connected by random neurons. The energy defined for a group of states (V, H) is defined as equation (4)

$$E(v, h | \theta) = -\sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} h_j \quad (4)$$

After determining the parameters, the joint probability distribution of (V, H) can be obtained based on the energy function as equation (5)

$$P(v, h | \theta) = \frac{1}{Z(\theta)} e^{-E(v, h | \theta)} \quad (5)$$

Since there is no connection between nodes in the same layer, the conditions between nodes are independent. The conditional probabilities of hidden elements and visual elements are described as equation (6) and (7)

$$p(h_j = 1 | v) = [1 + \exp(-\sum_i W_{ij} v_i - b_j)]^{-1} \quad (6)$$

$$p(v_i = 1 | h) = [1 + \exp(-\sum_j W_{ij} h_j - a_i)]^{-1} \quad (7)$$

The purpose of learning RBM is to find the model $\theta = \{W, a, b\}$. Maximum likelihood learning method can be used to obtain the parameter [11], the likelihood function is described as $L(\theta | v) = \prod_v p(v)$, the logarithmic likelihood probability C is derived, and then the parameter $\theta = \{W, a, b\}$ can be obtained by CD algorithm iteration.

For nonlinear multi-classification problems, this paper uses Softmax classifier to classify the diagnosis results. It is a generalization of Logistic classifier and can output the classification results in probability form [16]. The concrete structure of DBN is shown in the figure (1).

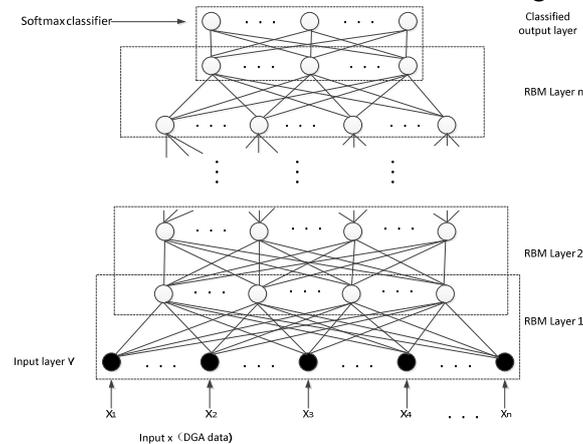


Figure 1. DBN classification model

2.3. The structure of FRVM-DBN

Many faults of power transformers are linearly separable. Therefore, the task can be transformed into a binary classification problem, and the number of training samples can be reduced by half, thus minimizing the number of rules that the DBN needs to learn. The combination of FRVM and DBN can significantly improve the efficiency of fault diagnosis.

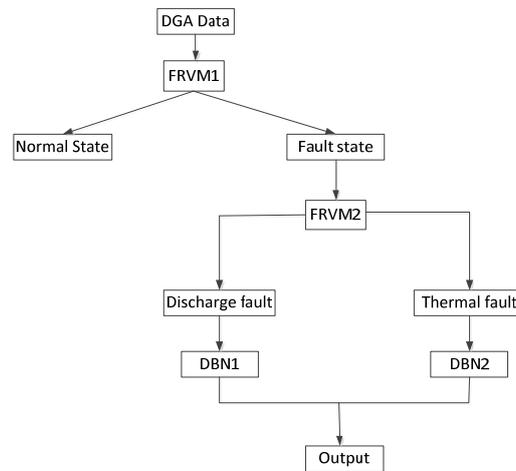


Figure 2. Structure of FRVM-DBN fault diagnosis system

3. Transformer fault classification based on FRVM-DBN

3.1. Feature parameter extraction

The H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , CO and CO_2 , are generally selected as characteristic parameters in engineering application of electric power transformer, and they are regarded as 7-dimensional vector groups. However, the concentration of characteristic gases varies greatly, so the content of the gases can not be used as input directly[12]. In order to reduce the impact of one dimension data on the whole, the content of characteristic gases is normalized according to equation (8), and the input range is limited to $[-1,1]$. In this paper, the following 4 methods are used to extract fault features as the feature pattern of input vectors.

$$x_{new} = \frac{2(x - x_{min})}{x_{max} - x_{min}} - 1 \quad (8)$$

(1) IEC ratio: This method uses the ratio of five characteristic gases as the judgment basis (C_2H_2 / C_2H_4 , C_2H_4 / C_2H_6 , CH_4 / H_2).

(2) Rogers ratios: This method also uses the ratio of five gas concentrations (CH_4/H_2 , C_2H_4/C_2H_6 , C_2H_2/C_2H_4) to form a three-dimensional vector.

(3) Doernenberg ratios: According to IEEE C57.104-2008, the method uses four gas concentrations for processing to form four-dimensional data (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_2/CH_4 , C_2H_6/C_2H_2).

(4) Non-code ratios: Because the commonly used ratio method contains less feature information, it is not conducive to transformer fault mode differential feature extraction. The main gases dissolved in the oil during transformer overheating and discharge faults vary greatly. Based on this property, the content of characteristic gases and their ratio to each other can be used as a basis for judging. Reference [19] proposes a non-coding ratio method, which consists of nine different combinations of gas ratios. Specific gas concentration ratios are: CH_4/H_2 , C_2H_4/C_2H_2 , C_2H_4/C_2H_6 , $C_2H_2/(C_1+C_2)$, $H_2/(H_2+C_1+C_2)$, $C_2H_4/(C_1+C_2)$, $CH_4/(C_1+C_2)$, $C_2H_6/(C_1+C_2)$, $CH_4+C_2/(C_1+C_2)$, $CH_4+C_2/(C_1+C_2)$, $C_2+C_2/(C_1+C_2+C_2)$. Among them, C_1 and C_2 represent hydrocarbon.

3.2. State coding and model training

Training is the key step of fault diagnosis, and training samples have a great impact on the results, so the samples must be reliable and representative. According to IEC 60599, it classifies faults into six types, and the data listed in Table 1 comes from IEC TC 10, which covers a variety of fault types. PD, D1, D2, T1, T2, T3, DT were used to express partial discharge, low energy discharge, high energy discharge, low temperature thermal, medium temperature thermal, high temperature thermal, discharge and thermal respectively. The following types of fault types are encoded.

Table 1. Code table of Transformer status

Fault type	Code
PD	(0 0 0 0 0 1)
D1	(0 0 0 0 1 0)
D2	(0 0 0 1 0 0)
T1	(0 0 1 0 0 0)
T2	(0 1 0 0 0 0)
T3	(1 0 0 0 0 0)

The training data used in this paper are mainly from the IEC TC 10 database and the DGA data published in publications. 2190 groups of samples reflecting various kinds of faults were selected from them. The samples are randomly divided into training set and test set, and the specific data arrangement is shown in Table 2.

Table 2. The partition of sample data

Fault type	Sample data	Training data	testing data
PD	300	240	60
D1	300	240	60
D2	280	224	56
T1	410	328	82
T2	400	320	80
T3	380	304	76
DT	120	96	24

3.3. Diagnostic steps

The proceeding of faults Diagnosis based on FRVM-DEN is as follow:

(1) Initialization parameters. For FRVM, the hyper parameter is set to $1/(N + 1)^2$, N represent the number of training data, the initial value of your noise variance is set to 0, and the maximum iteration value is set to 1000 times. For DBN, the network parameters W, a, B are initialized as random smaller numbers that obey Gaussian distribution, and the connection weights are random numbers that obey normal distribution $N(0, 0, 01)$. When the learning rate is set to 0.1, the relationship between the number of layers and the diagnostic accuracy is shown in Figure 3. As the number of layers increases, the accuracy increases. When the number of layers is greater than 5, the increase is not obvious. In order to improve the efficiency, the number of nodes of the neural network is set to 9-20-20-10-3, a total of five layers of neural network.

(2) Normalizing of input vectors. Output data include H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , CO, CO_2 . Because of the large differences between DGA data, the four methods described in Section 2.1 are used for standardization.

(3) Model training. In the training of FRVM, the discharge faults and overheating faults correspond to 1 and 0 respectively, and the discharge-overheating faults are taken as test data. We use CD (Contrastive Divergence) algorithm to train DBN networks by layer by layer. At the same time, the RBM parameters are adjusted properly by using the state of training samples and the reconstructed error of the calculated visual layer state to reduce the reconstructed error. Finally, we use BP algorithm to optimize the whole DBN network.

(4) Model testing. The fault classification system consists of four subsystems: two linear classification FRVM systems and two DBN systems for specific fault diagnosis. In order to test better, K fold cross validation is used to verify. Among them, the test data is divided into K, K-1 for training samples, the last for testing, its accuracy is the average of K tests. The output of DBN represents the probability of the corresponding fault type.

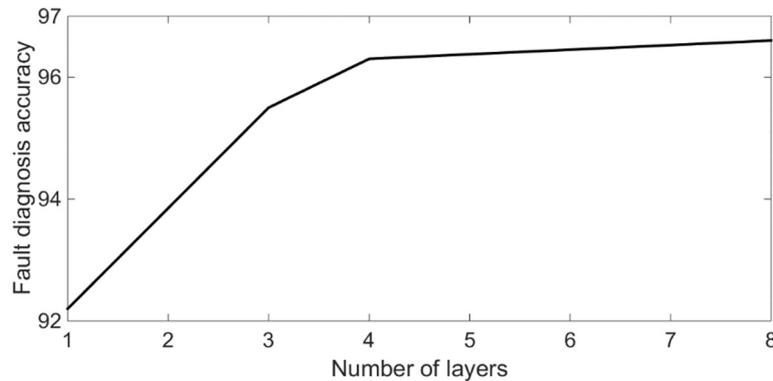


Figure 3. The relationship between RBM layers and diagnostic accuracy rate

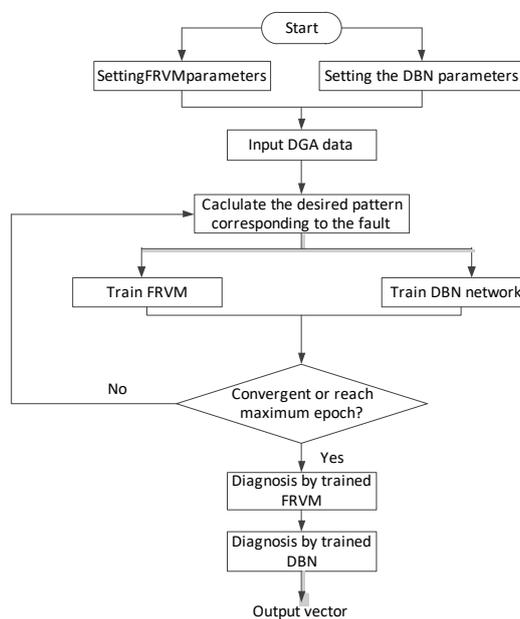


Figure 4. Diagnostic flow chart of FRVM-DBN

4. Case studies

4.1. Classification of discharge and overheating faults

This section analyzes three algorithms, FRVM, support vector machine and neural network (ANN). For SVM, penalty factor C and variance parameter γ are selected by cross validation and grid search. The range of grid search is 2^{-10} to 2^{10} , and the RBF is used as kernel. For ANN, the number of neurons in the input layer is the dimension of the input layer vector, so $N_1 = 4$ or 5 , the number of neurons in the hidden layer. The initialization values of the weights and thresholds of the neural network range from -0.5 to 0.5 . The transfer functions of the hidden layer and the output layer are selected as transfer functions and activation functions respectively. Considering that the training output of ANN is not consistent with each experiment, the K fold ($K=5, 6, 7$) cross validation is adopted.

Obviously, FRVM has the highest accuracy for all data sets that reach 99%. Therefore, FRVM is proved to be the best way to classify two classification errors. For different input patterns, the accuracy of binary classification results is also very different, and under the same classification algorithm, the accuracy of Non-code ratios are the highest, the following figure is the binary classification results.

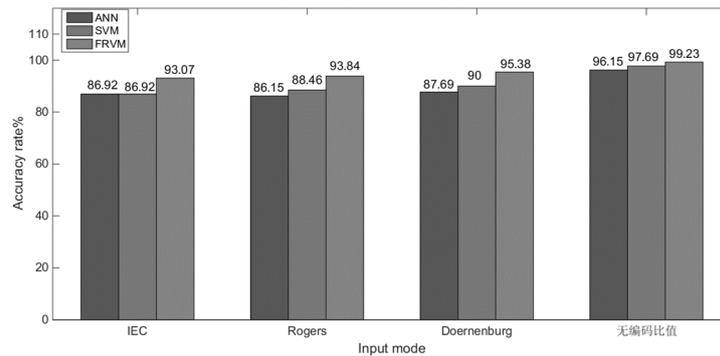


Figure 5. Linear classification results under different input modes

4.2. Fault diagnosing

The After classifying the discharge and overheating faults, DBN is used to further diagnose the fault types. In order to highlight the advantages of the proposed method, the accuracy and training time of SVM, DBN, ANN, FRVM-DBN are tested under different input modes. Table 3 gives seven kinds of test data of transformer.

Table 3 Gas concentration of fault transformers

NO	Gas concentration/($\mu\text{L}\cdot\text{L}^{-1}$)					Fault types
	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	
1	177	13.83	6.6	2.36	0	D1
2	654	55	34	20	0	PD
3	188	235	11	101	46.2	D2
4	67.66	59.7	66.64	4.98	0	T1
5	46.9	161.6	94.1	193.3	0.56	T2
6	90	149	32.4	486	19.2	T3
7	25.1	313	48	388	2.1	DT

The testing results are shown in Table 4. The IEC ratios method are the simplest and most efficient, but the encoding could make mistakes. The input vectors of other methods all adopt Non-cod ratios, because it contains more state information and can better reflect the difference information between sample data. In the same input mode, the training time of FRVM-DBN is less than that of DBN, because FRVM divides faults into two types first, which reduces the training data by half and reduces the feature quantity of DBN network. At the same time, because overheating and discharge faults are linearly separable, DBN diagnoses one kind of faults to reduce the misdiagnosis rate and improve the accuracy of fault diagnosis.

Table 4. Training time and accuracy of 4 algorithms

Method	Unable to Diagnosis	Training time/s	accuracy%
SVM	0	320	92
ANN	0	340	91.8
DBN	0	310	93.1
FRVM-DBN	0	285	96.2
IEC ratios	7	/	66.3

4.3. Multiple fault diagnosis

Transformer may have multiple faults superimposed at the same time, the output of fault probability is very close, and it is easy to blur the classification information between fault types. This kind of sample is easy to cause misdiagnosis. For all correctly diagnosed samples, the average maximum output

probability is 0.80, which is used as the FRVM-DBN threshold. That is, when FRVM-DBN is executed, if the maximum output probability of the DBN is less than the threshold, then other groups of the DBN are used, otherwise the output of the other groups is set directly to 0. Table 5 gives the diagnostic results of compound fault samples.

Table 5 The result of fault diagnosis

NO	Output of FRVM-DBN	DBN	ANN	SVM	Actual Faults
1	$\{[0.1 \ 0.87 \ 0.01]^T, [0 \ 0 \ 0]^T\}$	D1	D1	D1	D1
2	$\{[0.89 \ 0.01 \ 0.02]^T, [0 \ 0 \ 0]^T\}$	D1	D1	D1	PD
3	$\{[0.02 \ 0.03 \ 0.96]^T, [0 \ 0 \ 0]^T\}$	D2	D2	D2	D2
4	$\{[0 \ 0 \ 0]^T, [0.97 \ 0.03 \ 0.01]^T\}$	T3	T2	T3	T1
5	$\{[0 \ 0 \ 0]^T, [0.05 \ 0.85 \ 0.13]^T\}$	T2	T2	T2	T2
6	$\{[0 \ 0 \ 0]^T, [0.03 \ 0.02 \ 0.94]^T\}$	T3	T3	T3	T3
7	$\{[0.09 \ 0.37 \ 0.54]^T, [0.15 \ 0.59 \ 0.26]^T\}$	DT	T3	D2	DT

In the probability diagnostic vector set of table 5, for FRVM-DBN, $P = \{[P_{PD} \ P_{D1} \ P_{D2}]^T, [P_{T1} \ P_{T2} \ P_{T3}]^T\}$. As shown in Figure 2, when using FRVM-DBN for fault diagnosis, two DBNs classify the samples into two groups, one for thermal fault diagnosis and the other for discharge fault diagnosis. When FRVM-DBN is executed, the maximum output probability can be directly identified as the fault type. if the maximum output probability does not exceed 0.80, the discharge and overheating faults may occur simultaneously, and the samples are put into another set of DBN for testing. For example, in case 7 of Table 5, the maximum probability of DBN is less than 0.80, which indicates that the transformer fault overlaps. The maximum probabilities of the two groups are 0.54 and 0.59 respectively, indicating that high-energy discharge and medium-temperature thermal faults have occurred simultaneously. other methods are prone to Misdiagnosis, but FRVM-DBN can diagnose multiple faults

5. CONCLUSION

Aiming at the shortcomings of traditional transformer fault diagnosis methods, this paper presents a new method combining FRVM and DBN to diagnose power transformer faults. The conclusions are as follows:

The fault types are classified by FRVM firstly, and then the DBN is used for further analysis. Compared with using DBN alone, FRVM-DBN greatly reduces the feature information to be extracted, thus reducing the training time.

Compared with ANN, SVM, the hybrid FRVM-DBN has stronger learning ability that can extract implicit feature information, and well map the intrinsic relationship between DGA data and fault types. It also has a good diagnostic ability for complex faults, and the fault diagnosis accuracy is highest, which can better meet the needs of engineering.

References

- [1] Guo Xinchun, Song Qiong, FAN Xiuling. Transformer fault diagnosis based on semi-supervised classifying method [J]. High Voltage Engineering, 2013, 39(5): 1096-1100.
- [2] Sherif S. M. Ghoneim, Ibrahim B. M. Taha, and Nagy I. Elkalashy, "Integrated ANN-Based Proactive Fault Diagnostic Scheme for Power Transformer Using Dissolved Gas Analysis", IEEE Trans. Dielectr. Electr. Insul, Vol. 23, pp. 1838-1845, 2016.
- [3] Wang Yongqing, Lv Fangcheng, Li Heming, Fault diagnosis for power transformer banded on BN and DGA [J]. Transaction of china Eletro Technical Society, 2004,19(12):74-77(in Chinese).
- [4] The M. Duval, "Dissolved Gas Analysis: It Can Save Your Transformer," IEEEElectr. Insul. Mag., Vol. 5, pp. 22-27, 1989.

- [5] “IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers,” IEEE Std. C57.104-1991, 1992.
- [6] Tipping, M.E.: ‘Sparse Bayesian learning and the relevance vector machine’, *J. Mach. Learn. Res.*, 2001, 1, pp. 211–244
- [7] Hsu, C.-W., Lin, C.-J.: ‘A comparison of methods for multiclass support vector machines’, *IEEE Trans. Neural Netw.*, 2012, 13, (2), pp. 415–425
- [8] Khan S A, Equbal M D, Islam T. A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models[J]. *IEEE Trans on Dielectrics and Electrical Insulation*, 2015, 22 (1) : 590-596.
- [9] Zhagn Yiyi, Jiao Jian, Wang Ke et al. Power transformer fault diagnosis model based on support vector machine optimized by imperialist competitive algorithm[J]. *Electric Power Automation Equipment*, 2018(1):99-104.
- [10] Zhong Yongli, Yin Jinliang. Study on Application of Multi-Kernel Learning Relevance Vector Machines in Fault Diagnosis of Power Transformers[J]. *Proceedings of the CSEE*, 2013,(22):68-74.
- [11] Liang Yongliang, Li Kejun, Liu Lin, et al. A Transformer Diagnosis Method Based on optimized feature selection methods and fast relevance vector machine
- [12] Demir B , Erturk S . Hyperspectral image classification using relevance vector machines[J]. *IEEE Geoscience and Remote Sensing Letters*, 2007, 4(4): 586-590.