Cluster Computing
https://doi.org/10.1007/s10586-018-1804-0

Fault diagnosis for oil-filled transformers using voting based extreme learning machine

Liwei Zhang[1] · **Jian Zhai**[2]

Received: 10 December 2017 / Revised: 8 January 2018 / Accepted: 9 January 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
Extreme learning machine (ELM) based fault diagnosis for oil-filled transformers overcomes some drawbacks faced by that using traditional learning algorithms. Since the randomized hidden nodes are used and they remain unchanged during the training phase, some samples may be misclassified near the classification boundary. To reduce the number of such misclassified samples, fault diagnosis using voting based ELM (V-ELM) was proposed in this paper. The V-ELM-based diagnosis method incorporates multiple independent ELMs to improve the classification performance. Firstly, the user-specified parameter of individual ELM was chosen for dissolved gas analysis samples through experiment. Then, the unstable performance of individual ELM was demonstrated on testing samples. Finally, the network complexities and performance of V-ELM-based diagnosis were compared with original ELM approaches. Experimental results show that the proposed method achieves a much higher correct classification rate and the performance is more reliable.

Keywords Power transformers · Fault diagnosis · Dissolved gas analysis · Extreme learning machine · Majority voting method

1 Introduction

To satisfy both the increasing demand for power and the need to reduce carbon dioxide emissions, electrical systems will undergo a major evolution, improving reliability and reducing electrical losses, capital expenditures and maintenance costs [1]. The smart grid is the future for electrical systems, as it is designed to meet the four major electricity requirements of our global society: capacity, reliability, efficiency and sustainability.

Power transformers are considered as highly essential equipment of electric power transmission systems and often the most expensive devices in a substation. To improve the reliability of the electrical system, it will be required that power transformers must operate without failures [2].

✉ Liwei Zhang
zlwbd@outlook.com

Jian Zhai
jianzhaix@yeah.net

[1] School of Electrical Engineering, Northeast Electric Power University, Jilin, China

[2] State Grid Xi'an Electric Power Supply Company, Xi'an, China

Transformer online monitoring is the better tool to increase performance, reduce failure risks and cut maintenance costs. It is generally acknowledged that dissolved gas analysis (DGA) is normally the most powerful tool for diagnostic purposes, and for the detection of incipient faults in a transformer [3]. Around the world and during the years, several criteria have been proposed as evaluation schemes for the DGA, such as IEC 60599 [4] and IEEE Std C57.104-2008 [5]. All these conventional methods are simple and easy to implement, but their diagnostic accuracy is still limited and very sensitive to uncertainties in DGA data [6].

With the development of artificial intelligence (AI), various intelligent methods have been applied to handle the uncertainties in transformer fault diagnosis. Based on the DGA technique, an expert system was developed for transformer fault diagnosis as early as 1993 [7]. Other methods include artificial neural network (ANN) [8,9], Fuzzy system [10,11], support vector machine (SVM) [12,13], Hybrid System [3,14], etc. These novel diagnosis methods overcome the drawbacks of IEC method and improve the diagnosis accuracy. However, conventional learning methods on neural networks such as back-propagation (BP) and SVM methods apparently face some drawbacks: (1) slow learning speed,

Published online: 29 January 2018

(2) trivial human tuned parameters, and (3) trivial learning variants for different applications [15].

Extreme learning machine (ELM) is an emerging learning technique proposed for generalized single-hidden layer feed forward networks (SLFNs). ELM based transformer fault diagnosis overcomes some major constraints faced by conventional learning methods and computational intelligence techniques [16]. Since the randomized hidden nodes are used and they remain unchanged during the training phase, some samples may be misclassified in certain realizations, especially near the classification boundary [17]. Consequently, decisions based on a single realization of ELM may not be reliable and the classification results in different realizations may vary due to different nonlinear separation boundaries constructed with different random hidden node learning parameters.

To tackle the issue mentioned above, fault diagnosis of power transformers using voting based ELM (VELM) is proposed in this paper. In V-ELM, several individual ELMs are used, and all these individual ELMs are trained with the same dataset while the learning parameters of each ELM are randomly initialized independently. The final class label is determined by majority voting on all the results obtained by these independent ELMs. The rest is organized as follows. Section 2 reviews original ELM and VELM. In Sect. 3, the proposed fault diagnosis using VELM is described in detail. Section 4 discusses the user-specified parameter selection of individual ELM and comparison results of the proposed method with individual ELMs. Conclusions are finally drawn in Sect. 5.

2 Voting based extreme learning machine

2.1 Original ELM

Extreme learning machine (ELM) was originally developed for the single-hidden layer feedforward networks (SLFNs) and then extended to the "generalized" SLFNs [18]. The hidden layer in ELM need not be tuned. ELM randomly chooses the input weights and the hidden neurons' biases and analytically determines the output weights of SLFNs. Input weights are the weights of the connections between input neurons and hidden neurons and output weights are the weights of the connections between hidden neurons and output neurons. The typical SLFN architecture is shown in Fig. 1.

The output function of ELM for generalized SLFNs (take one output node case as an example) is

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} \beta_i h_i(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_L]^{\mathrm{T}}$is the vector of the output weights between the hidden layer of L nodes and the output node,

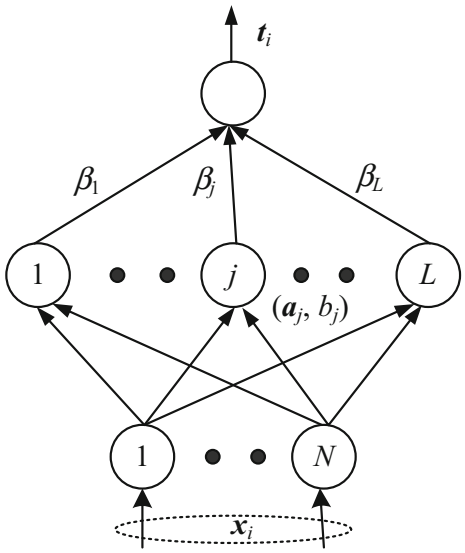

Fig. 1 Typical SLFN architecture

and $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x})]$ is the output (row) vector of the hidden layer with respect to the input $\boldsymbol{x}$. $\boldsymbol{h}(\boldsymbol{x})$ actually maps the data from the M-dimensional input space to the L-dimensional hidden-layer feature space (ELM feature space) $\boldsymbol{H}$, and thus, $\boldsymbol{h}(\boldsymbol{x})$ is indeed a feature mapping.

Given a set of training data $\{(\boldsymbol{x}_i, \boldsymbol{t}_i) | \boldsymbol{x}_i \in \mathbf{R}^d, \boldsymbol{t}_i \in \mathbf{R}^M, i = 1, \ldots, N\}$. Different from traditional learning algorithms, ELM tends to reach not only the smallest training error but also the smallest norm of output weights

$$\text{Minimize: } \|\boldsymbol{H}\boldsymbol{\beta} - \mathbf{T}\|^2 \text{ and } \|\boldsymbol{\beta}\| \tag{2}$$

where $\boldsymbol{H}$ is the hidden-layer output matrix

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}(\boldsymbol{x}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\boldsymbol{x}_1) & \cdots & h_L(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ h_1(\boldsymbol{x}_N) & \cdots & h_L(\boldsymbol{x}_N) \end{bmatrix} \tag{3}$$

and $\boldsymbol{T}$ is the expected output matrix

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{t}_N^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1M} \\ \vdots & \ddots & \vdots \\ t_{N1} & \cdots & t_{NM} \end{bmatrix} \tag{4}$$

The minimal norm least square method was used in the original implementation of ELM

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{5}$$

where $\boldsymbol{H}^{\dagger}$ is the *Moore–Penrose generalized inverse* of matrix $\boldsymbol{H}$.

The orthogonal projection method can be used to calculate the Moore–Penrose generalized inverse of $\boldsymbol{H}$ in two cases:

when $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ is nonsingular and $\boldsymbol{H}^{\dagger} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}$, or when $\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}$ is nonsingular and $\boldsymbol{H}^{\dagger} = \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{-1}$.

For multiclass classifier with multi-outputs, classifiers with M-class have M output nodes. If the original class label is p, the expected output vector of the M output nodes is $\boldsymbol{t_i} = [0, \ldots, 0, \overset{p}{1}, 0, \ldots 0]^{\mathrm{T}}$. In this case, only the pth element of $\boldsymbol{t_i} = [t_{i1}, \ldots, t_{iM}]^{\mathrm{T}}$ is one, while the rest of the elements are set to zero.

After ELM was trained, the given testing sample $\boldsymbol{x}^{\mathrm{test}}$ was taken as the input of the classifier. The index of the output node with the highest output value is considered as the predicted class label of the given testing sample. Let $f_i(\boldsymbol{x}^{\mathrm{test}})$ denote the output function of the ith output node, the predicted class label of sample $\boldsymbol{x}^{\mathrm{test}}$ is

$$label\left(\boldsymbol{x}^{\mathrm{test}}\right) = \arg \max_{i\in\{1,\ldots,M\}} f_i\left(\boldsymbol{x}^{\mathrm{test}}\right) \tag{6}$$

2.2 Voting based extreme learning machine

ELM constructs a nonlinear separation boundary in classification applications. Since the randomized hidden nodes are used and they remain unchanged during the training phase, some samples may be misclassified in certain realizations, especially for samples that are near the classification boundary. Therefore, decisions based on a single realization of ELM may not be reliable and the classification results in different realizations may vary due to different nonlinear separation boundaries constructed with different random hidden node learning parameters.

To tackle the issue mentioned above and improve the classification performance of ELM, Cao et al. proposed an algorithm referred to voting based extreme learning machine (V-ELM), which incorporates multiple independent ELMs and makes decision with a majority voting method [17].

In V-ELM, several individual ELMs with the same number of hidden nodes and the same activation function in each hidden node are used. As in ELM, the optimal number of hidden nodes for a given application usually locates in a small region and the variation of the network performance with respect to different number of hidden nodes selected from this region is usually very small. Hence, for the easy implementation, the number of hidden nodes is fixed for all the individual ELMs to the same value. All these individual ELMs are trained with the same dataset and the learning parameters of each ELM are randomly initialized independently. The final class label is then determined by majority voting on all the results obtained by these independent ELMs. Suppose that K independent networks trained with the ELM algorithm are used in V-ELM. Then, for each testing sample $\boldsymbol{x}^{\mathrm{test}}$, K prediction results can be obtained based on these independent ELMs. A corresponding vector $S_{K,x^{\mathrm{test}}} \in \mathbf{R}^M$ with dimension equal to the number of class labels is used to store all these K results of $\boldsymbol{x}^{\mathrm{test}}$, where if the class label predicted by the kth ($k \in [1, 2, ..., K]$) ELM is i, the value of the corresponding entry i in the vector $S_{K,x}$ test is increased by one, that is

$$S_{K,\boldsymbol{x}^{\mathrm{test}}}(i) = S_{K,\boldsymbol{x}^{\mathrm{test}}}(i) + 1 \tag{7}$$

After all these K results are assigned to $S_{K,x^{\mathrm{test}}}$, the final class label of $\boldsymbol{x}^{\mathrm{test}}$ is then determined by conducting a majority voting:

$$label\left(\boldsymbol{x}^{\mathrm{test}}\right) = \arg \max_{i\in\{1,\ldots,M\}} \left\{S_{K,\boldsymbol{x}^{\mathrm{test}}}(i)\right\} \tag{8}$$

In V-ELM, K independent ELMs are used where each ELM network is trained with the same dataset and the hidden node learning parameters are initialized independently. With a sufficiently large independent training number K, the number of ELMs that predicts the sample $\boldsymbol{x}^{\mathrm{test}}$ to its corresponding category is the largest one among all these K ELMs. Therefore, with a majority voting method, VELM could make a correct prediction with probability one.

The algorithmic description of the proposed V-ELM is presented in Fig. 2.

3 Transformer fault diagnosis using V-ELM

3.1 Feature vector

In a transformer, generated gases can be found dissolved in the insulating oil, in the gas blanket above the oil, or in gas collecting devices. Gases generated in oil-filled transformers can be used for qualitative determination of fault types, based on which gases are typical or predominant at various temperatures. The technique of dissolved gas analysis (DGA) is effective in detecting incipient fault of power transformers.

There are five dissolved gases in oil-filled transformers: Hydrogen (H_2), ethylene (C_2H_4), methane (CH_4), ethane (C_2H_6), and acetylene (C_2H_2), which are the by-products caused by internal faults [5]. In this paper, the concentration of the mentioned five gases are symbolized with $\varphi(H_2)$, $\varphi(CH_4)$, $\varphi(C_2H_6)$, $\varphi(C_2H_4)$ and $\varphi(C_2H_2)$ separately, and the attributes of each sample are normalized as $\{\varphi(H_2)/T,\ \varphi(CH_4)/T,\ \varphi(C_2H_6)/T,\ \varphi(C_2H_4)/T,\ \varphi(C_2H_2)/T\}$, which is taken as the feature vector in the gas-in-oil analysis. Here, T represents the total concentration of the five gases. Then, the input vector $\boldsymbol{x}$ of V-ELM could be described as

$$\begin{aligned}\boldsymbol{x} &= (x_1, x_2, x_3, x_4, x_5)\\ &= \left(\frac{\varphi(H_2)}{T}, \frac{\varphi(CH_4)}{T}, \frac{\varphi(C_2H_6)}{T}, \frac{\varphi(C_2H_4)}{T}, \frac{\varphi(C_2H_2)}{T}\right)\end{aligned} \tag{9}$$

Here, $x_i \in [0, 1]$, $i = 1, 2, \ldots, 5$, corresponding the five attributes of the input vector $\boldsymbol{x}$. Hence, the number of input nodes N of each ELM in V-ELM is set as 5.

Algorithm V-ELM

Given:
a training set $\{(\boldsymbol{x}_i, \boldsymbol{t}_i) | \boldsymbol{x}_i \in \mathbf{R}^d, \boldsymbol{t}_i \in \mathbf{R}^M, i=1,\ldots,N\}$
the hidden node output function $G(\boldsymbol{a}, b, \boldsymbol{x})$
hidden node number L
independent training number K
zero valued vector $S_K \in \mathbf{R}^M$, M is the number of classes

Training phase:
(1) Set k=1
(2) **while** ($k \leqslant K$) **do**
Randomly assign the learning parameters (a_j^k, b_j^k) (j=1,2,…,L) of the kth ELM
(3) Calculate the hidden layer output matrix $\boldsymbol{H}^k$
(4) Calculate the output weight $\boldsymbol{\beta}^k : \boldsymbol{\beta}^k = \left(\boldsymbol{H}^k\right)^{\dagger} \boldsymbol{T}$, where $\boldsymbol{T}$ is the target output matrix
(5) k=k+1
(6) **end while**

Testing phase:
(1) **for** any testing sample $\boldsymbol{x}^{\text{test}}$
(2) set k=1
(3) **while** ($k \leqslant K$) **do**
Using the kth trained basic ELM with learning parameters ($\boldsymbol{a}_j^k$, b_j^k,$\boldsymbol{\beta}_j^k$) to predict the label of the testing sample $\boldsymbol{x}^{\text{test}}$, say, as i where $i \in [1,2,...,M]$
(4) then $S_{K,\boldsymbol{x}^{\text{test}}}(i) = S_{K,\boldsymbol{x}^{\text{test}}}(i) + 1$
(5) k=k+1
(6) **end while**
(7) The final class label of the testing sample $\boldsymbol{x}^{\text{test}}$ is

$$label\left(\boldsymbol{x}^{\text{test}}\right) = \arg \max_{i \in \{1,\cdots,M\}} \left\{S_{K,\boldsymbol{x}^{\text{test}}}(i)\right\}$$

(8) **end for**

Fig. 2 Algorithmic description of V-ELM

3.2 Types of faults

The two main reasons of oil deterioration in operating transformers are thermal and electrical failures [19]. Classification of faults in IEC Publication 60599 is according to the main types of faults that can be reliably identified by visual inspection of the equipment after the fault has occurred in service. Detectable faults in IEC Publication 60599 by using DGA are discharges of low energy (D1), discharges of high energy (D2), partial discharges (PD), thermal faults below 300°C (T1), thermal faults above 300°C (T2), and thermal faults above 700°C (T3). In this paper, the fault types are symbolized with $\{i | i = 1, 2, \ldots, 6\}$ corresponding to fault type D1, D2, PD, T1, T2, and T3. Then, the fault type of each input sample could be described as

$$y = \{i | i = 1, 2, \ldots, 6\} \tag{10}$$

Hence, the number of output nodes M of each ELM in V-ELM is set as 6.

3.3 Hidden-node output functions

According to [15] and [18], almost all nonlinear piecewise continuous functions can be used as the hidden-node output functions, and thus, the feature mappings used in ELM can be very diversified. For example, such nonlinear piecewise continuous functions can be as follows.

(1) Sigmoid function

$$g(\boldsymbol{a}\boldsymbol{x} + b) = \frac{1}{1 + \exp(-(\boldsymbol{a}\boldsymbol{x} + b))} \tag{11}$$

(2) Sine function

$$g(\boldsymbol{a}\boldsymbol{x} + b) = \sin(\boldsymbol{a}\boldsymbol{x} + b) \tag{12}$$

(3) Hardlim function

$$g\left(\boldsymbol{a}\boldsymbol{x}+b\right)=\begin{cases}1, & \boldsymbol{a}\boldsymbol{x}+b\geq 0\\ 0, & \boldsymbol{a}\boldsymbol{x}+b<0\end{cases} \tag{13}$$

(4) Triangular basis function

$$g\left(\boldsymbol{a}\boldsymbol{x}+b\right)=\begin{cases}1-\left|\boldsymbol{a}\boldsymbol{x}+b\right|, & -1\leq \boldsymbol{a}\boldsymbol{x}+b\leq 1\\ 0, & \text{otherwise}\end{cases} \tag{14}$$

(5) Radial basis function

$$g\left(\boldsymbol{a},b,\boldsymbol{x}\right)=\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{a}\|^2}{b^2}\right) \tag{15}$$

Sigmoid and Radial basis functions are two of the major hiddenlayer output functions used in the feedforward neural networks and RBF networks, respectively. Interestingly, ELM with Hardlim functions can have good generalization performance as well.

3.4 Independent training numbers

For the V-ELM to work properly, a sufficiently large independent training number K is required. However, it may not be practically feasible to use V-ELM if the required K is too large, say well above 100 as the computational time of V-ELM increases with K proportionally, roughly about K times of that for a single ELM. The actual value of K depends on the application at hand. For practical applications, it is recommended to implement V-ELM by choosing K between 5 and 35, starting from 5 and gradually increasing K until 35 or a satisfactory validation result is obtained before K reaches 35. For V-ELM in this study, 7 independent ELMs are used for voting, i.e., $K = 7$.

3.5 Diagnosis model

The process of the proposed fault diagnosis using V-ELM is illustrated as Fig. 3.

The specific steps of fault diagnosis based on V-ELM are described as follows.

(1) Data processing. All the attributes (except for expected fault types) of the fault samples are normalized into the range [0,1] according to formula (9), and then the fault samples are randomly divided into training samples and testing samples in proportion. For instance, the gas concentrations of one fault sample are $\varphi(H_2) = 279\mu L/L, \varphi(CH_4) = 41\mu L/L, \varphi(C_2H_6) = 42\mu L/L, \varphi(C_2H_4) = 9.7\mu L/L, \varphi(C_2H_2) = 34\mu L/L$, and the corresponding fault type is discharges of high energy (D2). Then, the corresponding feature vector $\boldsymbol{x}$ is (0.6877, 0.1011, 0.1035, 0.0239, 0.0838), the expected output y is 5, and the input vector is (5, 0.6877, 0.1011, 0.1035, 0.0239, 0.0838).

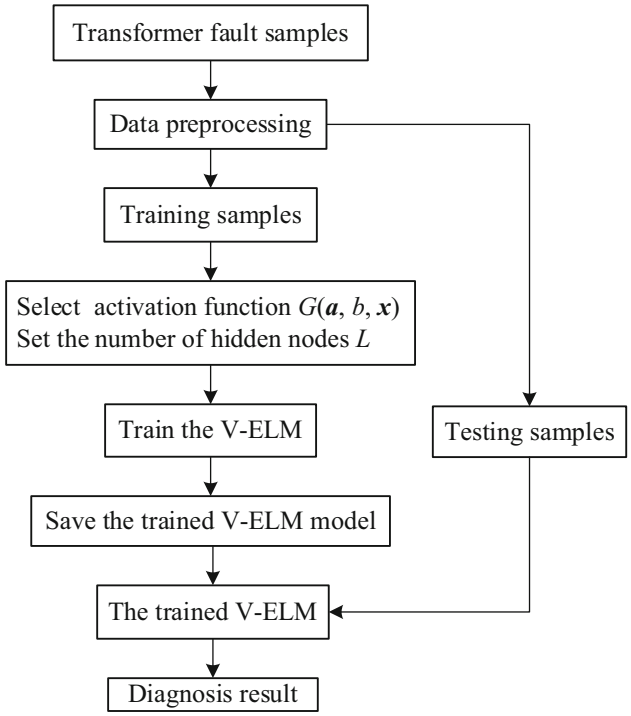

Fig. 3 Fault diagnosis procedure using KELM

(2) Training phase. Suppose that K independent networks trained with the ELM algorithm are used in V-ELM. Each individual ELM in V-ELM is with the same number of hidden nodes and the same activation function in each hidden node. All these individual ELMs are trained with the same training samples, meanwhile the learning parameters of each ELM are randomly initialized independently.

(3) Testing phase. The final fault types of testing samples are determined by majority voting on all the results obtained by these independent ELMs, according to formula (7) and (8).

4 Experiment results and discussion

In this section, the performance of transformer fault diagnosis based on V-ELM is compared with that based on the original ELM with different hidden-node output functions. All simulations are carried out in MATLAB R2015b environment running in Core(TM) i5-4300U CPU with 4GB RAM.

In this study, 387 DGA samples from real-world fault transformers are chose as experimental data. These samples were divided into two parts: 256 samples were taken as training data randomly and the rest 131 samples as testing data.

The training and testing data of all datasets are fixed for all trials of simulations.

4.1 User-specified parameter of individual ELM

As mentioned in Sect. 3.3, the hidden-node activation function used in individual ELM can be very diversified, such as Sigmoid function, Sine function, Hardlim function, Triangular basis function, and Radial basis function. In addition, in the training phase of individual ELM, all the hidden-node parameters are randomly generated based on uniform distribution. The only user-specified parameter is the number of hidden nodes L.

In order to achieve good generalization performance, the hidden-node activation function and the number of hidden nodes L need to be chosen appropriately for individual ELM in V-ELM. Simulations on DGA data set have been conducted with different activation functions mentioned above and with the number of hidden nodes L varying from 1 to 300. Simulation results are given in Figs. 4 and 5.

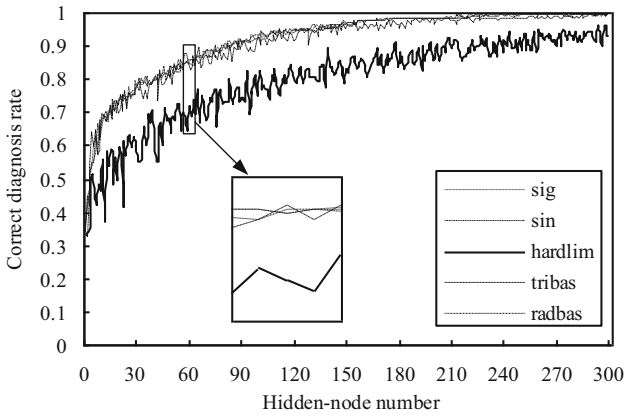

Fig. 4 Correct diagnosis rates on training dataset using different active functions

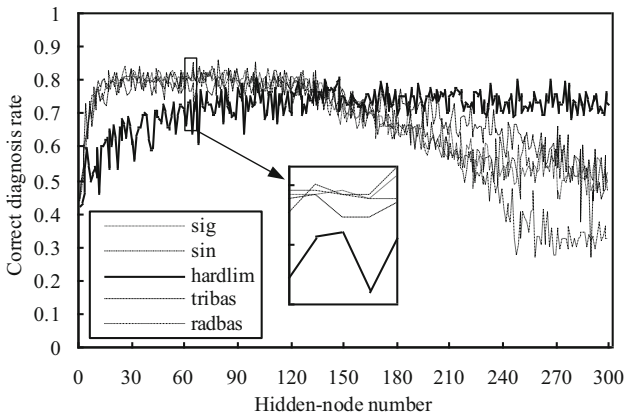

Fig. 5 Correct diagnosis rates on testing dataset using different active function

It can be seen in Fig. 4 that the correct diagnosis rate of each ELM on training dataset was close to 1 fast as the hidden-node number increased, which was with any one of the other four hidden-node activation functions except for hardlim function. Meanwhile, the correct rate of ELM with hardlim function increased slowly.

It can be observed from Fig. 5 that the correct diagnosis rate of each ELM on testing dataset, which was with any one of the other four hidden-node activation functions except for Hardlim function, increased with the hidden-node number increasing, remained stable in a certain broad interval and then decreased apparently. Meanwhile, the correct rate of ELM with Hardlim function increased slowly and tended to stable.

In ELM, the optimal number of hidden nodes for this given application usually locates in a certain broad region and the variation of the network performance with respect to different number of hidden nodes selected from this region is usually very small.

In this paper, Sigmoid function was used in V-ELM and the number of hidden nodes L was set to 100.

4.2 Performance of ELM with sigmoid function

As mentioned in the above subsection, the hidden-node activation function of individual ELM in V-ELM was Sigmoid function and the hidden-node number L was 100. Fifty trials have been conducted. Testing accuracies of the corresponding ELM on testing dataset with fifty trials are shown in Fig. 6.

Figure 6 shows that the performance of individual ELM with the same activation function and the same hidden-node number varies among 50 trials of simulations. The standard deviations of correct diagnosis rates is 2.04%. This is due to the fact that the hidden node learning parameters in ELM are randomly assigned and they remain unchanged during the training procedure.

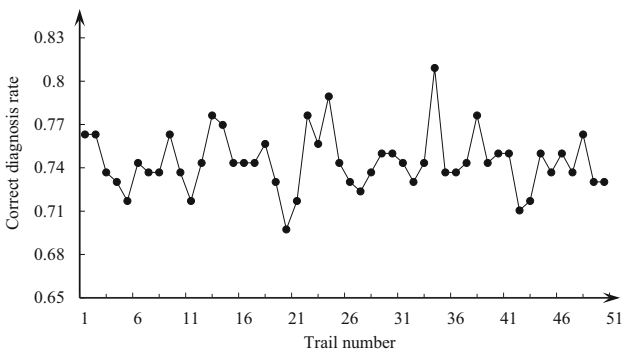

Fig. 6 Correct diagnosis rates of 50 trails on testing dataset

4.3 Performance of V-ELM

In order to improve the performance of ELM-based diagnosis method, the majority voting method is introduced into V-ELM incorporating multiple independent ELMs. For V-ELM, 7 independent ELMs are used for voting, i.e., $K = 7$. The testing performance of individual ELM in V-ELM and V-ELM are shown in Fig. 7.

From Fig. 7, it can be seen that the correct diagnosis rate of individual ELM in V-ELM is diverse from each other, while V-ELM wins the highest testing success rates in DGA datasets and the enhancement of the success rate is apparent. Take one instance to illustrate the process of making decision with a majority voting method as follow. For example, one of the normalized fault samples is (2, 0.11853, 0.037716, 0.072198, 0.70151, 0.070043), i.e., the fault type y is discharges of high energy (D2) and the feature vector $\boldsymbol{x}^{\text{test}}$ is (0.11853, 0.037716, 0.072198, 0.70151, 0.070043). Among the class labels of $\boldsymbol{x}^{\text{test}}$ determined by these independent ELMs, fault type D2 was obtained by three ELMs, fault type PD was obtained by another two ELMs, meanwhile fault type T2 was obtained by the other two ELMs. Then the final class label of $\boldsymbol{x}^{\text{test}}$ was D2 determined by conducting a majority voting, which is actual fault type. In this instance, V-ELM made the correct decision by incorporating multiple independent ELMs.

Table 1 shows the comparisons of the network complexities and performance of ELM and V-ELM with different activation functions, including the number of hidden nodes, the training time, the correct diagnosis rate and the standard deviations. Average results with 50 trials of simulations are reported in this subsection. The highest correct diagnosis rate and the lowest deviation valve are shown in bold face.

Through the comparison of ELMs with different activation functions, it can be seen that ELM with Sine function wins the highest success testing rate and the lowest deviation valve. While the success testing rate of V-ELM with Sigmoid function is higher than that with Sine function, and the deviation valve is lower. As observed from this table, the proposed V-ELM diagnosis outperforms ELM, and the standard deviations of V-ELM are much lower than individual ELM. The training time costed by V-ELM is linearly increased with a proportion to the number of independent training times used for voting. It can be found from this table that the computational time of using V-ELM is about 7 times on average that of using ELM when $K = 7$ in the above simulations.

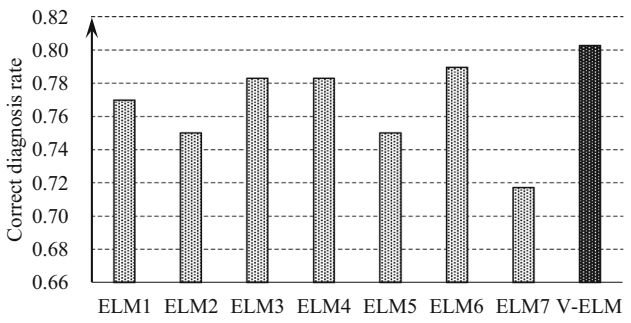

Fig. 7 Performance comparation among independent ELMs and V-ELM

5 Conclusions

In this paper, a fault diagnosis using V-ELM has been proposed for oil-filled transformers to improve the performance that using ELM. The V-ELM-based diagnosis method incorporates multiple independent ELMs to improve the classification performance. Based on the simulation results, comparisons and discussions, we have the following conclusions:

(1) Except for hardlim function, the other four activation functions mentioned in this paper are all suitable for V-ELM-based fault diagnosis. Particularly, V-ELM with Sigmoid function tends to has better performance.
(2) The optimal number of hidden nodes in ELM for this given application locates in a certain broad region, while the network performance with the same activation function and the same hidden-node number varies in a relatively narrow band.
(3) Compared with the original ELM algorithm, the incorporation of the voting method enables fault diagnosis using V-ELM achieve a much higher correct classification rate and the performance is more reliable.

Table 1 Comparisons of network complexities and performance

Algorithm	Activation function	Nodes	Training time	Correct rate	Dev.
ELM	Sigmoid function	100	**0.0428125**	0.744211	0.020360
ELM	Sine function	100	0.0465625	0.749605	0.015637
ELM	Triangular basis function	100	0.0493750	0.705658	0.027071
ELM	Radial basis function	100	0.0440625	0.742105	0.017882
V-ELM	Sine function	100	0.3037500	0.762632	0.012876
V-ELM	Sigmoid function	100	0.3178125	**0.764737**	**0.010597**

Acknowledgements The authors acknowledge the Doctoral Scientific Research Foundation of Northeast Electric Power University (no. BSJXM-201401), China.

References

1. Miceli, R., Favuzza, S., Genduso, F.: A perspective on the future of distribution: smart grids, state of the art, benefits and research plans. Energy Power Eng. **5**(1), 36–42 (2013)
2. Zeinoddini-Meymand, H., Vahidi, B.: Techno-economical lifetime assessment of power transformers rated over 50 MVA using artificial intelligence models. IET Gener. Transm. Distrib. **10**(15), 3885–3892 (2016)
3. Bakar, N., Abu-Siada, A., Islam, S.: A review of dissolved gas analysis measurement and interpretation techniques. IEEE Electr. Insul. Mag. **30**(3), 39–49 (2014)
4. Degeratu, S., Rotaru, P., Rizescu, S., et al.: Condition monitoring of transformer oil using thermal analysis and other techniques. J. Therm. Anal. Calorim. **119**(3), 1679–1692 (2015)
5. Abu-Siada, A., Islam, S.: A new approach to identify power transformer criticality and asset management decision based on dissolved gas-in-oil analysis. IEEE Trans. Dielectr. Electr. Insul. **19**(3), 1007–1012 (2012)
6. Mansour, D.E.A.: Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases. IEEE Trans. Dielectr. Electr. Insul. **22**(5), 2507–2512 (2015)
7. Mani, G., Jerome, J.: Intuitionistic fuzzy expert system based fault diagnosis using dissolved gas analysis for power transformer. J. Electr. Eng. Technol. **9**(6), 2058–2064 (2014)
8. Bhalla, D., Bansal, R.K., Gupta, H.O.: Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis. Int. J. Electr. Power Energy Syst. **43**(1), 1196–1203 (2012)
9. Illias, H., Chai, X., Abu Bakar, A.: Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. J. Int. Meas. Confed. **90**, 94–102 (2016)
10. Khan, S., Equbal, M., Islam, T.: A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models. IEEE Trans. Dielectr. Electr. Insul. **22**(1), 590–596 (2015)
11. Huang, Y.C., Sun, H.C.: Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. IEEE Trans. Dielectr. Electr. Insul. **20**(3), 974–981 (2013)
12. Li, J., Zhang, Q., Wang, K., et al.: Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. IEEE Trans. Dielectr. Electr. Insul. **23**(2), 1198–1206 (2016)
13. Wei, C., Tang, W., Wu, Q.: Dissolved gas analysis method based on novel feature prioritisation and support vector machine. IET Electr. Power Appl. **8**(8), 320–328 (2014)
14. Ghoneim, S.S.M., Taha, I.B.M., Elkalashy, N.I.: Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis. IEEE Trans. Dielectr. Electr. Insul. **23**(3), 1838–1845 (2016)
15. Huang, G.B., Zhou, H., Ding, X., et al.: Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B **42**(2), 513–529 (2012)
16. Malik, H., Mishra, S., Mittal, A.: Selection of most relevant input parameters using waikato environment for knowledge analysis for gene expression programming based power transformer fault diagnosis. Electr. Power Compon. Syst. **42**(16), 1849–1861 (2014)
17. Cao, J., Lin, Z., Huang, G.B., et al.: Voting based extreme learning machine. Inf. Sci. **185**(1), 66–77 (2014)
18. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing **70**(1–3), 489–501 (2006)
19. Roncero-Clemente, C., Roanes-Lozano, E.: A multi-criteria computer package for power transformer fault detection and diagnosis. Appl. Math. Comput. **319**, 153–164 (2018)

Liwei Zhang received the B.S.E.E. degree from Agricultural University of Hebei Province, Baoding, China, in 2007, and the Ph.D. degrees in electric power engineering from North China Electric Power University, Beijing, China, in 2014. He is currently a lecturer in the Department of Power Engineering, Northeast Electric Power University, Jilin, China. His research interests are in machine learning, pattern recognition, and condition monitoring and diagnosis of electrical equipment.

Jian Zhai received the B.S.E.E. degree and the M.S. degrees in communication engineering from North China Electric Power University, Baoding, China, in 2007 and 2010, respectively. Since 2010, he has been an electrical engineer in State Grid Xi'an Electric Power Supply Company, Xi'an, China. His research interests are in condition monitoring and diagnosis of electrical equipment and distribution network planning.