A Fault Diagnosis System of Power Transformers Using Acoustic Characteristics and Neural Network

Mingxin Geng¹, *, Chuang Fan¹, *, Kun Wang², b, Xiao Zhang², c and Zhijun Yang³,d

¹State Grid Shaanxi Electric Power Research Institute, Xi'an, China
²State Grid Xi'an Electric Power Supply Company, Xi'an, China
³State Grid Shaanxi Maintenance Company, Xi'an, China

*Corresponding author e-mail: 683945.good@163.com, a1126475605@qq.com, b18740445745@163.com, c598402556@qq.com, d675102817@qq.com

Abstract. As one of the most important facilities in the power system, power transformers undertake the crucial tasks of voltage transformation, power distribution and transmission. In the process of operation, the power transformers may have discharges, overheating, insulation degradation, winding and core loosing, solid pollution of insulation oil and some other faults. In order to address the aforementioned issues, a novel fault diagnosis system for power transformer is proposed. Through using the acoustic characteristics of the power transformer and establishing the model of the neural network, the proposed system is demonstrated with high accuracy in the experiment.

1. Introduction

The power transformers, as one of the most important facilities, take on the key tasks of voltage transformation, power distribution and transmission in the power system.

In the process of running for the power transformers, it is possible to have discharges, overheating, insulation degradation, winding and core loosing, solid pollution of insulation oil and other faults. Therefore, in-depth study on fault diagnosis methods of power transformers is of great significance to the stable operation of power system [1-3]. With the continuous development and improvement of machine learning theory, the nonlinear mapping ability, self-learning ability and fault-tolerant ability of neural network are constantly enhanced, and the application of neural network in power transformer fault diagnosis is becoming an apparent trend [4-7]. Based on the analysis technology of dissolved gas in power transformer’s oil, [8] used the content of the value of $2H$, $4CH$, $26CH$, $24CH$, $22CH$, CO and $2CO$ to train the neural network. However, in practical applications, the method requires contact or non-contact measurement of dissolved gas content in power transformer’s oil, which is cumbersome to operate, costly and difficult to realize on-line monitoring.

In this paper, we propose a fault diagnosis system of power transformers based on acoustic characteristics and the neural networks. The system can extract the frequency-domain characteristics of power transformer from the sound signals of the power transformers in operation, and then use the characteristics to train the GRU (Gated Recurrent Unit) neural networks. Particularly, the whole process of fault diagnosis of the power transformers is free from the interference of electric field and magnetic field.
2. System Design

The proposed fault diagnosis system is shown in Fig. 1, including sound acquisition device, pre-processing module, model building module and fault diagnosis module. Firstly, the sound acquisition device not only collects and records the sound when the power transformer works normally and when various types of faults occur, but also records the corresponding relationship between the sound and the status of the power transformer to be diagnosed. Note that the status include the normal status and various types of fault ones. Secondly, the pre-processing module performs Butterworth low-pass filtering, signal de-noising, feature extraction and data dimensionality reduction on the sound signal of power transformer. Thirdly, the model building module establishes the neural network model. Lastly, the fault diagnosis module uses the sound signal of power transformer to be diagnosed to make fault diagnosis based on the neural network model.

![Diagram of the proposed fault diagnosis system](image)

**Figure 1.** The frameworks of the proposed fault diagnosis system.

The workflow of power transformer fault diagnosis system mainly includes the following steps:

1. The sound acquisition device collects the sound signal of the power transformer.
   - The sound acquisition device collects and records the sound when the power transformer works normally and when various types of faults occur, and records the corresponding relationship between the sound and the state of the power transformer. Power transformer states consist of normal state and various types of fault state. The sampling frequency $f_s$ of the sound acquisition device shall be greater than or equal to the threshold value denoted as $f_t$, where $f_t=2000$Hz.

   In this work, the fault diagnosis problem of power transformer is transformed to be the classification problem of GRU neural network, and the GRU neural network is trained by means of supervised learning. The sound of power transformer is processed as the training data of GRU neural network, and the output of GRU neural network corresponds to the state of power transformer. In order to ensure that the GRU neural network can make correct prediction on the various states of the power transformer, the sound of the power transformer should include the sound when the power transformer is working normally and having various faults.

2. The pre-processing module pre-processes the sound signals collected from the power transformer.
   - (2-1) Butterworth low pass digital filter is used to filter the sound signal of power transformer.
     - The frequency domain energy of the power transformer sound is mainly concentrated in the low-frequency part, and the Butterworth low-pass digital filter carries out low-pass filtering on the sound to reduce the interference of the high-frequency part in the ambient noise. In order to achieve better
approximation of the passband and stopband, the order of Butterworth low-pass digital filter is \( N=8 \), the cut-off frequency of the passband is \( f_p=1000\text{Hz} \), the initial stopband cut-off frequency is \( f_s=1200\text{Hz} \), the minimum attenuation of the fluctuation within the passband is \( \Delta p = 1\text{dB} \), and the minimum attenuation within the stopband is 50dB.

(2-2) Signal de-noising. Wavelet decomposition is carried out for the processed sound signal, and the wavelet coefficient after that is processed as a threshold, and then the sound of power transformer is reconstructed by using the wavelet coefficient after threshold processing.

Let the sound signal processed by step (2-1) be \( S(n) \), and use db4 wavelet to decompose the five-layer wavelet. The specific process of threshold processing is as follows:

Assuming that the wavelet coefficient obtained after wavelet decomposition of sound signal is \( w_{i,j} \), and the processed wavelet coefficient is \( \hat{w}_{i,j} \). The threshold is set to \( \lambda \). If \( |w_{i,j}| \geq \lambda \), then \( \hat{w}_{i,j} = w_{i,j} \); If \( |w_{i,j}| < \lambda \), then \( \hat{w}_{i,j} = 0 \). Experientially, the threshold \( \lambda \) is set to 0.025. After that, the processed wavelet coefficients \( \hat{w}_{i,j} \) are used to reconstruct the sound \( \hat{S}(n) \).

(2-3) Feature extraction. The sound of the power transformer is normalized and processed by frame and window, and the frequency domain features are extracted for each frame to construct the one-dimensional feature matrix.

When the sound acquisition device is used to collect and record the sound of the power transformer, the distance and relative Angle between the sound acquisition device and the power transformer will affect the sound quality of the power transformer. Therefore, the reconstructed sound \( \hat{S}(n) \) in step (2-2) is normalized to make the sound amplitude in the same interval.

The specific process of normalization is as follows: In the signal de-noising processing in step (2-2), the reconstructed sound signal of the power transformer is \( \hat{S}(n) \), which consists of \( x_0, x_1, x_2, x_3, \ldots, x_n \) in the time domain. Assuming that the maximum and minimum value among them are \( x_{\text{max}} \) and \( x_{\text{min}} \), respectively. Then the normalize \( x_i \in \{x_0, x_1, x_2, x_3, \ldots, x_n\} \) can be represented as \( \frac{2 \times (x_i - x_{\text{min}})}{x_{\text{max}} - x_{\text{min}}} - 1 \).

The specific process of framing and windowing is as follows: Supposing the normalized sound signal is \( S(n) \), set the frame length as \( T \), and the frame shift as \( \alpha \). Frame segmentation means to the interception of a segment with a time length of \( T \) as a frame. The overlap between the tail of the previous frame and the head of the next frame is the frame shift \( \alpha \). The end time of the j-th frame \( (j \geq 1 \text{ and } j \text{ is an integer}) \) \( t_{\text{end}} \) is \((j-1) \times (1-\alpha) \times T \), and the beginning time \( t_{\text{start}} \) of the frame is \((j-1) \times (1-\alpha) \times T \). In the process of framing, if the time length of the remaining part of \( S(n) \) is smaller than the frame length \( T \), the part will be discarded and no framing operation will be carried out on the part. Note that the frame length \( T \) is 0.5s, and frame shift is 1/3.

Add window to the \( S(n) \) after framed. Denote j-th frame as \( f_j(n) \), and use window function to process \( f_j(n) \). Specifically, hamming window can be used to add windows, as shown in equation (1):

\[
\text{hm}(n) = \begin{cases} 
0.54 \times 1 - 0.85 \times \cos \left( \frac{2 \pi n}{N-1} \right), & 0 \leq n \leq N - 1 \\
0, & \text{else}
\end{cases}
\]
Where $N$ is the window size. Assuming that $f_j(n)$ after windowed is $\hat{f}_j(n)$, then $\hat{f}_j(n) = f_j(n) \ast h_m(n)$, where $\ast$ represents convolutional operation.

Each frame after adding window is processed by feature extraction. This paper uses the frequency domain characteristics of electric power transformer fault diagnosis. There are five approaches can be used to extract the frequency domain characteristics of $\hat{f}_j(n)$. Specifically, one can uses the DFT, STFT, MFCC, SC processing to get the frequency domain characteristics of $\hat{f}_j(n)$ named DFT, STFT, MFCC, SC features, respectively. And one also can use any of the above four kinds of approaches combination to extract the frequency domain features.

The results obtained by DFT and SC processing of $\hat{f}_j(n)$ are one dimensional matrix, but two dimensional matrix for STFT and MFCC method. In order to ensure the structure of the extracted features in the frequency domain keep the same, we need to transfer the STFT and MFCC features to be one-dimensional as DFT and SC features. In specific, both of the transformation methods are converted from two-dimensional matrix to one-dimensional matrix in a row-based manner.

In view of the extraction of multiple frequency-domain features, the first step is to determine whether the two-dimensional matrix needs to be transformed into a one-dimensional matrix. If the transformation is required, the transformation shall be carried out in the way described above, and the multiple one-dimensional matrices shall be directly splicing into the one-dimensional matrix. Otherwise, multiple one-dimensional matrices can be directly spliced into one-dimensional matrices.

After the above process of constructing the frequency-domain characteristic in one-dimensional matrix, a corresponding feature is obtained to represent each frame.

(2-4) The sample matrix is constructed and then processed by one-dimensional PCA.

In the practical operation process, the frequency domain characteristics of one dimensional matrix contains the amount of data is too large, direct expression is that the vector is very long, this will make the follow-up process a large amount of calculation. Moreover, not all data in the frequency-one-dimensional domain characteristic matrix is favorable for GRU neural network classification. Based on this assumption, this article uses the one dimensional PCA algorithm to reduce the data volume, which includes constructing the sample matrix and applying 1-D PCA.

For each one-dimensional matrix obtained in step (2-3), the data in it is extracted as a row of the sample matrix, which means, each row of data in the sample matrix has a unique frequency-domain characteristic one-dimensional matrix corresponding to it. One-dimensional PCA algorithm is used to process the sample matrix, and the first $M$ principal components whose contribution rate was greater than the threshold $\text{PCA}_{th}$ are selected from the processing results, and $\text{PCA}_{th} = 0.7$. After the sample matrix is processed by one-dimensional PCA algorithm, a two-dimensional matrix is obtained. The first $M$ column is selected from the two-dimensional matrix as the data dimensionality reduction matrix. Each row in it is the result of the dimensionality reduction processing of the corresponding one-dimensional frequency domain characteristic matrix.

(3) Establish neural network model.

Power transformer sound is a kind of time series. In the field of neural network, LSTM (Long-Short Term Memory) neural network is commonly used to process the time series. However, due to the complexity of the LSTM neural network model, the training time of the neural network model is long. GRU neural network is a simplification of LSTM neural network and can shorten the training time of the neural network model. In this paper, GRU neural network is used to process the sound signal of power transformer.

GRU neural network includes GRU neurons. Fig. 2 is the structural diagram of a single GRU neuron. $x_t$ represents the input of GRU neurons at time $t$, $h_{t-1}$ and $h_t$ are the outputs of GRU neurons at time $(t-1)$ and $t$, respectively. And the input of GRU neurons at time $t$ includes $x_t$ and $h_{t-1} \cdot \sigma$.
\( z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \), \( r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \), \( \tilde{h}_t = \text{tanh}(W \cdot [r_t \times h_{t-1}, x_t]) \)

and \( h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \), where \( W_z \) and \( W_r \) represent the weight of the update gate, and the reset gate respectively. And \( W \) is the input weight of \( \text{tanh} \).

**Figure 2.** The structural diagram of a single GRU neuron.

The establishment process of neural network model is as follows.

(3-1) Determine the number of neurons in the input layer and output layer.

The data dimension reduction matrix includes \( M \) columns, that is, each row of the data dimension reduction matrix contains \( M \) value. We take each row of the matrix as the input of the GRU neural network, so the number of neurons in the input layer is \( M \). Assuming that there are \( N \) fault types in total, the output layer of GRU neural network is set as the full connection layer, and the number of neurons in the output layer is set as \( (N+1) \).

(3-2) Determine the number of hidden layers and the number of GRU neuron nodes included in each hidden layer.

It is generally believed that if there is only one hidden layer, the neural network can fit any function as long as there are enough neuron nodes in the hidden layer. In addition, increasing the number of hidden layers can reduce the error of neural network. However, increasing the number of hidden layers and the number of neurons contained in each hidden layer will make the neural network more complex, and the training time of the neural network will be longer, which may also lead to overfitting. According to this, the GRU neural network we use includes two hidden layers. Let the number of GRU neurons contained in the first and the second hidden layer be \( h_1 \) and \( h_2 \), respectively. The first hidden layer is connected to the input layer and the second hidden layer, and the second hidden layer is connected to the first hidden layer and the output layer. The calculation of \( h_1 \) and \( h_2 \) is shown in equations (2) and (3), where \( M \) is the number of neurons in the input layer, and \( (N+1) \) is the number of neurons in the output layer.

\[
h_1 = \left\lfloor \sqrt{M \times (N+1)} \right\rfloor \tag{2}
\]

\[
h_2 = \begin{cases} 
\frac{h_1}{2}, & \frac{h_1}{2} \geq N + 1 \\
N + 1, & \frac{h_1}{2} < N + 1
\end{cases} \tag{3}
\]
(3-3) Weight Initialization.

The neural network has the problem of initial value sensitivity, and improper initialization of the weights will lead to the neural network's occurrence of training speed slow, gradient explosion, and falling into local optimal solution, etc. Therefore, we use Glorot Initialization method to initialize the weights, and the specific process is as follows:

Let the weight of the update gate and the reset gate of the GRU neurons in the first hidden layer be \( w_{iz} \), \( w_{ir} \), respectively. And then initialize them according to the distribution of formula (4) and (5):

\[
\begin{align*}
\mathbf{W}_{iz} & \sim U \left[ -\frac{\sqrt{6}}{\sqrt{M+h_1}}, \frac{\sqrt{6}}{\sqrt{M+h_1}} \right] \\
\mathbf{W}_{ir} & \sim U \left[ -\frac{\sqrt{6}}{\sqrt{M+h_1}}, \frac{\sqrt{6}}{\sqrt{M+h_1}} \right]
\end{align*}
\]  

(4) \hspace{1cm} (5)

Let the input weight of the tanh of the GRU neurons in the first hidden layer be \( w_{i\text{tanh}} \), and then initialize it according to formula (6):

\[
\mathbf{W}_{i\text{tanh}} \sim U \left[ -\frac{\sqrt{6}}{\sqrt{M+h_1}}, \frac{\sqrt{6}}{\sqrt{M+h_1}} \right]
\]  

(6)

In similar, the weight of update gate and the reset gate denoted as \( w_{2z} \) and \( w_{2r} \), and the input weight referred to \( W_{2\text{tanh}} \) of the tanh of the GRU neurons in the second hidden layer can be initialized according to formula (7)-(9):

\[
\begin{align*}
\mathbf{W}_{2z} & \sim U \left[ -\frac{\sqrt{6}}{\sqrt{h_1+h_2}}, \frac{\sqrt{6}}{\sqrt{h_1+h_2}} \right] \\
\mathbf{W}_{2r} & \sim U \left[ -\frac{\sqrt{6}}{\sqrt{h_1+h_2}}, \frac{\sqrt{6}}{\sqrt{h_1+h_2}} \right] \\
\mathbf{W}_{2\text{tanh}} & \sim U \left[ -\frac{\sqrt{6}}{\sqrt{h_1+h_2}}, \frac{\sqrt{6}}{\sqrt{h_1+h_2}} \right]
\end{align*}
\]  

(7) \hspace{1cm} (8) \hspace{1cm} (9)

The output layer of the GRU neural network is set as the full connection layer, and the weight of the output layer is set as \( w_{out} \), then it is initialized according to equation (10):

\[
\mathbf{W}_{2out} \sim U \left[ -\frac{\sqrt{6}}{\sqrt{h_2+(N+1)}}, \frac{\sqrt{6}}{\sqrt{h_2+(N+1)}} \right]
\]  

(10)

(3-4) Train the neural network.
Assuming that the reduced data matrix is:
\[
\mathbf{\bar{X}} = \begin{bmatrix}
\bar{x}_{11} & \bar{x}_{12} & \ldots & \bar{x}_{1M} \\
\bar{x}_{21} & \bar{x}_{22} & \ldots & \bar{x}_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
\bar{x}_{n1} & \bar{x}_{n2} & \ldots & \bar{x}_{nM}
\end{bmatrix}
\] (11)

Each row in \(\mathbf{\bar{X}}\), denoted as \(\bar{x}_i\), is the an input sample for training of the GRU neural network, such as \(\bar{x}_1 = [\bar{x}_{11} \ \bar{x}_{12} \ \ldots \ \bar{x}_{1M}]\), \(\bar{x}_2 = [\bar{x}_{21} \ \bar{x}_{22} \ \ldots \ \bar{x}_{2M}]\). The vector denoted as \(\bar{y}_i\) is used to represent the state of the power transformer, which includes \((N+1)\) number. The value in \(\bar{y}_i\) includes \(N\) zeros and 1 one, and the different states of the power transformer are distinguished by the different positions of the number 1. For example, \(\bar{y}_1 = [1 \ 0 \ 0 \ \ldots \ 0 \ 0]\) means that the power transformer is in normal working state. \(\bar{y}_2 = [0 \ 1 \ 0 \ \ldots \ 0 \ 0]\) means that the power transformer is in the state of fault type one. The mean variance is used as the error function. The weights of the GRU neural network are updated with the stochastic gradient descent algorithm. In the update process, if the change of mean variance within a certain number of steps is less than a threshold \(\text{loss}_{\text{thr}}\), the process of updating weights of the GRU neural network will be stopped to prevent over-fitting. After the weights of the GRU neural network stop updating, the weights of the GRU neural network calculated in the last step are saved. The certain number of steps is set to 5, and the value of \(\text{loss}_{\text{thr}}\) is 0.013458.

4) Perform power transformer fault diagnosis.

Power transformer fault diagnosis is a real-time continuous process. Sound acquisition device is used to continuously collect and record the sound of the power transformer to be diagnosed. Please refer to step (1) for the collection process. Low-pass filtering, signal de-noising, feature extraction and data dimensionality reduction are carried out for the sound of power transformer to be diagnosed collected and recorded in step (4). See step (2) for the processing process. Suppose that \(\hat{X}\) is the reduced data matrix, and \(\hat{x}_i\) is one row of \(\hat{X}\). Using vector denoted as \(\hat{y}_i\) to represent the state of power transformer, which has \((N + 1)\) elements including including \(N\) zeros and 1 one, and through the position of number 1 to distinguish different state of power transformer, which write as \(\text{state}_i\). For each \(\hat{x}_i\), there is a corresponding \(\text{state}_i\). The process of determining the state of the power transformer \(\text{state}_i\) is as follows: Firstly, with input \(\hat{x}_i\), \(\hat{y}_i\) is obtained through the GRU neural network. Then, the Euclidean distances of \(\hat{y}_i\) and every \(\text{state}_j\) are calculated, and the smallest distance is chosen as \(\text{state}_{\text{min}}\). Finally, the state of the corresponding power transformer is that of the power transformer predicted by the network. For example, after inputting \(\hat{x}_1\) into the GRU neural network, \(\hat{y}_1 = [0 \ 0.987 \ 0 \ \ldots \ 0 \ 0]\) is obtained, then, it is found that \(\hat{y}_1\) is closed to \(\text{state}_2 = [0 \ 1 \ 0 \ \ldots \ 0 \ 0]\) through calculating the Euclidean distance, and the state of the corresponding power transformer is the state of fault type 1. Finally, it is predicted that the power transformer has a fault type 1.

From the aforementioned analysis, we can find that the power transformer fault diagnosis system in this paper has the following advantages:

Use the frequency-domain characteristics of the power transformer to diagnose the fault of power transformer, which is free from the interference of electric and magnetic fields and does not affect the normal operation of power transformer. The operation is simple and the cost is low.
The use of GRU neural network shortens the training time of the neural network and improves the efficiency of fault diagnosis of power transformers.

Use Glorot Initialization method to initialize the GRU neural network and solve the initial value sensitivity problem.

During the training process of the GRU neural network, the weight update should be stopped timely by judging the range of variance of mean square deviation, so as to avoid over-fitting of the GRU neural network and improve the accuracy of fault diagnosis of power transformers.

3. Simulation and Result
In this paper, the fault diagnosis system was compared with the fault diagnosis method in [6], and four tests are conducted. The test samples are 100, 80, 60 and 40, respectively. The accuracy of each test is shown in Fig. 3. The average accuracy of the proposed fault diagnosis system is 0.905, and only 0.735 for the comparison method. Apparently, the average accuracy and accuracies in each test of the fault diagnosis system in this paper is higher than those of the method in the [6].

![Figure 3. Fault diagnosis accuracies of the proposed and the compared method.](image)

4. Conclusion
Considering the power transformer fault diagnosis problem, we put forward a fault diagnosis system based on the acoustic characteristics and the neural network. The system is able to process the sound of the power transformer, including low-pass filtering, signal de-noising, feature extraction and data dimension reduction, and it can build a neural network model. At the time of diagnosis, based on the neural network model and the audio signals of the power transformer, fault is effectively diagnosed. In particular, the experimental results show that our proposed fault diagnosis system is practical and it can achieve higher accuracy than the compared method.

Acknowledgments
This work was financially supported by State Grid Shaanxi Electric Power Company Science and Technology Project (5226KY170018).

References
[1] Qi Chen. Analysis of the Fault Diagnosis of Power Transformer [J]. Automation Application, 2018 (01): 115+120.
[2] Zhongjie Liu. Discussion on the Fault Diagnosis Method of Power Transformer [J]. Energy Technology and Management, 2017, 42 (04): 187-188.
[3] Renmin Zhang, Xueyun Huang, Xia Cao. Study on the Fault Diagnosis Method of Power Transformer [J]. Outlook of Science and Technology, 2014 (18): 108.
[4] Linfan Liu. Survey of research on Fault Diagnosis of Power Transformer Based on Machine Learning [J]. Electronic world, 2017 (15): 9-10.

[5] Xia Fei, Luo Zhijiang, Zhang Hao, etc. Application of mixed neural network in transformer fault diagnosis [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31 (01): 118-124.

[6] Longlong Chen, Bo Wang, Ling Yuan. Fault Diagnosis Method Using Neural Network for Power Transformer [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2018, 10 (02): 199-202.

[7] Zibing Tan, Xiuchao Huang, Jianwei Zhong. Power Transformer Fault Diagnosis Based on BP Artificial Neural Network [J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2018, 36 (01): 89-92.

[8] Xin Shi, Yongli Zhu. Application of Deep Learning Neural Network in Fault Diagnosis of Power Transformer [J]. Electric Power Construction, 2015, 36 (12): 116-122.