

# A Partial Discharge Fault Identification Algorithm based on SGWT Neural Network

Wei Zhang, Qiuli Wu, Yurong Deng

*Guangxi Electric Power Company Electric Power Science Research Institute,  
Nanning, 530023, China  
zhang\_w.sy@gx.csg.cn*

## **Abstract**

*Based on the second generation wavelet and information entropy, in this paper, we recognize the partial discharge pattern using the second generation wavelet (SGWT) and adaptive BP. Firstly, feature extraction of discharge signals are obtained using the SGWT and information entropy. Then, the extracted features are feed into the training BP network. The learning algorithm employed the conjugate gradient methods and the adaptive adjustment to train the error for BP network. Finally, we get the optimum training network, and the simulation results verified the feasibility of the algorithm.*

**Keywords:** *Partial discharge, wavelet entropy, conjugate gradient, SGWT*

## **1. Introductions**

The degree and type of partial discharge is an important factor used to determine abnormal electrical equipment insulation, It not only reflects the insulating state level, but also to estimate the life of electrical equipment in accordance with partial discharge type. Since the 1990s, the pattern recognition method began to be used to identify the type of partial discharges and used instead of the visual spectrum to determine discharge[1-3]. In recent years, Artificial neural network (ANN) has been increasingly used to identify the type of partial discharge. Mainly includes the use of artificial neural network for partial discharge data processing[4,5]; To construct the method of fault space division method and the combination of neural network to identify the types of partial discharge[6]; Based on the expert system to collect the theory and experience, and to identify the types of partial discharge[7]. Multilayer feedforward neural network is widely used at present, but the feature vector dimension classifier network input high, low efficiency, easy to fall into the local minimum, and the procedure is complex, difficult to promote.

In recent years, with the maturity of wavelet theory, wavelet theory has been used in many fields, and achieved good results. Wavelet theory is applied when a power failure is detected, the use of wavelet mutation signal processing capabilities, the feature vector by wavelet decomposition of the signal as the input vector neural network[8,9]. Since the second generation wavelet transform(SGWT) constructed based only on the time domain, unlike traditional wavelet-based Fourier transform needs, so it's better than the traditional wavelet calculation speed.

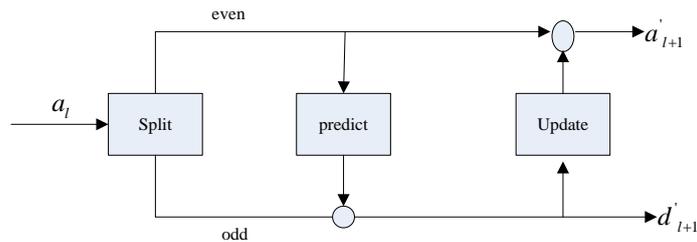
This article will second generation wavelet transform and adaptive error function are combined and used to identify the type of partial discharge. Test results show that the improved algorithm can improve the convergence speed neural network training time, optimizing the efficiency of the traditional methods and reduce the pattern recognition results mean square error, and ultimately improve accuracy.

## 2. Feature Extraction based SGWT

The basic steps SGWT structure includes decomposition and reconstruction, and reconstruction that is anti of decomposition. The partial discharge type identification algorithm that is based on the decomposition SGWT with the characteristic parameter extraction, and compression of high-dimensional feature space to low-dimensional feature space to effectively design classifiers. Since the characteristic parameter extraction involves only part of the decomposition, it details the decomposition process.

### 2.1 Decomposition Process

Decomposition of SGWT include decomposition, prediction and updating. As shown in Figure 1.



**Figure 1. Decomposition of SGWT**

**Step 1: (decomposition)** also known as inert wavelet transform process. Usually split by parity order to define the approximate coefficient of the l layer  $a_l$ , in the decomposition process, the  $a_l$  is divided into even and odd samples of  $a_{l+1}$  samples of  $d_{l+1}$ :

$$a_{l+1} = a_l(2n)$$

$$d_{l+1} = a_l(2n+1) \quad (1)$$

**Step 2: (prediction)** can be used to eliminate redundant decomposed. Need to construct prediction operator P, and act on  $a_{l+1}$ , to predict the details of the component  $d'_{l+1}$ :

$$d'_{l+1}(n) = d_{l+1} - \sum_{i=1}^M p(i)a_{l+1}(n - M_d + i - 1) \quad (2)$$

In the formula: P operator  $p(i)$  is a prediction coefficient, M is the length P,  $M_d = \frac{M}{2} - 1$ ,  $M = 2D$ ; D is a positive integer, *i.e.*, even with the M samples of the adjacent odd samples to predict the new  $d'_{l+1}$  to form a prediction error, defined  $d'_{l+1}$  for the detail coefficients of wavelet coefficients that;

**Step 3: (update)** to find a better subset of  $a'_{l+1}$ . Construction update operator U, acting on the detail coefficients  $d'_{l+1}$ , and superimposed to the even samples  $a_{l+1}$ .

$$a'_{l+1}(n) = a_{l+1} + \sum_{j=1}^N u(j)d'_{l+1}(n - N_d + j - 2) \quad (3)$$

In the formula:  $u(j)$  is an update operator U in the coefficient, N is the length of the updating operator,  $N_d = \frac{N}{2} - 1$ ,  $N = 2D'$ . And  $D'$  is a positive integer, that is predicted by the adjacent even samples N detail coefficients, the updated coefficients  $a'_{l+1}$  are approximately of  $a_l$ . So continue to make the same decomposition 111, that can get to detail coefficients and approximate coefficients of the next level.

The reconstruction of SGWT include restoration updates, recovery forecast and synthesis.

## 2.2 Feature Extraction

Setting the partial discharge signal (S) after SGWT by decomposition, the decomposition scale of the high-frequency component of the j layer factor is dj, the low-frequency component coefficient is aj, carried out after a single branch reconstructed signal component is Dj and Aj . The original signal at m-th layer can be

$$S = \sum_{i=1}^m D_i + A_m$$

expressed as a superposition of each component: . The feature vector of comprises is a energy entropy and coefficient entropy by decomposed of SGWT, specifically defined as follows:

### (1) Energy entropy

Set  $E = E_1, E_2, \dots, E_m$ , which is the spectrum of the signal S m-layer scale, then E is a division of a signal energy. And the total energy of the signal E is equal to the sum of each scale energy Ej. Set  $p_j = E_j / E$ , then  $\sum_j p_j = 1$ , its wavelet energy entropy is:

$$W = -\sum_j p_j \ln p_j \quad (4)$$

In the formula:  $E_j = \sum_k |D_j(k)|^2$ . W reflects the variation of the energy in the frequency signal.

### (2) Coefficient entropy

The signal decomposition can be decomposed sequence of each scale, regarded as the division of the signal, wavelet coefficients entropy is defined as:

$$W_j = -\sum_{i=1}^{N/2^j} \mu_j(i) \ln \mu_j(i) \quad (5)$$

$$\mu_j(i) = \frac{d_j(i)}{\sum_{i=1}^{N/2^j} d_j(i)}$$

j is the number of decomposition, N is the length of the signal.

$d_j(i)$  is an exploded detail j-th coefficient.

## 2.3 Extraction Step

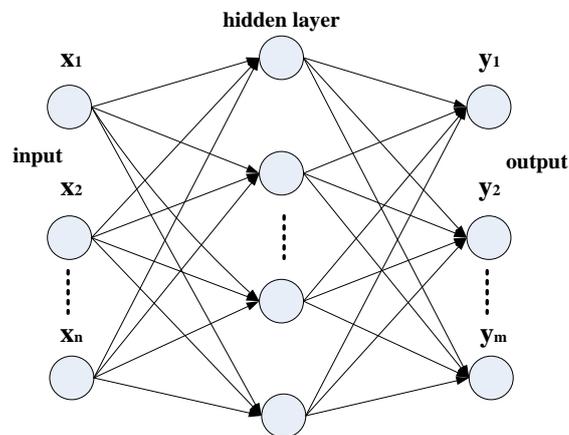
Feature extraction specific steps are as follows:

- i) Collection of various partial discharge signal.
- ii) Signals collected by SGWT decomposition, the decomposition level is 5.
- iii) extracting the layers of detail coefficients  $d_j$ , calculate the energy  $(E_1, E_2, \dots)$  of each high-frequency part measure  $(\mu_j(i))$  each division.
- iv) According to the formula (4) calculate the signal wavelet energy entropy(W);
- v) According to the formula (5) calculate the signal wavelet coefficients entropy(Wj), and gain normalization.
- vi) the  $[W, Wj]$  is the feature vector of the acquired signal.

## 3. The BP Algorithm of Adaptive Adjustment

### 3.1 Standard BP Algorithm

Feedforward neural networks of classic BP network is currently the most widely used, the diagram shown in Figure 2.



**Figure 2. Schematic Diagram of BPNN**

In an example of three-layer BP neural network, set the input nodes is  $n$ , hidden layer nodes is  $q$ , the output nodes is  $m$ . It is the number of input samples is  $P$ , the  $i$ -th value of the  $P$  input samples representative of the  $x_{pi}$ ,  $v_{ki}$  represents the weights of the  $i$ -th node of input layer to the  $k$ -th node of hidden layer,  $w_{jk}$  represents the weights of the  $k$ -th node of hidden layer to the  $j$ -th node of output layer.

The output of the  $k$ -th node of hidden layer as follows:

$$z_{pk} = f(\text{net}_{pk}) = f\left(\sum_{i=0}^n v_{ki} x_{pi}\right) \quad (6)$$

In the formula:  $k = 1, 2, \dots, q$

The output of the  $j$ -th node of the output layer:

$$y_{pj} = f(\text{net}_{pj}) = f\left(\sum_{i=0}^q w_{ki} z_{pk}\right) \quad (7)$$

In the formula:  $j = 1, 2, \dots, m$ ;

$f$  is a standard sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

f derivatives satisfy formula (7):

$$f' = f(1 - f) \quad (9)$$

Then the global error function as:

$$E = \sum_{p=1}^P E_p = \frac{1}{2} \sum_{p=1}^P \sum_{j=1}^m (t_{pj} - y_{pj})^2 \quad (10)$$

In the formula:  $E_p$  is the error of the p-th samples, the  $t_{pj}$  is the ideal output.

By using the gradient descent method for solving, to get the weight amount of correction values hidden layer and output layer neurons as formula (11), the weight amount of correction values hidden layer and input layer neurons as formula (12).

$$\Delta w_{jk} = \eta \sum_{p=1}^P \delta_{pj} z_{pk} \quad (11)$$

In the formula:  $\delta_{pj} = (t_{pj} - y_{pj}) \square y_{pj}(1 - y_{pj})$

$$\Delta v_{ki} = \eta \sum_{p=1}^P \delta_{pk} x_{pi} \quad (12)$$

In the formula:  $\delta_{pk} = (\sum_{j=1}^m \delta_{pj} w_{jk}) z_{pk}(1 - z_{pk})$

Although the structure and algorithm of feedforward network classic BP neural network are very simple and stable, but it identifies the type of partial discharge drawback is the poor generalization ability[10]. Structure and parameter selection is only on the basis of the experimental data, what the best choice of what kind of structure and the type of parameters as input is lack of theoretical basis.

### 3.2 Adaptive Adjustment Error

By the formula (11) and (12) shows that the error signal has a factor of  $y_{pj}(1 - y_{pj})$  and  $z_{pk}(1 - z_{pk})$ , That is the error signal output layer and hidden layer has an S-function derivative( $f(1 - f)$ ).When the neuron output  $y_{pj}$  and  $z_{pk}$  close to 1 or 0,  $f' \rightarrow 0$ , leading to weight adjustment becomes slow, the need to increase the number of iterations to reduce the convergence rate. The solution to this weakness is A multiplied by the weights will more than 1, with amplification factor, thus reducing the number of iterations adaptive weighting factors. And improve neural network training process convergence rate.

For this purpose, will rewrite the formula (11) and the formula (12):

$$\delta'_{pj} = -\gamma * \ln(y_{pj}(1 - y_{pj})) \cdot \delta_{pj} \quad (13)$$

$$\delta'_{pk} = -\gamma * \ln(z_{pk}(1 - z_{pk})) \cdot \delta_{pk} \quad (14)$$

From the above two formula indicates, when  $y_{pj}$ ,  $z_{pk}$  is close to 1 or 0,  $\delta'_{pj}$  and  $\delta'_{pk}$  tends to infinity, This shows that when neurons are close to 1 or 0, the error signal will become large, in order to adapt to the weight adjustment, so as to reduce the

number of iterations, improve the rate of convergence. And, the parameter  $\gamma$  may suppress or control weight adjustment process, in which  $\gamma = 1.5$ .

### 3.3 Conjugate Gradient Descent Method

Since the gradient descent method of classic BP algorithm in the training process will fall into the local minimum, so this algorithm is the use of conjugate gradient method in the training process. Conjugate gradient method is an improvement on the gradient descent method, its basic principle is to construct 1 groups conjugate direction of have been identified gradient, and thus the direction to search extrema of objective function. When implemented, the first set of search directions

$$P_0 = -g_0 \quad (15)$$

And then determine the new search directions:

$$P_k = -g_k + \beta_k P_{k-1} \quad (16)$$

In the formula:  $g_k = -\frac{\partial E}{\partial W_{ij}}$ ,  $\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$

The last available weight adjustment method:

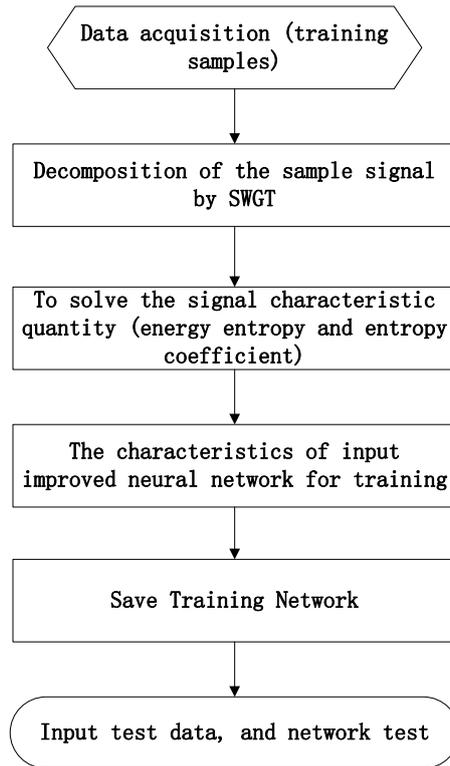
$$w_{k+1} = w_k + \lambda P_k \quad (17)$$

$\lambda$  is a learning rate, generally between 0-1, the algorithm selects 0.003.

## 4. Recognition Step

The SWGT with improved BP algorithm are combined and used to identify partial discharge patterns.

- i) The PD samples for SWGT decomposition, and the resulting wavelet energy entropy and entropy coefficient as the feature vector representation of signals.
- ii) The BP network input feature vector after modification. That is the network trained using conjugate gradient method, error in the training process is calculated by formula (13), (14).
- iii) Get the trained network, and save the weights and threshold.
- iv) Using the test samples to test.



**Figure 3. Flow Chart Recognition**

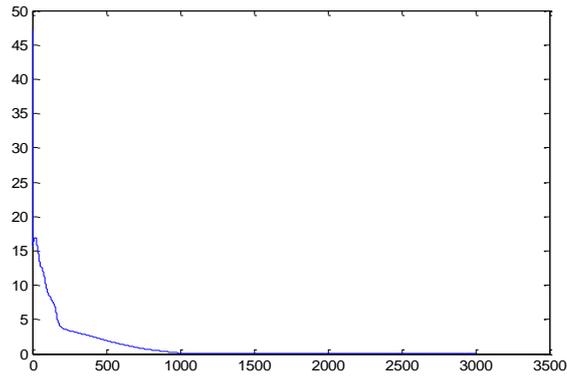
## 5. Simulation

In order to verify the algorithm specially collected three kinds of fault discharge types, each type of discharge were collected 40 sets of data. Each type of discharge data are divided into two parts; the first part of the 30 group of data as training samples, 10 groups of data of the second part as the test sample. Using the classical BP neural network, the conjugate gradient wavelet BP neural network, this paper described to training data, and comparing the results of the test.

The neural network is three layer structure; the input layer to the hidden layer activation function, the hidden layer to the output layer activation function for both S type functions use. The input feature vector of classical BP neural network belongs to the common  $\phi-q-n$ , so the nodes in input layer is 3; while the wavelet BP network input layer number will depend on the number of decomposition of wavelet for signal; the method presented in this paper are the 5 layer decomposition of signals, so the method of the input layer node number is 6.

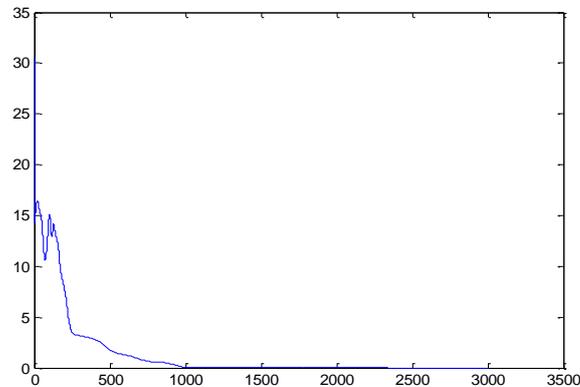
The method of the input layer node number as 6, the number of hidden layer nodes are set to 10; and since the three discharge mode (simulation of tip discharge, internal discharge, surface discharge), so in the course of training the network output is three. In simulation training, using [0 0 1] vector to represent the tip discharge, using [0 1 0] vector to represent the internal discharge, surface discharge using [1 0 0] vector to represent. The simulation result shows that the method presented in this paper in two aspects of the recognition rate and speed of convergence are better than those of the other two methods.

Figure 4 shows, in order to accurately identify the type of discharge, when 30 groups of data samples collected by classical BP neural network training, need to iterate 2980 times to the objective function converge.



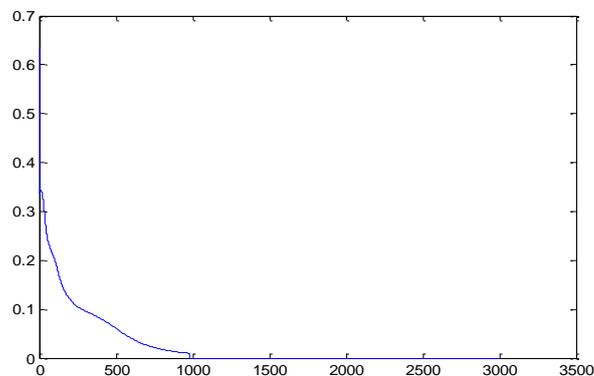
**Figure 4. The Iterations of Classic BP Neural Network Model**

Figure 5 shows that the conjugate gradient method introduced into the wavelet-based BP neural network, the number of iterations of training data at the same reduced to about 2378 times.



**Figure 5. The Iterations of Conjugate Gradient Wavelet BP Neural Network Model**

Figure 6 shows that the proposed method carries on the training data, reduces the number of iterations to about 1358 times. Thus, if the prespecified error remains unchanged, the convergence speed of BP neural network combination regulation by conjugate gradient and adaptive error will be significantly enhanced.



**Figure 6. The Iterations of the Proposed Method in this Paper**

Table 1 shows the number of iterations and recognition of the classical BP neural network, the wavelet based BP neural network and the method proposed in the paper in three kinds of discharge type recognition rate in the training. Table 1 shows the number of iterations and recognition of the classical BP neural network, the wavelet based BP neural network and the method proposed in this paper in the discharge recognition training.

**Table 1. Comparison of Results**

Comparison of content	the classical BP neural network	the wavelet based BP neural network	the method proposed in the paper
Iterations	2980	2378	1358
recognition	85%	90%	96.7%

## 6. Conclusion

Characteristics of partial discharge type is not consistent, and the lack of unified standard for calibration of severity and type. Because the wavelet signal characterization ability is very good, and the calculation of SWGT faster than the early wavelet feature extraction, so the amount of discharge signal by SWGT. In order to improve the network training phase of the training speed, and not to fall into the local extremum and partial saturation state, so the selection method of conjugate gradient descent method. And in the process of calculating the derivative of S type functions need to be adaptive adjustment of the signal error. Through simulation verification of the three kinds of actual discharge data, proved the convergence speed and the recognition rate of the proposed method in this paper is better than the classical BP algorithm of neural network, wavelet neural network BP algorithm. Also shows that the adaptive BP neural network is introduced into the SWGT can be used for the partial discharge fault recognition

## References

- [1] Gulski E, Krivds A. A Combined ANN and Expert System Tool for Transformer Fault Diagnosis [J]. IEEE Trans. on EI, **1996**, 3(2):207-212
- [2] Kranz H G. Diagnosis of partial discharge signals using neural networks and minimum distance classification [J]. IEEE Trans. on EI, **1993**, 28(6): 1016-1024
- [3] Gulski E, Morshuis P H F. *et al.* Automized Recognition of Partial Discharges in Cavities [J]. Japanese Journal of Applied Physics, **1990**, 29(7): 1329-1335
- [4] Danikas M G, Gao N, Aro M. Partial discharge recognition using neural networks: a review [J]. Electrical Engineering, **2003**, 85(2): 87—93.
- [5] T. Hong, D. Hilder, M. T. CFang. PD Classification by Modular Neural Networks Based on Task Decomposition [J]. IEEE Trans. on D.I, **1996**, 3(2): 112 - 114
- [6] Du Boxue, Wei Guozhong. Partial discharge classification based on wavelet and fractal theory [J]. Power System Technology, **2006**, 30(13): 76-80 (in Chinese).
- [7] Hozumi N, T. Okamoto, T. Imajo. Discrimination of Partial Discharge Patterns Using Neural Network [J]. Proc. of the 3rd IC-PADM, Tokyo, **1991**: 476-478
- [8] Si Wenrong, Li Junhao, Li Yanming, *et al.* Digital detection grouping and classification of partial discharges signals at DC voltage [J]. IEEE Trans. on Dielectrics and Electrical Insulation, **2008**, 15(6): 1663-1674.
- [9] QIAN Yong, HUANG Chengjun, CHEN Chen, JIANG Xiuchen. A Multi-wavelet Based Algorithm for Extracting Diversiform Partial Discharge Signals [J]. Automation of Electric Power Systems, **2007**, 31(7): 75-80. [10] PAN Chong, CHEN Weigen, YUN Yuxin, DU Lin, SUN Caixin. Fault Diagnosis of Power Transformers Based on Genetic Algorithm Evolving Wavelet Neural Network [J]. Automation of Electric Power Systems, **2007**, 31(13): 88-92.

## Authors



**Wei Zhang;** Mainly engaged in electrical equipment condition monitoring and diagnostic research work.