

Recognition of Transformer High Frequency Partial Discharge Based on Time Domain Feature

Meng Liu*, Ying Lin*, Qingdong Zhu*, Junxin Wang*, Yi Yang*, Chao Gu*, Wenbing Zhu*, Wenjie Zheng*, Demeng Bai*, Jiafeng Qin*, Jian Wang*, Fengda Zhang*, Zhuangzhuang Li*

*State Grid Shandong Electric Power Research Institute, No. 2000 Wangyue Road Jinan China

liumengup@126.com

Abstract— Power transformer is an important equipment of power system. The inner high voltage will cause transformer insulation deterioration fault, which can cause large-scale power outage. Partial discharge (PD) is an important phenomenon of transformer insulation degradation, so the research, diagnosis and identification of transformer partial discharge is very necessary. In this paper, recognition of partial discharge based on a time domain feature extraction is proposed. We collected PD signals through a high-voltage partial discharge device, using high-frequency current sensors. The discharge types include four kinds of defects: oil gap defect, floating defect, tip defect and surface defect. Wavelet transform was used to remove noise interference, and convolutional Neural Network was utilized to extract and identify signal features.

Keywords— Partial discharge recognition, high frequency, time domain, CNN

I. INTRODUCTION

High voltage and informatization is the main research points of the strong smart power grid, and the power transformer is the most basic and core equipment of the power system. With the gradual increase of the voltage level, the transformers insulation degradation will rapidly develop into a fault, cause a large-scale blackout and a serious damage of the production and life. Therefore, it is very important to monitor the insulation status of the transformer, and partial discharge is an important reference for the insulation performance of the transformer. The recognition of the partial discharge types helps to assess the form and degree of transformer insulation deterioration, find hidden dangers in time, and avoid transformer failures.

The types of transformer partial discharge monitoring mainly include: ultrasonic, ultra-high frequency (UHF), Electrochemical, and high frequency (HF).

Ultrasonic comes from partial discharge pulse. Due to the expansion and contraction of the discharge area within a short period of time, the surrounding matter vibrates instantaneously. The frequency band of the ultrasonic peak frequency generated by partial discharge is 50KHz ~ 200KHz[1]. The ultrasonic is sensitive to PD signal, but the interference of mechanical vibration and the noise in the environment will serious.

Ultra-high Frequency electromagnetic is also generated in the process of partial discharge, and bandwidth of the obtained UHF electromagnetic wave is usually 300MHz-3000MHz[2]. According to the characteristics of the partial discharge, Hikita [3] et al used the ultra-high frequency method to measure the partial discharge signal of GIS for the first time. Gao Hongwu et al [4] found that the oil-paper barrier has a certain influence on UHF detection. Different layers of insulating paper have an influence on the measured waveform amplitude and the power spectrum of a specific section.

Electrochemical is based on the chemical reaction of insulating organic matter after the partial discharge. This process produces new chemical substances, like hydrocarbons, CO₂ and other substances, and the composition ratios of the new chemical substances is useful to detects the PD types. These characteristic gases constitute the dissolved gas analysis (DGA) method [5], which can be used to diagnose some common faults in transformers. This kind of method is usually used as detection due to the stability of serious discharge, but it also has the disadvantage of low sensitivity, and it is not sensitive to low-energy discharge.

High-frequency electromagnetic is the only partial discharge measurement method with specific standards IEC60270-2000, which can be tested both offline and online. This advantage makes it the most widely used. Generally, the pulse current method is calibrated first, and then the parameters are measured. The narrow-band pulse current measurement and the frequency band width is 9~30kHz, and it has better anti-interference performance. However, this method can only measure low-frequency component pulse signals, and the signal feature is also insufficient. Wide pulse current is a measurement bandwidth of several MHz to tens of MHz. By using a sensor to expand the bandwidth, a signal characteristic with more discharge details can be obtained.

Researchers have studied in the field of partial discharge recognition for many years, and have proposed many methods [6], including Support vector machine (SVM), Artificial neural network (ANN), Random forest (RF), Deep learning, and so on.

SVM is a two-classification model, and its basic model is a linear classifier with the largest interval defined in the feature space. The significant factor of the performance is the kernels of SVM, which include linear-kernel, polynomial kernel,

Gaussian or radial basis function (RBF) kernel and so on. Firuzi et al combined SVM with local binary pattern (LBP) and histogram of oriented gradient (HOG) features of PRPD in transformer applications [7]. Raymond et al extracted PRPD features by PCA, then recognised the PD using SVM. The result shows high noise tolerance capability and better diagnosis accuracy [8]. Ying Lin et al present a feasible algorithm to reconstruct signals in a GIS partial discharge graph through image processing[9]

ANN is a classical PD diagnostics method which is inspired by biological neural networks. Lumba et al investigated the BPNN for PD recognition using statistical features of PRPD [10]. Sukma et al proposed an comprehensive BPNN PD classification method, using waveform parameters and PRPD of different types of sensors as input [11]. Soltani et al proposed an ANN denoising algorithm to remove the influence of noise in applications. This kind of method does not require prior knowledge of the signals and can achieve better performance[12].

RF is proposed by Leo Breiman using multiple trees to train and predict samples. Iorkyase et al utilized WPD to extract location features from PD signals using three R-F sensors and RF regression to localize PD [13]. Wang et al used isolation forest (IF), linear prediction cepstrum and PCA to separate multisource PD signals in GIS and transformer [14]. Peng et al proposed an RF based optimal feature selection for PD classification in HV cables [15].

Deep learning is a further development of neural networks, learning the internal laws and representation levels of sample data. Duan et al developed an end-to-end stacked sparse auto-encoder (AE) method for PD classification, the PD current signal captured by high frequency circuit transform (HFCT). The experimental results shows that single sparse AE combined with a soft-max classifier is sufficient to achieve satisfactory performance [16]. Adam et al [17] invented a transformer PD classification using raw data of PD pulse waveforms and the LSTM model. Song [18] proposed a CNN method to recognise PD using UHF PRPS raw data as input. The initial parameters of CNN are from unsupervised learning of an AE model. The experimental results shows a improve classification accuracy compared to conventional ML based methods based on SVM and BPNN.

Above all, high frequency PD recognition using deep learning method has wide prospect. This work focus on the partial discharge recognition based on a time domain feature extraction and CNN. Wavelet transform was used to remove noise interference, and convolutional Neural Network was utilized to extract and identify signal features.

II. TECHNOLOGY AND IMPLEMENTATION

A. PD signal acquisition system

Transformer PD signal is collected using the system shown in Figure 1. The PD devices are deployed inside the transformer, and the structures of four types of PD are shown in Figure 2. The high-frequency sensors are employed at the transformers ground wire to couple the high-frequency pulse current signal. The real-time waveform and synchronous

voltage signal are collected by the oscilloscope. The high-frequency partial discharge sensor adopts TWHCT-7850 with the 3MHz-30MHz frequency band, and the oscilloscope is R&S RTO-2044.



Figure 1. The PD simulation device, high frequency sensors and oscilloscope.

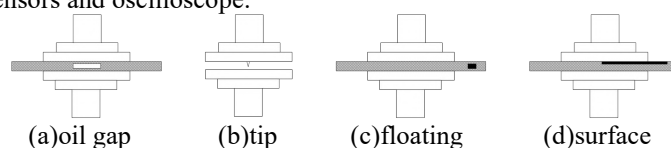


Figure 2. The four type PD defect

B. Method and Approach

The signal collected by the oscilloscope is shown in Figure 3. It can be seen from the figure that the noise has a great influence. In order to reduce the noise influence, the wavelet transform is used. Wavelet transform is an effective transform analysis method, which can refine the high frequency and low frequency of the signal through the expansion and translation operation. The Daubechies5 wavelet with orthogonality and tight support is used here. As a sparse base, this wavelet is not easy to produce large smoothing errors, that is, the reconstructed signal is smoother.

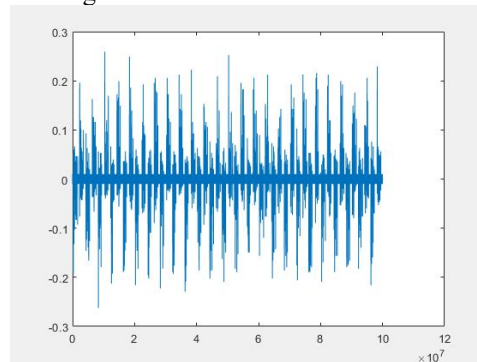


Figure 3. The raw PD signal

Use wavelet transform to perform a wavelet decomposition on the difference of the received signal to obtain low-frequency and high-frequency components, and then threshold the high-frequency components to reconstruct the processed low-frequency and high-frequency components. The reconstructed signal eliminates the influence of fluctuation and noise, and keeps the original basic waveform of the signal unchanged. Then the corresponding fragments of the different parts of signals are spliced into one-dimensional signal.

After the signal pre-processing, a deep learning network is built for PD recognition. Using supervised learning methods, using convolutional neural networks to extract signal features.

First, the input data needs to be enhanced. In order to reduce the sampling workload and the over-fitting of the CNN model, the training set data can be appropriately expanded. Data enhancement is a technique often used in the field of image deep learning, including colour, scale, translation, flipping, and so on. The signal enhancement method in this system uses the sliding of the signal interception window. The clipping window is appropriately slid on the time axis, and the information of different time windows can be used as input data to achieve the purpose of data enhancement. The data enhancement method can also avoid the influence of some irrelevant features on classification to a certain extent.

The first layer is convolutional layer, and the convolutional kernel $kernel^{(c1)}$ is a 1×3 array, step length is 1, channels is 64. The bias $b^{(c1)}$ is a 1×64 array. The input is pre-processed and enhanced signal, then the output after passing through the first convolutional layer is:

$$f(conv(\Delta s, kernel^{(c1)}) + b^{(c1)}), \quad (1)$$

where $conv()$ is the convolutional calculation, and $f()$ is activation function. The dimension of output and input is consistent using padding, and the purpose of the activation function is to perform a nonlinear combination in a linear network to generate a nonlinear decision boundary. The activation function used is the ReLu (Rectified Linear Units) function:

$$f(x) = \text{ReLu}(x) = \max(0, x). \quad (2)$$

The second layer is pooling layer, using the maximum pooling method. The pooling window size is 1×3 , and the step size is 2, that is, the maximum value within the window size range is selected as the input feature of the latter layer. By partitioning and sampling the data, the pooling layer can reduce the large matrix to a small matrix while maintaining the basic characteristics, to reduce the amount of calculation and prevent over-fitting.

The third layer is the local response normalization layer (local response normalization), using the concept of lateral inhibition in neurobiology. This introduces a competition mechanism into local neurons, that is, activated neurons will inhibit neighbouring neurons. The result is the large value larger, and has benefits to improve the generalization ability of the model. The calculation method of local response normalization is:

$$LRN(x) = x / (k + \alpha \sum_{j=\max(0, i-n/2)}^{\min(N-1, i+n/2)} (x)^2)^\beta, \quad (3)$$

where x is input and N is amount of kernels, n , k , α and β are hyperparameters, n is the number of neighbouring kernel mappings, and k is the offset. In this structure, n is set to 4, k is set to 1, α is set to 0.001, and β is set to 0.75. The output at this layer is expressed as:

$$out^{(c1)} = LRN(pooling(\text{ReLu}(conv(\Delta s, kernel^{(c1)}) + b^{(c1)}))), \quad (4)$$

Then the next connect layers is the second set of convolutional layer, pooling layer and local normalization layer, the output can be expressed as:

$$out^{(c2)} = LRN(pooling(\text{ReLu}(conv(out^{(c1)}, kernel^{(c1)}) + b^{(c1)}))), \quad (5)$$

Through these convolutional layers and related layers, the features in the original data can be extracted. The two fully connected layers are then connected to achieve a linear transformation from one feature space to another feature space. By weighted combination of the previously extracted features, the hidden features extracted from the original data are mapped to the sample label space. The number of hidden units in the first fully connected layer is 384, and the number of hidden units in the second fully connected layer is 192. The output of the final network structure is expressed as:

$$out^{(l4)} = \text{ReLu}(W^{(l4)} \text{ReLu}(W^{(l3)} out^{(c2)} + b^{(l3)}) + b^{(l4)}) \quad (6)$$

where $W^{(l3)}$ and $b^{(l3)}$ is the weight and bias of first fully connected layers, $W^{(l4)}$ and $b^{(l4)}$ is the weight and bias of second fully connected layers. The last part of the network structure is the fully connected layer and the soft-max layer. The fully connected layer further restricts the features to the order of the number of classifications, that is, the number of hidden units is the number of classifications C , and the output is expressed as:

$$out^{(sm5)} = W^{(l3)} out^{(l4)} + b^{(l3)} \quad (7)$$

The soft-max layer uses a multi-class logistic regression method, which can convert the input to a probability of 0-1, that is, to achieve the conversion of features and probabilities. The calculation formula is expressed as:

$$out^{final} = \text{softmax}(out^{(sm5)}) = \begin{bmatrix} p(\text{label} = 1 | out^{(sm5)}) \\ p(\text{label} = 2 | out^{(sm5)}) \\ \vdots \\ p(\text{label} = C | out^{(sm5)}) \end{bmatrix} \\ = \frac{1}{\sum_{c=1}^C e^{out_c^{(sm5)}}} \begin{bmatrix} e^{out_1^{(sm5)}} \\ e^{out_2^{(sm5)}} \\ \vdots \\ e^{out_C^{(sm5)}} \end{bmatrix} \quad (8)$$

where out^{final} is a $C \times 1$ vector, and indicates the category the input sample belongs to. The index with the largest value in the vector is the predicted label of the sample. The fully connected weights and bias parameters are trained to make the output of the fully connected layer have the highest probability of corresponding to the real label after passing through the soft-max layer.

After the construction of the entire network, a cost function needs to be defined to measure the difference between the predicted classification and the true classification. Cross entropy is used as a method to measure this difference, which can be expressed as:

$$cross_entropy = - \sum_c out^{final} \log(\text{label}) \quad (9)$$

where the label is the one-hot code corresponding to the real label value. For a batch, the above formula needs to be summed or averaged. In order to increase the generalization ability of the network and prevent over-fitting, on the basis of formula (9), an L2 regularization term is added, and the cost function is finally expressed as:

$$\text{cost_function} = \text{cross_entropy} + \lambda \sum_j W_j^2 \quad (10)$$

Where W_j is the weight parameter of the neural network and is the coefficient. The L2 norm regularization can limit the parameters to be too large or too small. By minimizing the entire cost function, model weight parameters and bias parameters can be learned.

The stochastic gradient descent method is adopted as training method. Each iteration of the stochastic gradient descent method uses a set of samples, and then updates the parameters according to the gradient. The batch size of each training set here is 128 samples. In addition, the learning rate is an important hyperparameter. Generally, it is hoped that the learning rate in the early stage of training will be larger, so that it can converge faster, and the learning rate will be reduced in the later stage of training, which can easily achieve the best advantage. Therefore, the learning rate decays with the number of iterations is used in the system. Rate. The initial learning rate is 0.1, and it decays to 10% every 1000 iterations. Before learning, the network parameters need to be initialized. Initialization has a certain impact on the convergence speed of the model and the final performance. In order to minimize the disappearance of the gradient or the explosion of the gradient, the truncated normal initialization is used in the data initialization. This initialization can be randomly selected from the normal distribution of the specified mean and variance, and reselect if the difference between the value and the mean is outside the range of two times the variance.

Through the above processing and settings, the entire network can be trained to a suitable state on the training set, and feature extraction and classification recognition can be realized. There are four types of partial discharge to recognize, oil gap defect, floating defect, tip defect and surface defect. For each type, 200 samples are sampled to form the training set, and 50 samples are sampled to form the test set. Therefore, the total training set has 800 samples and the test set has 200 samples. Network training is carried out through the training set, and the results are verified on the test set. The accuracy of item recognition reaches 91.8%. Compared with BP neural network[19] and OS-ELM algorithms, which the accuracy are 87.8% and 91.4% respectively, our methods get higher accuracy.

C. Conclusion

A partial discharge recognition method was introduced in this paper. Four kinds of PD signals were collected through a high-voltage partial discharge device and high-frequency current sensors. We use the wavelet transform to remove interference of the time-domain signal, and CNN was used to extract features and recognition. The accuracy of recognition is 91.8%.

III. REFERENCES

- [1] Boczar T, "Identification of a specific type of PD from acoustic emission frequency spectra," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 8, pp. 598-606, 2001.
- [2] Hikita M, Ohtsuka S, Matsumoto S. "Recent trend of the partial discharge measurement technique using UHF electromagnetic wave detection method," *IEEE Transaction on Electrical and Electronic Engineering*, vol. 2, pp. 432-435, 2007.
- [3] Xinyan Feng, Qingchen Meng, Kai Li, Fengchao Wang, "Discussion on Fault Location in the Partial Discharge Detection of GIS with the UHF Method" *Shandong electric power*, vol. 43, pp. 72-74, 2016
- [4] GAO Hong-wu, LI Jian, CHENG Chang-kui, JIN Zhuo-rui, WANG You-yuan, "Impact of Oil-Paper Barrier on Detected UHF Signal of Partial Discharge in Transformer" *high Voltage Apparatus*, vol. 45, pp: 118-121, 2009
- [5] M. Tsuchie, M. Kozako, M. Hikita and E. Sasaki, "Modeling of early stage partial discharge and overheating degradation of paper-oil insulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 3, pp. 1342-1349, June 2014.
- [6] S. Lu, H. Chai, A. Sahoo and B. T. Phung, "Condition Monitoring Based on Partial Discharge Diagnostics Using Machine Learning Methods: A Comprehensive State-of-the-Art Review," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, no. 6, pp. 1861-1888, December 2020.
- [7] K. Firuzi, M. Vakilian, B. T. Phung, and T. R. Blackburn, "Partial Discharges Pattern Recognition of Transformer Defect Model by LBP & HOG Features," *IEEE Trans. Power Deliv.*, vol. 34, no. 2, pp. 542-550, Apr. 2019.
- [8] W. J. K. Raymond, H. A. Illias, and A. H. A. Bakar, "High noise tolerance feature extraction for partial discharge classification in XLPE cable joints," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 1, pp. 66-74, Feb. 2017.
- [9] Ying Lin, Yingtao Sun, Xinyan Feng, Xi Sun, YuFeng Chen, "Signal Recovery of GIS Partial Discharges Graph based on Imae Pocessing" *Shandong electric power*, vol. 42, pp. 19-22.
- [10] L. S. Lumba, U. Khayam, and R. Maulana, "Design of pattern recognition application of partial discharge signals using artificial neural networks," *Int. Conf. Electr. Eng. Informatics (ICEEI)*, pp. 239-243, 2019.
- [11] T. R. Sukma et al, "Classification of partial discharge sources using waveform parameters and phase-resolved partial discharge pattern as input for the artificial neural network," *Int. Conf. Cond. Monit. Diagnosis (CMD)*, 2018.
- [12] A. A. Soltani and A. El-Hag, "Denoising of radio frequency partial discharge signals using artificial neural network," *Energies*, vol. 12, no. 18, p. 3485, Sep. 2019.
- [13] E. T. Iorkyase et al, "Improving RF-based partial discharge localization via machine learning ensemble method," *IEEE Trans. Power Deliv.*, vol. 34, no. 4, pp. 1478-1489, Aug. 2019.
- [14] Y. Wang et al, "Separating multi-source partial discharge signals using linear prediction analysis and isolation forest algorithm," *IEEE Trans. Instrum. Meas.*, pp. 1-1, Jul. 2019.
- [15] X. Peng et al, "Random forest based optimal feature selection for partial discharge pattern recognition in HV cables," *IEEE Trans. Power Deliv.*, vol. 34, no. 4, pp. 1715-1724, Aug. 2019.
- [16] L. Duan et al, "Identification of partial discharge defects based on deep learning method," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1557-1568, Aug. 2019.
- [17] B. Adam and S. Tenbohlen, "Classification of multiple PD sources by signal features and LSTM networks," *IEEE Int. Conf. High Volt. Eng. Appl. (ICHVE)*, 2019.
- [18] H. Song, J. Dai, G. Sheng, and X. Jiang, "GIS partial discharge pattern recognition via deep convolutional neural network under complex data source," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 2, pp. 678-685, Apr. 2018.
- [19] Xingshun Miao, Liang Zou, Zhiping Song, Xiaolong Wang, Junbo Wang, Bin Liu, "Study on PD Pattern Recognition of Power Transformer Considering External Corona Interference Signal" *2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 1-4, 2020.
- [20] J. Fubao, Z. Shanjun, Z. Yuanxiang and L. Bin, "Pattern Recognition of Development Stage of Creepage Discharge of Oil-Paper Insulation under AC-DC Combined Voltage based on OS-ELM," *2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 15-17, 2020.



Meng Liu was born in 1989 in Shandong province, China, and got the Ph.D in information and communication engineering 2019. Now I am a engineer in state grid and focus on partial discharge recognition research. Partial discharge, deep learning, and signal processing is the main interest of research.



Wenjie Zheng was born in 1989, majored in information and communication engineering, he is now an engineer of state grid shandong electric research institute. He has focus on the Visualization of transmission channels for many years.



Ying Lin was born in 1986 in shandong province, china, and got her Ph.D in zhejiang university in information and communication engineering, now she is a engineer in stae grid and focus on image recognition research.



Demeng Bai was born in 1989, majored in information and communication engineering, he is now an engineer of state grid shandong electric research institute. He has focus on the Equipment status evaluation for many years.



Qingdong Zhu was born in 1987, received a PhD in 2014, mainly work on the state monitoring evaluation and diagnosis of power equipment.



Jiafeng Qin was born in 1991, majored in information and communication engineering, he is now an engineer of state grid shandong electric research institute. He has focus on the electric management for many years.



Junxin Wang was born in 1986, majored in Electrical engineering, he is a employee in state grid in shandong province, and focus on the implementation of new electrical technology.



Jian Wang was born in 1983, majored in electrical engineering, she is now an engineer of state grid shandong electric research institute, she has focus on the Transformer insulation for many years.



Yi Yang, was born in 1986, majored in information and communication engineering, she is now an engineer of state grid shandong electric power research institute, she has focus on Equipment status evaluation for many years, is very experienced in the field of intelligent transportation inspection.



Technology Center.

Fengda Zhang was born in Jinan City, Shandong Province. I am a postgraduate student majoring in Computer Science and Technology at Shandong University. My main research direction is the design and application of object evolution and relevance model based on spatio-temporal framework. I work in the State Grid Shandong Electric Power Research Institute, and my job is an intelligent operation and inspection position in the Power Transmission and Transformation



Chao Gu, was born in 1971, majored in electrical engineering, he is now the director of Power Transmission Technology Center, he has focus on the study of Transformer insulation for many years.



Zhuangzhuang Li was born in 1993 and graduated from Shandong University and majored in Electronic and communication Engineering with a master's degree. Now I am working in State Grid Shandong Electric Power Research Institute, and my main research direction is power big data and application of artificial intelligence technology.



Wenbing Zhu was born in 1980, major in electrical engineering, he is now the deputy director of power transmission technology center, and focus on partial discharge recognition research. Partial discharge, deep learning, and transformer insulation is the main interest of research.