Transformer Partial Discharge Pattern Recognition Based on Random Forest

Shijun Wang1,*, Chang Ping1 and Guobin Xue1
1The Institute of Economic Technology for Grid Power Company, Gansu, China

*Corresponding author e-mail: songli0212@163.com

Abstract. As the deficiencies of the classifier algorithm commonly used in the partial discharge pattern recognition, this paper studied the application of RF in transformer partial discharge pattern recognition. Firstly, extracting the statistical characteristics from partial discharge test data to establish the discharge samples; Then, using ten folds method to judge the algorithm performance and compare the recognition accuracy of BP neural network, support vector machine, KNN, CART and RF algorithm. The results showed that the accuracy of the discharging pattern classifier basic on RF algorithm was the highest. The main differences between the different discharge modes were discussed by CART algorithm which composes RF.

1. Introduction
The power transformer bears the task of energy transmission and transformation in the power system, and is the pivotal equipment of the power system. Its operating status is related to the safe operation of the power system. Partial discharge can reflect the deterioration of transformer insulation. The detection of partial discharge can evaluate the internal insulation of the transformer, effectively identify the internal discharge type of the transformer, and has guiding significance for its maintenance and repair work [1,2]. At present, in the identification of transformer discharge mode, the commonly used methods are neural network, SVM, KNN and so on.

The neural network adjusts the input and output weights and offsets in the network through different kinds of samples to determine the mapping relationship between feature quantities and categories. Currently, BP neural network [3] and probabilistic neural network [4] are more commonly used. Support Vector Machine (SVM) is based on the theory of statistical theory and structural risk minimization, and theoretically can achieve a certain generalization ability. Because this method has the advantages of overcoming dimensionality disaster, over-fitting, small sample, etc., it is widely used in the field of PD mode recognition [5,6]. KNN is the closest to K homogeneous training samples in a feature space in a test sample, and this test sample falls into this category. Because its principle is simple, the application is good [7-9]. RF is a combined classifier model that consists primarily of a set of decision tree classifiers. Each decision tree can select once to determine the best sort result. RF can quickly process high-dimensional data, and has the advantages of no feature selection, no over-fitting, no noise sensitivity, and strong noise tolerance.

In this paper, the RF algorithm was used to identify the inter-turn discharge and oil wedge discharge. The training results of the basic classification CART algorithm that constitutes RF were discussed, and the main differences between different discharge modes were pointed out.
2. RF algorithm

2.1. CART Algorithm

The CART algorithm uses binary recursive segmentation technology to construct prediction rules and presents them in the form of binary trees. It has superior anti-noise performance and is easy to understand and use. The CART algorithm mainly contains two important ideas: first, recursively dividing the independent variable space; second, using the verification data for pruning.

1) Partition the independent variable space

All sample sets are used as root nodes, X1, X2, ..., Xp are used to represent explanatory variables, and Y is used to represent categorical variables. The segmentation variable is selected by the segmentation function and the segmentation threshold is determined. The purpose of dividing the nodes is to continuously reduce the impurity I of the child node, and the formula is as shown in formula (1).

\[
I = 1 - \sum_{k=1}^{K} P_k^2
\]  

(1)

In this formula, K is kind, Pk is the proportion of the observation point that belongs to the kth class. Repeat the above steps until the entire X space is divided into small rectangles that do not overlap each other.

2) Pruning with validation data

Step 1 constructed a tree with the lowest impurity level, but the tree constructed at this time is easily over-fitting. In order to avoid over-fitting, the tree generated from the training must be pruned with a validation data set, the purpose of which is to obtain a best pruning tree. CART uses the cost complexity J standard to pruning, and its expression is shown in equation (2).

\[
J = \text{Err}(T) + \alpha L(T)
\]  

(2)

where Err(T) is Err error rate of the validation data set, L(T) is the number of leaf nodes of tree T, and \( \alpha \) is the penalty cost for each node.

The pruning is stopped when the pruning sequence contains errors within one standard deviation of the smallest error tree, at which point the optimal tree is obtained. The minimum error rate E expression is as shown in equation (3).

\[
E = \sqrt{E_{\text{min}}(1-E_{\text{min}})/N}
\]  

(3)

where \( E_{\text{min}} \) is the error rate of the minimum error tree, and N is the number of validation data sets.

2.2. RF Algorithm

The RF repeatedly and randomly extracts N samples in a centralized way with a return from the original N training sample sets, and k times to obtain k sets of sample sets, and uses the decision tree algorithm to form RF according to the extracted sample sets. The classification error of RF depends on the classification ability of each tree and the correlation between trees and trees [10]. The larger the strength of each classification decision tree, the better the classification performance of RF; the larger the correlation between the tree and the tree, the worse the classification performance of RF. Therefore, increasing the classification strength of each decision tree and reducing relevancy between the tree can effectively decrease the total error rate of RF. A schematic diagram of the RF construction process is shown in Figure 1.
The generation process of RF is as follows:

1) Extracting $k$ training sample sets in the original sample set $n$ by the Bootstrap method, the number of samples in each training set is $n$;

2) Learning separately for $k$ training sets, generating $k$ decision tree models. In order to improve the difference between the decision trees, the extracted feature sets are randomly selected. Suppose there are $M$ input variables and randomly extract $F$ features, the segmentation features of each node of CART are randomly selected in $F$ feature sets for all features;

3) Combine the results of the $k$ CART decision trees to form the last result.

Figure 2 shows the classification boundaries for the two categories for the three CART algorithms. It can be seen that the three boundaries are relatively coarse for both categories. The RF algorithm divides the boundaries of the three CART integrations into two categories, and the final RF partitioning boundary is shown in Figure 3. It can be seen from Figure 3 that through integration, more precise division boundaries can be obtained, and the accuracy of classification can be improved.

It has been proved in the literature that as the number of CART increases, the generalization error of RF will converge to a fixed value, indicating that RF has good scalability for unknown samples and the ability to prevent over-fitting.

3. Partial discharge data extraction

The electric field between the electrodes inside the transformer is mostly a slightly uneven electric field and an uneven electric field. The slightly uneven electric field appears between the high and low voltage windings, the winding to the core or the outer casing, the high voltage lead to the outer casing, the electrostatic screen to the winding, etc.; the uneven electric field generally appears at the corner of the inner core of the iron core without electrical shielding, The structure of the lead-to-clamp member and the like, the corner of the lead to the tank wall, the upper and lower iron yokes of the electrostatic screen, the crimping of the winding end to the insulating platen, and the like. In this paper, the inter-turn model [11,12] was used to represent a slightly uneven electric field, and a spherical - plate
oil wedge model was used to represent the uneven electric field. The structure of the inter-turn model and oil wedge model is shown in figure 4.

![Inter-turn model and Oil wedge discharge model](image)

**Figure 4.** Inter-turn model Oil wedge model and structure

Ninety cases of diurnal model discharge samples and oil wedge model discharge samples were collected at different stages of discharge development. In this paper, four partial spectra were obtained by processing the partial discharge signals, which were: two-dimensional $\phi$-$n$ spectrum, two-dimensional $\phi$-$q_{\text{sum}}$ spectrum, two-dimensional $\phi$-$q_{\text{ave}}$ spectrum and two-dimensional $\phi$-$q_{\text{max}}$ spectrum. The discharge diagram of each type is shown in Table 1.

For the four spectral extractions of the two types of discharges, the positive and negative half-cycle skewness $S_k$, the steepness $K_n$, the local peak point $P_e$, the mean value $M_v$ of the discharge phase, and the standard deviation $\sigma$ were extracted. In order to characterize the spectral difference between the positive and negative half of the spectrum, the cross-correlation coefficient $CC$, the correction factor $QF$ and the modified cross-correlation coefficient $MCC$ [13] were extracted. The extracted data was used as a learning sample of the classifier, and each classifier was trained to form an identifier for the inter-turn discharge and the oil wedge discharge mode.

| Table 1. Discharge patterns of each discharge model |
|---------------------------------------------------|
| $\phi$-$n$ spectrum | $\phi$-$q_{\text{sum}}$ spectrum | $\phi$-$q_{\text{ave}}$ spectrum | $\phi$-$q_{\text{max}}$ spectrum |
| inter-turn discharge | ![inter-turn discharge spectrum](image) | ![inter-turn discharge spectrum](image) | ![inter-turn discharge spectrum](image) |
| oil wedge discharge | ![oil wedge discharge spectrum](image) | ![oil wedge discharge spectrum](image) | ![oil wedge discharge spectrum](image) |

4. Classification results

4.1. Classification results of RF

In this paper, the data set used Bootstrap to generate $k$ data subsets, each of which had the same capacity as the original sample. Each data subset randomly selected $F$ feature numbers for CART decision tree algorithm learning, and finally generated RF. In this paper, the performance of the classifier constructed by RF was tested by the ten-fold method. The discharge data was divided into 10 folds. Each fold data was used as the test set in turn, and the remaining data was used as the training set to learn the classifier. The final accuracy was the average of 10 accuracy. The accuracy of the number of different decision trees and the number of different random features $F$ are shown in Table 2.
It can be seen from Table 2 that the RF algorithm has the highest accuracy when \( F = 5 \). As \( F \) increases, the accuracy of the RF algorithm decreases, because the increase in \( F \) reduces the difference between decision trees. Compared with the influence of \( F \) on the accuracy of RF algorithm, \( k \) has little effect on the accuracy of RF algorithm, but the increase of \( k \) value will prolong the learning time and prediction complexity of RF algorithm. Therefore, when \( F = 5 \) and \( k = 100 \), the RF algorithm has the shortest learning time and the highest recognition accuracy.

In order to verify the effectiveness of RF algorithm in partial discharge pattern recognition, this paper used BP neural network, SVM, KNN, CART classification algorithm to identify partial discharge pattern. Among them, there were 52 nodes of BP neural network input layer, 1 layer of hidden layer, 3 nodes of output layer, corresponding to 3 patterns to be identified, and the number of hidden layer nodes was set to 9 according to reference [14]. The SVM used the RBF core and the polynomial kernel for training. The CART algorithm took 1/3 of the training set as a validation set for pruning, using all the features of the entire training set for training. The recognition accuracy of each algorithm is shown in Table 3. It can be seen from Table 3 that the RF algorithm has the highest recognition accuracy for each discharge model.

### Table 3. Accuracy of algorithms

| Classifier          | Accuracy of each discharge model/% | Total accuracy/% |
|---------------------|-----------------------------------|-----------------|
| BPNN                | 95.56 93.33                       | 94.445          |
| SVM (RBF kernel)    | 96.67 83.33                        | 90              |
| SVM (polynomial kernel) | 96.67 94.44                        | 95.555          |
| KNN                 | 94.44 90                           | 92.22           |
| CART                | 91.11 93.33                        | 92.22           |
| RF                  | 97.78 97.78                        | 97.78           |

### 4.2. Classification Results of CART

The discharge recognition model constructed by CART algorithm also has a good recognition effect. The CART algorithm is constructed using all the features of the entire training set, and the discriminant model is obtained as shown in Figure 5. According to Figure 5, the following rules can be obtained:

In the \( \varphi-n \) spectrum, when \( M_{V^+} < 55.959 \), the discharge mode is inter-turn discharge.

In the \( \varphi-n \) spectrum, when \( M_{V^+} \geq 55.959 \), the discharge mode is oil wedge discharge.
It can be known from the rules that the main difference between the inter-turn discharge and the oil wedge discharge is the positive half-cycle discharge phase average Mv+ of the $\varphi$-n spectrum, and the Mv+ of the $\varphi$-n spectrum of the oil wedge discharge is larger than the inter-turn discharge. This is because the discharge voltage of the oil wedge model is lower than the discharge voltage of the inter-turn model, so at the end of the positive half cycle, the oil wedge discharge is more likely to occur, so that the phase average value Mv+ of the corresponding positive half cycle is larger.

5. Conclusion
In this paper, the discharge signals of transformer inter-turn discharge and oil wedge discharge were extracted under laboratory conditions, and RF, BP neural network, SVM, KNN and CART algorithms were applied to partial discharge pattern recognition. The results showed that RF has higher recognition accuracy. The three discharge modes were discussed using the classification results of the CART algorithm. The results showed that the main difference between the inter-turn discharge and the oil wedge discharge was the positive half-cycle discharge phase average Mv+ of the $\varphi$-n spectrum, and the Mv+ of the $\varphi$-n spectrum of the oil wedge discharge was larger than the inter-turn discharge.

Acknowledgments
This article has been funded by the Science and Technology Project of the State Grid Gansu Electric Power Company (52272816001D) and expresses its sincere gratitude to them.

References
[1] J. Guo, G.N.Wu, X.Q. Zhang, et al. The Actuality and Perspective of Partial Discharge Detection Techniques[J]. Transactions of China Electrotechnical Society, 2005, 20(2): 29-35.
[2] M. D. Judd, G. P. Cleary, C. J. Bennoch. Applying UHF partial discharge detection to power transformers[J]. IEEE Power Engineering Review, 2002, 22(8): 57-59.
[3] P. Wang, L. H. Fu, G. Z. Hu. Recognition of Partial Discharge Distribution Patterns Based on Morlet Wavelet Transform and Neural Network[J]. High Voltage Apparatus, 2005, 41(5): 366-369.
[4] J. S. Yuan, H. K. Shang. Pattern recognition based on principal component analysis and probabilistic neural networks for partial discharge of power transformer[J]. Electric Power Automation Equipment, 2013, 33(6): 27-31.
[5] X. Ji, Y. Li, Z. Wang, et al. Partial discharge pattern recognition of XLPE cable connector based on support vector machine[C]. International Conference on Electrical and Control Engineering. IEEE, 2011:2422-2425.
[6] J. C. Chan, H. Ma, T. K. Saha. Partial discharge pattern recognition using multiscale feature extraction and support vector machine[C]. Power and Energy Society General Meeting. IEEE, 2013:1-5.
[7] F. Liu, G. Zhang, X. Yao, et al. Recognition of PD mode based on KNN algorithm for converter transformer[J]. Electric Power Automation Equipment, 2013, 33(5): 89-93.

[8] N. Pattanadech, P. Nimsanong. Effect of Training Methods on The Accuracy of PCA-KNN Partial Discharge Classification Model[C]. Tencon IEEE Region 10 Conference, 2014:1-5.

[9] K. Wang, J. Li, S. Zhang, et al. Time-Frequency Features Extraction and Classification of Partial Discharge UHF Signals[C]. International Conference on Information Science, 2014, 2:1231-1235.

[10] K. N. Fang, J. B. Wu, J. P. Zhu, et al. A Review of Technologies on Random Forests[J]. Statistics & Information Forum, 2011, (03): 32-38.

[11] F. Liu, G. Zhang, X. Yao, et al. Recognition of PD mode based on KNN algorithm for converter transformer[J]. Electric Power Automation Equipment, 2013, 33(5): 89-93.

[12] L. Breiman. Random Forests[J]. Machine Learning, 2001, 45(1):5-32.

[13] N. Yuan. Analysis and Fault Feature Extraction for Transformer Partial Discharge Signal[D]. Beijing, China: North China Electric Power University, 2013.

[14] H. Y. Shen, Z. X. Wang, C. Y. Gao, et al. Determining the number of BP neural network hidden layer units[J]. Journal of TIANJIN University of Technology, 2008, (05): 12-15.