Research on online diagnosis method of insulator contaminated discharge based on XGBoost algorithm

Ben Tan1, Yishi Yue1,*, Xingxin Chen2, Fuyong Huang1, Ziyi Cheng2, Yafeng Chao1, Zezhong Sun1, Yanhui Zou1
1. State Grid Hunan Electric Power Research Institute, Changsha, Hunan, China
2. China Three Gorges University, Yichang, Hubei, China
*a*yueyshndky@qq.com

Abstract. Insulators are one of the commonly used power equipment in overhead transmission lines. Long-term operation of insulators will cause end deterioration, pollution flashover and other problems due to the discharge phenomenon caused by surface contamination, after which equipment failure huge losses to the power system may occur. In order to accurately monitor the discharge status of insulators, an online diagnosis method for insulator contamination discharge based on XGBoost algorithm is proposed in this paper. The discharge status of insulators can be judged by analyzing the discharge acoustic signals of the insulators running under different pollution levels, so as to realize accurate monitoring of insulators. The results show that this method can effectively judge the discharge state of the insulator thus early warnings can be given the safe and stable operation of the power system can be ensured.

1. Introduction
Overhead transmission line insulators are exposed to strong electromagnetic fields for a long time, they will suffer from defects such as surface deterioration due to loss and aging, or long-term contamination will cause pollution flashovers, causing large-area line outages, seriously threatening the reliability and safety of the power system [1-2]. Contamination accumulates on the surface of the insulator, forming a moist polluted area in a humid environment, which leads to a decrease in the insulation performance of the insulator [3]. In this case, it is very easy to cause pollution flashover. With the higher voltage levels of modern power systems, the current research results on insulators are difficult to meet the operational requirements of actual lines. It is necessary to further study the pollution flashover monitoring methods of insulators, so as to monitor and early warning of pollution flashovers in time [4-6]. Existing studies have proposed methods based on mathematical morphology, principal component analysis, and random forest based on insulator discharge sound by monitoring the discharge sound of insulators, but related methods are mainly based on the analysis of data frequency domain characteristics, which are easily affected by the environment noise [7-9]. This paper proposes an on-line diagnosis method for the insulator discharge state. Through the study of the acoustic information of the operating insulator in different discharge states, the accurate judgment of the discharge state is achieved. The results show that the method can effectively monitor the discharge status of insulators under different pollution levels and find hidden equipment hazards in time.
2. Experiment

2.1 Test system
This test collects the discharge acoustic signals of wet polluted ceramic insulators at 10 different distances under the same pollution level and the discharge acoustic signals of dry polluted glass insulators at the same distance under 5 different pollution levels with different discharge states (corona discharge, surface discharge, flashover). The designed artificial pollution discharge test system is shown in Fig.1, which consists of sound collection device, capacitor voltage divider, test transformer, voltage regulator and insulators. In this experiment, the insulator string consists of two insulators. Fig.2(a) shows the test device at the testing site and Fig.2(b) is the flashover discharge phenomenon of wet polluted ceramic insulators.

![Diagram of test system](image1)

Fig. 1 Diagram of test system

![Test device at the testing site](image2)

(a)Test device at the testing site

![Flashover phenomena of polluted insulators](image3)

(b)Flashover phenomena of polluted insulators

Fig. 2 Photos of test device and phenomena of polluted insulators

2.2 Production of polluted insulators
The pollution degree of the sample is expressed by salt deposit density (SDD) and the non-soluble deposit density (NSDD). Combined with the actual pollution situation, the NSDD-SDD ratio in this test is 5:1 and the insulator strings are numbered in sequence and smeared with different degrees of SDD. The sodium chloride (chemically pure) and the weight of diatomaceous earth required for each sample are calculated according to SDD and the insulator surface area of the insulator. After drying, the sodium chloride is placed in a small ceramic bowl, stirred evenly and applied to the surface of the sample insulators. Using this kind of quantitative smearing of stained samples is simple and easy to do and does not require random inspection. The above-mentioned 2 pieces of smeared insulators are selected as a string of insulator to which power frequency voltage is applied. The sound is record and collected under different discharge states using sound collection devices. The salt deposit and non-soluble deposit required to be painted on the upper and lower surfaces of the insulators with different SDD is shown in
Table 1. The contaminated glass insulators are shown in Fig. 3. Typical acoustic signal of corona discharge of the polluted insulator at 3 meters is shown in Fig. 4.

Table 1 Deposit painted on the upper and lower surfaces of the insulators with different SDD

| No. of samples | SDD (mg/cm²) | Salt deposition (mg) | Non soluble deposit density (mg) | Salt deposition (mg) | Non soluble deposit density (mg) |
|----------------|--------------|----------------------|---------------------------------|----------------------|---------------------------------|
| 1              | 0.01         | 6.5                  | 32.4                            | 8.6                  | 43.1                            |
| 2              | 0.05         | 32.4                 | 162                             | 43.1                 | 215.5                           |
| 3              | 0.10         | 64.8                 | 324                             | 86.2                 | 431                             |
| 4              | 0.20         | 129.6                | 648                             | 129.3                | 646.5                           |
| 5              | 0            | 0                    | 0                               | 0                    | 0                               |

Fig. 3 Contaminated glass insulators

Fig. 4 The corona discharge sound at 3m

2.3 Introduction to XGBoost algorithm

The XGBoost (Extreme Gradient Boosting) algorithm is a very effective algorithm in machine learning. Its basic idea is to perform a second-order Taylor expansion on the objective function, and then use the second-order derivative information of the function to train the tree model. In the training process, the complexity of the tree model is added as a regular term to the optimization goal, which makes the generalization ability of the obtained model higher. When the XGBoost algorithm model has K trees, its objective function O is expressed as below.

\[
O = \sum_{i=1}^{n} (l(y_i, \hat{y}_i)) + \sum_{k=1}^{K} \Omega(f_k)
\]  

(1)

Where \( y_i \) is the true value of the i-th target; \( \hat{y}_i \) is the predicted value of the i-th target; \( l(y_i, \hat{y}_i) \) describes the difference between \( y_i \) and \( \hat{y}_i \); \( n \) is the number of samples; \( \Omega(f_k) \) is the complexity of the tree model where the k-th sample feature parameter is located; \( K \) is the total number of sample feature parameters. By optimizing in the gradient direction, the model residual of the learner is continuously reduced and a new tree model is obtained. Therefore, when the objective function is approximated by the second-order Taylor expansion, the training objective function \( O \) in the \( t \)-th iterative step calculation is transformed into the expression below.

\[
O^{(t)} \approx \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + 0.5 h_i f_i^2(x_i) + \Omega(f_i)
\]  

(2)

The first and second derivative of the prediction error with respect to the current model are

\[
g_i = \partial_{\hat{y}_i} l(y_i, \hat{y}_i^{(t-1)})
\]

(3)

\[
h_i = \partial_{\hat{y}_i}^2 l(y_i, \hat{y}_i^{(t-1)})
\]

(4)

Since at the \( t \)-th iteration, the \( t-1 \)th model residual is known, the constant term is removed and the equation (3) is expanded, and the objective function is written in the form of accumulating leaf nodes:
\[O(i) \approx \sum_{j=1}^{T} \left[ Gw_j + 0.5(H + \lambda)\omega_j^2 \right] + \gamma T \]  

(5)

Where \( i \) is the set of samples on each leaf; \( J_j = \{ i | q(x_i) = j \} \); \( q(x_j) \) is the tree structure function; \( T \) is the number of leaf nodes in the split tree; \( \lambda \) and \( \gamma \) are the weighting factors used to control the proportion of corresponding part. According to the specific application scenario, the corresponding error function is constructed to train the specific tree model. Therefore, it is completely feasible and necessary to introduce the XGBoost algorithm into fault diagnosis and build an adaptive fault split tree. Similar to the information gain and Gini index in the decision tree, the XGBoost algorithm calculates the gain \( \Theta \) of the selected parameter every time when adding a segmentation to a leaf:

\[ \Theta = \frac{1}{2} \left[ \frac{G^2_i}{H_L + \lambda} + \frac{G^2_R}{H_R + \lambda} - \frac{(G_i + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \]  

(6)

where \( L \) and \( R \) represent the left subtree and the right subtree respectively; \( G^2_i /( H_L + \lambda) \) is the information score of the left subtree; \( G^2_R /( H_R + \lambda) \) is the information score of the right subtree; \( (G_i + G_R)^2/(H_L + H_R + \lambda) \) is the current undivided information score. The gain \( \Theta \) is essentially to statistically purify the important information contained in the data to reduce the uncertainty of the information before and after the tree leaves are split. When the entire tree is split, the amount of information of the leaf nodes that are finally formed is the largest, and the importance of information is also the highest.

3. Test Results and Discussions

The acoustic signals of wet polluted ceramic insulators at different positions are collected. According to the acoustic signal data, the minimum, maximum, standard deviation and distance are taken as the characteristic parameters of the signal and the different stages of the discharge (corona discharge, surface pollution discharge, flashover) as the target, some of the sample parameters used to train the model are shown in Table 2. The acoustic signals of wet polluted ceramic insulators at different distances include 30 sets of data, and the distance between the sound collector and the insulator ranges from 2 meters to 11 meters. The samples collected at the same distance include the acoustic signals of three different discharge states of insulators. The minimum, maximum, and standard deviation of the acoustic signals in the three different discharge states have an obvious trend of gradual increase and the maximum value and standard deviation in the flashover discharge stage have obvious sudden changes. The first 22 sets of data are used as the training set to train the model, and the last 8 sets of data are used as the test set.

The acoustic signals of dry contaminated glass insulators at different levels of contamination are also collected. According to the acoustic signal data, the minimum, maximum, standard deviation and contamination density are taken as the characteristic parameters of the signal, the different stages of the discharge as the target, some of the sample parameters used to train the model are shown in Table 3. The sample No.5 is related to the clean insulator. The acoustic signals of insulator discharge with the same SDD include three different discharge states. The minimum, maximum and standard deviations of the acoustic signal of different discharge states have the same tendency as the wet polluted ceramic insulators, and there are obvious sudden changes in the minimum and standard deviations of the signals during the flashover discharge stage. The first 9 sets of data are used as the training set to train the model, and the last 6 sets of data are used as the test set.

The XGBoost model is built and trained using Python 3.7. In order to illustrate the accuracy of XGBoost's online diagnosis of the discharge phase, it is compared with Random Forests (RF) and Gradient Boosting Decision Tree (GBDT) models. Table 4 shows the diagnosis of discharge status of wet polluted ceramic insulators using different models, where 0 stands for corona discharge, 1 stands for surface discharge and 2 stands for flashover. The accuracy of the XGBoost-based online diagnosis method for wet polluted ceramic insulators at different positions reach to 87.5% and it is higher than
both of the accuracy of the random forest and the accuracy of the gradient boosting decision tree, which are 62.5% and 75.0% separately. The accuracy of the XGBoost-based online diagnosis method for dry polluted glass insulators at different pollution levels reach to 83.3% while the accuracy of the random forest is 83.5% and the accuracy of the gradient boosting decision tree is 66.67%. In summary, XGBoost has a high accuracy rate for on-line diagnosis of insulator discharge and hidden perils in insulator operation can be efficiently found using this method.

| No. | Min(Pa)   | Max(Pa)  | Standard deviations | Distance(m) | Stage of discharging | Data sets       |
|-----|-----------|----------|---------------------|-------------|----------------------|----------------|
| 1   | -0.30455  | 0.24757  | 0.05609             | 10          | Flashover            | Training Set   |
| 2   | -0.09469  | 0.0856   | 0.0269              | 10          | Surface discharge    |                |
| 3   | -0.09954  | 0.10311  | 0.03116             | 10          | Corona discharge     |                |
| 4   | -0.0959   | 0.09688  | 0.02678             | 11          | Surface discharge    |                |
| 5   | -0.2004   | 0.14071  | 0.03698             | 11          | Corona discharge     |                |
| 6   | -2.04833  | 1.96725  | 0.03011             | 2           | Flashover            |                |
| 7   | -0.08645  | 0.08281  | 0.01836             | 2           | Surface discharge    |                |
| 8   | -0.07653  | 0.08177  | 0.02015             | 2           | Corona discharge     |                |
| 9   | -2.43661  | 1.68577  | 0.03043             | 3           | Flashover            |                |
| 10  | -0.09734  | 0.08402  | 0.02523             | 3           | Corona discharge     |                |
| 11  | -1.28774  | 1.5653   | 0.03773             | 4           | Flashover            | Training Set   |
| 12  | -0.17952  | 0.12549  | 0.02791             | 4           | Surface discharge    |                |
| 13  | -0.11228  | 0.09198  | 0.0262              | 5           | Flashover            |                |
| 14  | -0.10761  | 0.10248  | 0.02803             | 5           | Surface discharge    |                |
| 15  | -0.14042  | 0.18598  | 0.02905             | 6           | Flashover            |                |
| 16  | -0.12947  | 0.13546  | 0.03866             | 7           | Surface discharge    |                |
| 17  | -0.16776  | 0.12537  | 0.03636             | 7           | Corona discharge     |                |
| 18  | -0.10046  | 0.09296  | 0.02322             | 8           | Flashover            |                |
| 19  | -0.13725  | 0.14504  | 0.03783             | 8           | Surface discharge    |                |
| 20  | -0.12399  | 0.1369   | 0.0401              | 9           | Flashover            |                |
| 21  | -0.19717  | 0.17456  | 0.04874             | 9           | Surface discharge    |                |
| 22  | -0.37346  | 0.45766  | 0.07764             | 9           | Corona discharge     |                |
| 23  | -0.08916  | 0.07593  | 0.02575             | 5           | Corona discharge     |                |
| 24  | -0.16388  | 0.2117   | 0.04297             | 7           | Flashover            |                |
| 25  | -0.1361   | 0.13627  | 0.0374              | 3           | Surface discharge    |                |
| 26  | -0.0865   | 0.09804  | 0.02559             | 6           | Surface discharge    |                |
| 27  | -0.08858  | 0.08408  | 0.02323             | 4           | Corona discharge     |                |
| 28  | -0.14371  | 0.15196  | 0.03112             | 6           | Corona discharge     |                |
| 29  | -0.13823  | 0.14925  | 0.0382              | 8           | Corona discharge     |                |
| 30  | -0.70725  | 1.21069  | 0.05022             | 11          | Flashover            |                |
Table 3 Signal characteristics of dry polluted glass insulators

| No. | Min(Pa)     | Max(Pa)     | Standard deviations | SDD (mg/cm²) | Stage of discharging | Data sets   |
|-----|-------------|-------------|---------------------|--------------|-----------------------|-------------|
| 1   | -1.51922    | 1.20106     | 0.08                | 957          | Surface discharge     | Training Set|
| 2   | -0.59779    | 0.41504     | 0.02                | 307          | Surface discharge     | Training Set|
| 3   | -0.0771     | 0.06886     | 0.01                | 0.05         | Corona discharge      | Training Set|
| 4   | -1.42009    | 1.23036     | 0.11                | 0.10         | Flashover             | Training Set|
| 5   | -0.08725    | 0.10588     | 0.02487             | 0.1          | Corona discharge      | Training Set|
| 6   | -0.49952    | 0.49531     | 0.03077             | 0.2          | Surface discharge     | Training Set|
| 7   | -1.43808    | 1.47741     | 0.10808             | 0.2          | Flashover             | Testing Set |
| 8   | -1.58075    | 1.24599     | 0.10431             | 0            | Flashover             | Testing Set |
| 9   | -0.13535    | 0.12745     | 0.03273             | 0            | Corona discharge      | Testing Set |
| 10  | -0.702      | 0.81745     | 0.02845             | 0            | Surface discharge     | Testing Set |
| 11  | -1.36888    | 1.47349     | 0.08948             | 0.05         | Flashover             | Testing Set |
| 12  | -1.01352    | 0.71763     | 0.0372              | 0.01         | Flashover             | Testing Set |
| 13  | -0.86849    | 0.61019     | 0.03073             | 0.1          | Surface discharge     | Testing Set |
| 14  | -0.12404    | 0.12185     | 0.03313             | 0.01         | Corona discharge      | Testing Set |
| 15  | -0.05438    | 0.06061     | 0.01417             | 0.2          | Corona discharge      | Testing Set |

Table 4 Diagnosis of the discharge status of insulators

| No. | Wet polluted ceramic insulators | Dry polluted glass insulators |
|-----|--------------------------------|------------------------------|
|     | 23 24 25 26 27 28 29 30 10 11 12 13 14 15 | 10 11 12 13 14 15 |
| True value | 0 2 1 1 0 0 0 2 1 2 2 1 0 0 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| XGBoost  | 0 2 1 1 0 0 1 2 1 1 2 1 0 0 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| RF      | 1 2 1 1 0 0 1 1 1 1 1 2 1 0 1 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| GBDT    | 0 2 1 2 0 0 1 2 1 1 1 1 0 0 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |

4. Conclusion

Based on the results and discussions presented above, the conclusions are obtained as below:

(1) An online diagnosis method of insulator discharge state based on XGBoost algorithm is proposed in this paper. By taking the minimum, maximum, standard deviation, distance and contamination density of the discharge acoustic signal of the insulator in operation as characteristic parameters as the input of the XGBoost algorithm, the discharge state of the insulator is predicted.

(2) The accuracy rate of the online diagnosis method for wet polluted ceramic insulator discharge reaches 87.5%, and the accuracy rate of the online diagnosis method for dry polluted glass insulator discharge reaches 83.3%, which is higher than RF and GBDT.

(3) The contamination status of insulators can be effectively monitored by the method so that hidden dangers in the operation of insulators can be discovered in time and the safe, reliable and stable operation of the power system can be ensured.

Acknowledgment

This work is supported by Science and Technology Project of State Grid Hunan Electric Power Company (5216A520000T). We sincerely express our thankfulness here.

References

[1] Arshad, Nekahi A, McMeeking S G, et al. Effect of pollution severity and dry band location on the flashover characteristics of silicone rubber surface[J]. Electrical Engineering, 2017, 99(3):1053-1063.

[2] Ma H, Guo Y, Zhang X, et al. Moisture content detection of insulator contamination based of hyperspectral technology[J]. High Voltage Engineering, 2020, 46(4):1396-1404.
[3] Liu Y, Jiang X, Zhang H, et al. Influence of Insoluble Contaminants on the Pollution Flashover of Insulators [J]. Proceedings of the Csee, 2017(11): 3331-3338.

[4] Wu J, Fu H, Zheng K, et al. Influence of Non-uniform Pollution on AC Flashover Performance of LXY4-160 Suspension Glass Insulator String [J]. Transactions of China Electrotechnical Society, 2020(17): 13-20.

[5] Pei C, Research of external insulation strength of contaminated-insulator monitoring method based on acoustic emission [D]. Wuhan University, 2009.

[6] Li Z, Study for On-line Monitoring Method and Device of Polluted Insulator Electrical Insulation Strength Based on Acoustic Emission [D]. Wuhan University, 2013.

[7] GB/T 4585-2004, Artificial pollution test of high voltage insulator for AC system [S]. 2004.

[8] A X M, A J S, B D W, et al. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning [J]. Electronic Commerce Research and Applications, 2018, 31: 24-39.

[9] Zhu K, Ma H, Wang J, et al. Optimization Research on Abnormal Diagnosis of Transformer Voiceprint Recognition based on Improved Wasserstein GAN [J]. Journal of Physics Conference Series, 2021, 1746:012067.

[10] Sun J, Zhang H, Li Q, et al. Contamination degree prediction of insulator surface based on exploratory factor analysis-least square support vector machine combined model [J]. High Voltage, 2020.

[11] Chen T, Tong H, Benesty M. xgboost: Extreme Gradient Boosting [J]. 2016.

[12] Mitchell R, Frank E. Accelerating the XGBoost algorithm using GPU computing [J]. PeerJ Computer Science, 2017, 3(1):e127.