

# Research on Insulator Leakage Current Data Analysis Based on DBClustering

Baoyi Wang<sup>a</sup>, Chenjing Lv<sup>b</sup>, Shaomin Zhang<sup>c</sup>

School of Control and Computer Engineering, North China Electric Power University,  
Baoding,071003,China

<sup>a</sup>email:wangbaoyiqj@126.com,<sup>b</sup>email:lvchenjing@163.com,<sup>c</sup>email:zhangshaomin@126.com

**Key words:** insulator leakage current, feature extraction, density clustering

**Abstract.** In order to analyze the operating condition of the leakage current of the insulator effectively, a cluster structure clustering algorithm based on density named DBClustering is designed. In this paper, four characteristic values of leakage current are extracted from the two aspects of leakage current in time domain and frequency domain, then the leakage current data sample set is formed, after that DBClustering algorithm was used to cluster and analyze this set. By extracting and analyzing the leakage current data collected in the laboratory, corresponding experiments show that the algorithm is superior to the traditional density clustering algorithm in terms of time and space consumption, and it is helpful to analyze the state of leakage current assessment and flashover studies.

## Introduction

Transmission line insulators due to changes in the external environment and long-term operation is easy to malfunction, then it followed by the frequent occurrence of pollution flashover accidents and the occurrence of power failure [1]. These accident because of insulator flashover caused great economic losses. Leakage current is the physical quantity which can reflect the operating state of the insulator [2]. The analysis of the leakage current can help to understand the operation status of the insulator, and effectively carry out the flashover warning.

At present, the research on leakage current of insulators mainly focuses on extracting and analyzing the characteristics of leakage current in time-domain and frequency-domain and the relationship between the characteristic value and flashover, but the literatures on leakage current are less. In paper[3], the fast Fourier transform is used to study the spectrum characteristics of leakage current at different pollution degrees, and four characteristic quantities of leakage current in the safety zone are extracted, and the fitting relationship with the pollution degree is analyzed. Paper[4] discusses the relationship between the contamination degree and the fractal dimension of the insulator, and proved that the law of the fractal dimension of the leakage current can effectively predict the development trend of contamination discharge and the occurrence of pollution flashover. In paper [5],by using HHT and wavelet analysis, the leakage current in time-frequency is analyzed, the fundamental and high-order harmonic electrical characteristics of leakage current in the danger zone is analyzed ; in paper[6] two characteristic value of leakage current is extracted by using fractal theory and the data is clustered by data mining algorithm, but the extracted by this paper is not enough to fully represent the characteristic of leakage current. The basic DBSCAN algorithm is used to cluster the data, and the efficiency of the algorithm is needs to be improved

In this paper, a cluster structure clustering algorithm based on density named DBClustering is designed. Four characteristic values of leakage current are extracted from the two aspects of leakage current in time domain and frequency domain, then the leakage current data sample set is formed. Clustering algorithm was used to cluster and analyze this set, aims at analyzing the state of leakage current Assessment.

## Feature Extraction of Leakage Current Signal

Signal feature extraction is the pre-processing stage of signal analysis, which is the dimensionality reduction of signal data. We extract the characteristic values of leakage current from both time domain and frequency domain.

Leakage current signal RMS is one of the most effective characteristic parameter of all leakage current signal features [7]. To obtain the effective value of the leakage current signal, we calculate the current sampling value within a certain period of time by using the following formula:

$$I_e = \sqrt{\frac{1}{T} \int_0^T x^2(t) dt} = \sqrt{I * I' / N} \quad (1)$$

It is difficult to identify the current operating state of the insulator by only the leakage current rms. The three characteristic values of the leakage current rms (The mean, maximum, and standard deviation of the rms values) can distinguish the three operating states of the insulator: Normal, development and flashover. The mean, maximum and standard deviation of the leakage current rms are calculated as follow formulas:

$$I_{me} = \frac{\sum_{i=1}^N I_e(i)}{N} \quad (2)$$

$$I_{maxe} = \max(I_e(i)) \quad (3)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (I_e(i) - I_{me})^2}{N}} \quad (4)$$

Three characteristic parameters describe the characteristics of current leakage current signal from different angles, and fully reflect the many device state characteristics which can not be reflected by single effective value.

Leakage current data in the frequency domain contains a wealth of information, the fractal dimension value based on the power spectral density represents the state monitoring data characterization of the frequency domain [9]. The larger the power fractal dimension, the greater the signal fluctuation, the weaker correlation between adjacent points in the signal, and the higher frequency component contained in the signal.

The AR model has the advantages of good linearity and good performance [10] in the calculation of power estimation. The formula is as follows:

$$P_x(e^{j\omega}) = \sigma^2 / |1 + \sum_{k=1}^p a^k e^{-j\omega k}|^2 \quad (5)$$

$\sigma^2$  is the variance of the excited white noise;  $P_x(e^{j\omega})$  is the power spectral density;  $a^k$  is the model parameter.

First of all, the power spectrum estimation of AR model needs to get the parameters  $\sigma^2$ ,  $a^1$ ,  $a^2$ , ...,  $a^k$  of AR by Yule-Walker equation by Levinson-Dubin recursive algorithm. Then, each point is described on the double logarithmic power spectrum  $\log p(\omega)$ - $\log \omega$  diagram, the straight line is fitted by the least square method, and the fitted straight line slope is K, The frequency domain fractal dimension  $D_h$  [11] based on modern power spectrum estimation can be expressed as

$$D_h = (5 + K) / 2 \quad (6)$$

## Design of Density - Based Cluster Structure Clustering Algorithm

When faced with massive leakage current data, we need to use clustering algorithm to deal with data effectively. The classification of current data clustering should have the following principle:

Because the geometry of the data in space is arbitrary, so the clustering algorithm can find any shape of the cluster; abnormal data does not affect the effect of clustering, clustering algorithm can deal with noise data and has strong anti-interference ability. Based on the above two points, we select the density-based clustering algorithm called DBSCAN [12].

Although DBSCAN has the above-mentioned advantages, but there is a problem of sensitive of input parameter  $\epsilon$ , MinPts. if the parameter is chosen improperly, it will cause the drop of quality of clustering, and time complexity is high, and consume memory is big. In order to overcome this shortcoming, the OPTICS cluster analysis method is proposed. Instead of explicitly generating data and clustering, OPTICS outputs the cluster ordering, which is a linear table for all analytic objects and represents the density-based clustering structure of the data [13].

OPTICS algorithm solves the problem that DBSCAN algorithm is sensitive to the input parameter  $\epsilon$ . However, OPTICS is far slower than DBSCAN because of the complicated processing method and the additional disk IPO operation.

Based on the idea of OPTICS clustering, this paper proposes a cluster structure clustering algorithm based on density (DBClustering)for clustering the data of insulator leakage current. Firstly, a tree storage structure is proposed, which maps the relationship of core reach ability among data points. Then, a sorted list of data clusters is constructed. The clustering results of leakage current are analyzed and the clustering results are evaluated according to this sorted list.

In this paper, we first design a CR-Tree construction algorithm, The algorithmic process is described below:

Algorithm 1 BulidCRTree(build CR-tree for D).

Input: Leakage current data set D;

Output: CR-Tree;

Step 1. Initializes a stack S;

Step 2. select a data point p unprocessed from D, find the point o that p directly core-reached, (Point o is not in the stack, if the stack has been pushed, then take smaller than the o-point relative to the core point instead of o);

Step 3. If the point o is not on the CR-tree, then point p is pressed to the stack S, deal with point o first, join the point o to CR-Tree, then update the point o to the father node of the point p;

Step 4. Then take the top element of stack to deal with, if the stack S is empty, then remove the next set of unprocessed data points from D to deal with. Go to step 1, until all points in dataset D are processed

In the case of no index, the time complexity of calculating the core distance is  $O(n^2)$  for all data points, and  $O(n)$  is the time complexity of computing the direct core distance for each data point. In general, the time complexity of algorithm 1 is  $O(n^2)$ .

CR-Tree contains the density relation among the data points, and the data can be extracted by simple transformation, which is shown in Algorithm 2.

Algorithm 2ExtraClu(extract clustering structure)

Input: CR-Tree;

Output: order points list L.

Step 1. Initialization data sorted list L is empty;

Step 2. Randomly we select a data points dp in the CR-Tree, and if the point dp has not been extracted (not inserted into the sorted point list L), the extraction of dp is started.

Step 3. Analyze the data points dp, if the father node of the point dp is empty, the point dp is directly inserted at the end of the list L; if the father node of the point dp is not empty, go to step 4;

Step 4. Analyze the data points dp, If the father node of point dp is in L, starting from the sequence number position of its father node in the list L, the core reachable distance of the data point in the list L is compared with the distance between the point dp, until the reachable distance of the point dp is greater than or equal to a point of list L's, then insert the point dp into that position; if it is greater than the core reachable distance of the point dp up to the end of L, go to step 5;

Step 5. The point  $dp$  is inserted at the end of the list, if the father node of point  $dp$  is not in list  $L$ , the father node of  $dp$  is nested processing first, go to step 3; If it is in, continue processing the next data point in CR-Tree to step 2.

The output data is sorted list  $L$  until all of the data points in the CR-Tree are processed.

### Experimental Test And Analysis

In this section, first we extract the characteristic data of the leakage current data of insulator, then form the sample set  $D$  of the leakage current data. The clustering efficiency and clustering efficiency of DBClustering algorithm are experimentally evaluated and compared with OPTICS and optimized OPTICS-G algorithm.

Experimental platform: 2.4GHz i3 processor, 4G memory, Window7 operating system and matlab software. Test data is from the leakage current feature extracted from previous experiment. the data set  $D$  ( $I_e$ ,  $I_{maxe}$ ,  $\sigma$ ,  $D_h$ ). The clustering effect of the algorithm is tested on  $D$ , and the processing time index (CPU time) and the space cost index (memory) of the algorithm are evaluated.

### Experiment on Feature Extraction of Leakage Current

Select a section of leakage current data monitored in the laboratory to calculate the Eigen values of time and frequency domain. The waveform of the collected data is shown in Fig.1.

In order to analyze the leakage current signal, we choose 2s as the length of the sample. First, calculate the effective value of the current recorded in each cycle (20ms). Then select 100 effective values of the same insulator's time stamp and calculate the three characteristic quantities of the leakage current: mean value, maximum value and standard deviation.

The frequency domain fractal dimension is calculated as follows. First, the power spectrum estimation is calculated, and the result is shown in Fig.2. According to the power spectrum estimation, the double logarithmic power spectrum is shown in Fig.3.

The slope of the curve fitted with least squares method is  $-0.4085$ , so the fractal dimension of frequency domain characteristic of power spectrum estimation is calculated as  $D_h = (5 - 0.4085) / 2 = 2.2956$ .

The time-domain and frequency-domain eigenvalues of the insulator leakage current data measured in the laboratory are calculated, and some results are shown in Table 1.

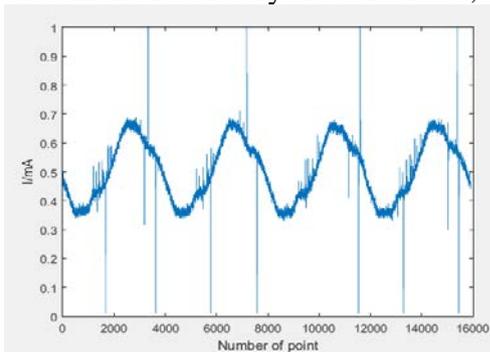


Fig.1 Measured insulator leakage current waveform data

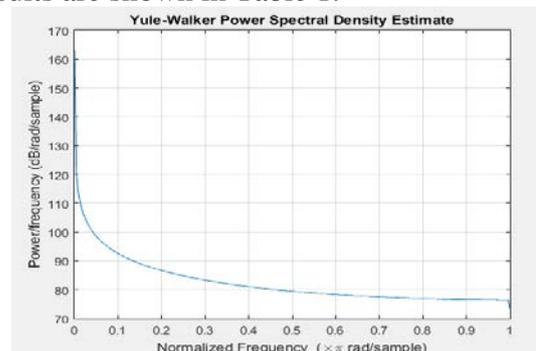


Fig.2 Modern power spectrum estimation curve

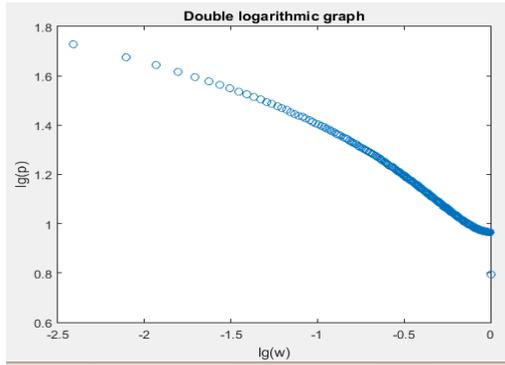


Fig.3 Double logarithmic graph

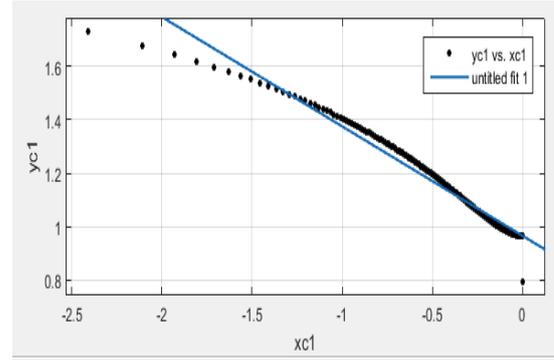


Fig.4 Curve fitting

**Clustering Experiment of DBClustering Algorithm**

Using DBClustering's ExtraClu algorithm to extract the data sort table of density cluster structure, Extra Clu extraction is as simple as the sorting process of OPTICS, but Extra Clu can only sort the data by comparing the core reachable distance of partial data from CR-Tree, and the extraction process is very fast. The resulting sorted list L is visualized as shown in Figure 5 (a), Figure 5 (b) shows the cluster structure of OPTICS. It can be seen that the clustering order in the cluster structure is different, but the independent cluster results can correspond to each other, and the internal cluster structure is basically similar, and the leakage current is divided into three categories: normal, development, flashover . corresponding to the three state of leakage current.

Tab.1 Fractal dimension value of measured insulator

Number	$I_{me}$	$I_{maxe}$	$\sigma$	$D_h$
1	0.5044	0.6629	0.2034	2.2956
2	0.5073	0.6606	0.2036	2.0084
3	0.6069	0.7663	0.2043	1.0812
4	0.6079	0.7846	0.2035	1.0247
5	0.6067	0.8829	0.3032	0.2876
6	0.6110	0.8998	0.3174	0.2327

Figure 6 shows the time costs of the OPTICS, OPTICS-G, and DBClustering algorithms. OPTICS consumes the most CPU time, far more than the other algorithms. When the amount of data is 2K, DBClustering processing is faster than OPTICS-G increased by 48.3%.

Figure 7 shows the memory consumption of the OPTICS and DB Clustering algorithms. On the D data set, DBClustering has a large space overhead, saving 41.6% of memory consumption compared with OPTICS.



(a)The cluster structure of DBClustering on D (b) The cluster structure of OPTICS on D

Fig.5 Effectiveness comparison of DBClustering and OPTIC

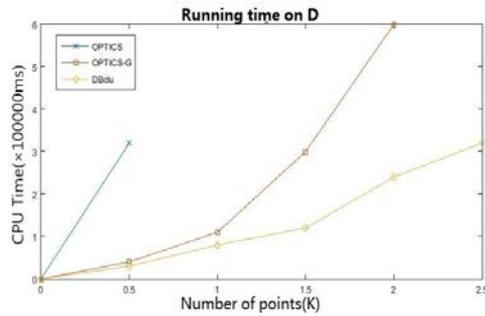


Fig.6 Efficiency comparison of algorithms

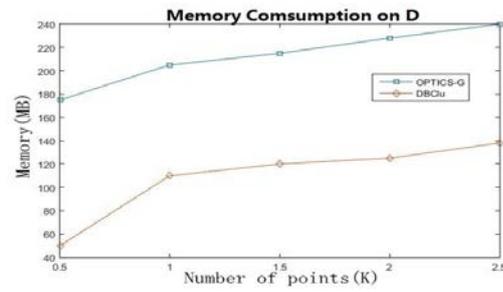


Fig.7 Memory comparison of algorithms

## Conclusion

In this paper, four characteristic values of leakage current (maximum value of leakage current, mean value, standard deviation, power spectrum fractal dimension) are extracted from the two aspects of leakage current in time domain and frequency domain, a cluster structure clustering algorithm based on density named DBClustering is designed, and it is helpful to analyze the state of leakage current assessment and flashover studies. A lot of monitoring data of insulator leakage current is often accumulated in the short term, how to analyze it in real-time and efficient has become a serious problem,. Cloud computing platform is a solution to solve this problem. The next step of this paper is to combine the new cloud platform to analyze the large data of leakage current and improve the efficiency of feature extraction and clustering algorithm.

## Acknowledgement

In this paper, the research was sponsored by Scientific research project of Hebei Province (Project No.Z2012077).

## References

- [1] Carry Melik, Zeng Dong, Lin Zhou, Ye Ziqing, Insulator leakage current monitoring, its application and advantages [J], 2014 China International Conference on Electricity Distribution (CICED), 2014.9:1-4.
- [2] Miao Pengchao, Xu Zhiniu, Summary of the Transmission Line Insulator Leakage Current Characteristics Extraction [J], High Voltage Apparatus, 2014,08:127-138;
- [3] Chen Weigen, Xia Qing, Li Jingyan, Yao Chenguo, Sima Wenxia, Time-frequency Characteristic Analysis of Leakage Current for New Characteristics of Insulators Contamination Prediction [J], High Voltage Engineering, 2010,05:1107-1112.
- [4] Chen Weigen, Xia Qing, Sun Caixin, et al. Research on fractal dimension of leakage current for discharge zones dividing and contamination forecasting of insulators [J]. Proceedings of the CSEE, 2011, 31(13): 121-127.
- [5] Zhao MingShu; Liu JianZheng, Insulator leakage current analysis based on time-frequency and HHT methods [J], 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), 2015.01:1365-1369.
- [6] Zhou Guoliang, Song Yaqi, Wang Guilian, Zhu Yongli, Research of Condition Monitoring Big Data Storage and Clustering [J], Transactions of China Electrotechnical Society, 2013, S2:337-344.
- [7] Gao Yongkuan, The research of extraction algorithms of insulators leakage current data characteristics [D], North China Electric Power University, 2013.

- [8]Zhao Shihua, Jiang Xingliang, Zhang Zhijin, Hu Jianlin, Flashover Voltage Prediction of Polluted Glass Insulators Based on the Characteristics of Leakage Current[J], Power System Technology, 2014,02:440-447.
- [9]Chen Weigen,Wang Wanping,Xia Qing, Research on Multi-fractal Characteristics of Leakage Current for Contamination Discharge of Insulators[J], Transactions of China Electrotechnical Society, 2013,01:50-56
- [10]Yao Zhifei, Jiang Wanlu,Zhu Yong, AR Model-based Power Spectrum Estimate Method and Application in Fault Detection[J], Machine Tool & Hydraulics, 2013,05:177-179+186.
- [11]Maximilian Hoecker, Kai Lars Polsterer, Sven Dennis Kugler, Vincent Heuveline, Clustering of Complex Data-sets using Fractal Similarity Measures and Uncertainties[J], 2015 IEEE 18th International Conference on Computational Science and Engineering.2015.10:82-91;
- [12]Jin Jianye, Ni Zhiwei, Wang Sha, Data Stream Clustering Algorithm Based on Density and Fractal Dimension[J], Computer Engineering, 2012,05:38-40.
- [13]Yu Yanwei, Wang Huang, Wang Qin, Zhao Jindong, Density-Based Cluster Structure Mining Algorithm for High-Volume Data Streams[J], Journal of Software, 2015,05:1113-1128.