An Edge Computing Method for Partial Discharge under the Power Internet of Things

Shaojing Wang¹, a, Maoxin Ren¹, Pei Cao¹ and Peng Xu¹

¹Electric Power Research Institute, Shanghai Municipal Electric Power Company, Shanghai, China

Corresponding author: ajiapengtian@sjtu.edu.cn

Abstract. The recognition of impulse interference for the partial discharge data on the edge computing platform can effectively reduce the false alarm rate in the risk assessment of gas insulated switchgear (GIS). Therefore, an edge computing method for partial discharge is proposed in the paper, which uses the similarity computing among the aggregation of data for each sensor after clustering to match homologous signals. Finally, the algorithm will jointly locate the interference signal and recognize the impulse interference signal by comparing the received signal strength indicator (RSSI) of multiple UHF sensors in/out of GIS equipment. The training set, being consisted of partial discharge and interference signal from the test, is used to contrast the algorithms on the edge computing platform, the result of which indicates that the accuracy rate of the recognition of the impulse interference with the proposed algorithm is 81.4% on the computing platform, with the average processing time for each sample of 7.15s.

1. Introduction

In the power Internet of Things, mass of sensory data collected through sensors is up-loaded to access points through aggregation nodes and finally access into the server of the cloud platform through the access controller and the control gateway [1]. If it is possible to send data to the network edge, rather than the data center or the Cloud, the data will be processed and analyzed on the data source [2], which could effectively reduce the computational burden of the server on the Cloud platform.

Consequently, the intelligent sensors with the capacity of the edge computing are indispensable in the power Internet of Things. It is the core supporting technology of the perception layer in the power Internet of Things, the in-vestment for which is far more than other technologies and increased continuously [3]. The edge computing splits the large-scale services which should have been processed by central nodes into several smaller parts which are easy to manage and then distribute such parts to fringe nodes for processing. Such structures relief the computational burden of central nodes and reduce the performance requirement for the processors of central nodes, with a high robustness. As the first-hand data is processed on the fringed nodes, the perception of the system to the external environment is improved greatly, which makes the system could responds to the change of the external environment quickly and effectively. Thus, it is essential for the study of the edge computing.

The total awareness of transmission and distribution equipments status is the precondition to ensure the safe operation of a power grid, improve the ruling ability of a power grid and realize the intelligent run of a power grid [3]. Therefore, in the paper [4], it proposes an effective GIS risk assessment
method based on the deep study for partial discharge. For the risk assessment of GIS equipments, it proposes to measure the partial discharge first for investigating the defects of equipment insulation. However, the false alarm always happens during the practical operation, which would interfere with the normal operation of the system to a large extent. The survey indicates that most of the false alarms of partial discharge are resulted from the impulse electro-magnetic interference similar to partial discharge. The interference signals are classified into three types, which are continuous periodical interference, white noise interference and impulse interference [5]. Comparing with other two interference signals, the processing of the impulse interference signal resulting in false alarms is very complex. The impulse interference mainly comes from corona discharge at the high voltage terminal, thyristor rectification, floating potential discharge and others. It shows that the interference could be resulted from partial discharge with the characteristics being similar to the partial discharge in the equipment, which is difficult to be recognized through the feature recognition of typical machine learning algorithm.

Up to now, the scholars at home and abroad have done lots of researches for interference signals and noise signals in the partial discharge data. However, such researches only focus on analyzing the white noise by using time-frequency domain, filtering periodic narrow-band interference and analyzing the characteristics of impulse interference which is very different from partial discharge by using the machine learning algorithm and others [6-9]. It is difficult to recognize the impulse interference (which is the partial discharge) in such method (such as floating interference and corona interfere).

The mean shift clustering algorithm is an algorithm based on the density calculation, with the characteristics of no need to set the classification number, good adaptability to the data of probability distribution and others, which is widely used for target tracking, fingerprinting method and other engineering [10-11].

In the paper, it proposes an effective edge computing method for partial discharge data based on mean shift clustering, aiming to recognize and process the interference of partial discharge data collected by sensors. It is capable to locate the same source of partial discharge on the edge computing platform in such method for recognizing and removing the impulse interference signal by adopting several sensors to jointly measure, combining with mean shift clustering algorithm and making comparison with the received signal strengths of sensors. Finally, the effective partial discharge data on the edge would be kept computing platform and then uploaded under the instruction of the other segments in the system.

2. The edge computing flow and algorithm for partial discharge

2.1. The edge computing flow for partial discharge

UHF sensors with partial discharge are used for the test to display the characteristics and intensity of partial discharge, to some extent, by detecting high-frequency electromagnetic wave of the partial discharge. As the impulse interference signal could not be processed by the analysis of traditional time-frequency domain or feature recognition, it is an effective method to use multiple sensors to jointly locate.

When processing data, it will convert the UHF PRPS data (phase resolved impulse series) into PRPD (phase revolved partial discharge) data and then arrange the mean shift clustering for recognizing different partial discharge and interference source. When the internal partial discharge matches with each partial discharge data of other sensors, if no matching signal is found, it means that such signal could be only found on the sensor which is probably a partial discharge signal. If there is matching signals, it shall recognize again. Due to the attenuation of electro-magnetic wave after passing through the shell of GIS equipment, it could locate by comparing with the received signal strength indicator (RSSI) of the sensors with partial discharge. If the amplitude from external sensors is large, it would indicate that the signal source comes from the outside of the equipment, which would be an interference signal; otherwise, it would be an internal signal with partial discharge. The data
processed by signals is kept on the edge computing platform and then take actions under the request of the other segments in the system. Refer to Figure 1 for other algorithm flow chart.

![Flow Chart](image)

**Figure 1.** The Edge Computing Method Flow Chart

2.2. Matching homologous signal and mean shift drifting algorithm

The UHF partial discharge data measured actually not only the effective internal partial discharge signals of the equipment, but also the external interference signals of the equipment. Therefore, it is necessary to recognize the different partial discharge sources in UHF PRPD data by clustering.

In the paper, it clusters the partial discharge data with mean shift clustering algorithm. Such algorithm includes three steps which are searching in the clustering center, clustering/merging the similar areas in the clustering center and merging small domains [12]. To be specific, it shall randomly select the central point $x$ in the characteristic space with $N$ sample points and calculate the average offset vector $M$ of the point set included in the high-dimensional ball $S_h$ centering in such point and with the radius of $h$, as indicated in the formula 1.

$$M(x) = \frac{1}{k} \sum_{x_i \in S_h} (x - x_i)$$

(1)

Whereas, $k$ – the number of sample points in High-dimensional ball $S_h$.

Then, move the central point $x'$ to the position of mean shift $x^{t+1}$ till the mean shift meets the set threshold condition, as indicated in the formula 2.

$$x^{t+1} = M^{t} + x'$$

(2)

Thus, it can be seen that the clustering method makes each central point move to the direction with high density, which is very suitable for the partial discharge clustering in PRPD data.

It is necessary to introduce the kernel function is introduced in case that the data in the characteristic space in lower dimension is not linearly separated. After introducing the kernel function, it could directly invoke the function to map the sample in the high dimensional space without the specific mapping relation so that the data set could be separated linearly. The average offset vector introducing the kernel function $m_k$ is indicated in the formula 3.

$$m_k(x) = \frac{1}{\sum_{i=1}^k x_i g(\|x - x_i\|/h)} \left[ \sum_{i=1}^k x_i g(\|x - x_i\|/h) - x \right]$$

(3)
Whereas, \( g(x) \) – negative derivative of the kernel function \( h \). \( k \) is same as the formula 1.

In such method, the clustering center in the sample could be determined at first. And then the similarity and distance of each center could be compared. The centers with close distance could form a new clustering center so as to complete the clustering computation.

After complete the clustering of UHF partial discharge PRPD data, the location of each clustering center for internal and external data could be calculated. Then, the distance \( r \) of super-parametric matching data could be defined to make similarity contrast with external data cluster by the reference of the clustering center of internal data. If one of the external clustering central points has a distance of less than \( r \) to the internal central point, such two clusters are named as a matching cluster. If several external clustering central points have distance of less than \( r \) to the internal central point, the two nearest clusters shall be a matching cluster.

After completing the matching, the internal data unmatched with any external clustering center would be directly recognized to be partial discharge signals. When comparing the amplitude of the internal and external clustering centers in the matching clusters, if the absolute value of \( y \) coordinate for the internal cluster-based data is less than the one for the external cluster-based data, it shall delete all points of the cluster in the internal partial discharge data. In order to saving the memory resource of the edge computing platform, only the internal processed partial dis-charge data file could be kept after the signal process.

3. Experimental verification

3.1. Experimental scheme

In the paper, it uses two UHF partial dis-charge sensors to make measurement, one of which is placed inside GIS experimental platform and another one placed outside the experimental platform, to collect data in combination with different defect models and external meaconing. During the experiment, both sensors transfer the partial discharge experimental data to the edge computing platform at the same time. Such transferred data is kept on the edge computing platform and then processed according to the data processing instruction of the edge computing platform.

During the experiment, five insulation defect models including suspended discharge, point discharge, discharge along dielectric surface, discharge in the insulation and particle discharge are installed inside the experimental platform in order, which stimulate suspended discharge interfere and corona discharge interfere signals outside the experimental platform. Each of the five defect models could measure 200 pairs of discharge data, 1000 pairs of experimental data in total. Refer to Figure 2 and 3 for the experimental data.

![Figure 2. Particle Discharge with Suspending ImpulseInterfere](image1)

!*Figure 2. Particle Discharge with Suspending ImpulseInterfere*

The experiment is executed in the operation system of Ubuntu 16.04.3 server with 32G external memory. The development language of algorithm is Python 3.5.

In the section 1.2 of the paper, it defines a super-parametric matching distance. During the debugging of the procedure, it found that the error rate of the experimental data reduces rapidly and
then increases slowly as the parameter increases, which is because many interference signals of no matching signal source are considered to be the internal partial discharge signal of the equipment when the matching distance is small, meanwhile some partial discharge signals of no matching source are considered to be interference signals after finding their signal sources when the parameter is large. In the case that the parameter is between 10 and 12, the data is processed at a lowest error rate. Considering that the larger the matching radius is, the heavy the computing burden of the edge computing platform is, so the final selected value shall be 10.

The experiment is completed offline. The interference recognition is made for measured data which is kept on the specified path of the edge platform through the average shift clustering algorithm.

3.2. Experimental scheme

3.2.1 Analysis of examples. In the Figure 4 and 5, it shows the same point discharge signal measured in the experiment. The partial discharge data includes suspended interfere signal. The signals in red are the point discharge inside the equipment and the signals in black are suspended interfere needed to be recognized and filtered.

![Figure 4. Diagram of Point Discharge PRPS with Suspended Interference](image1)

![Figure 5. Diagram of Point Discharge PRPD with Suspended Interference](image2)

In Figure 6, it shows the after-clustering effect, which is worth noting that the same partial discharge sources are not always classified into the same cluster, but the data in the same cluster definitely comes from the same partial discharge source.

![Figure 6. Clustering PRPD Data by Mean Shift](image3)

3.2.2 Contrastive analysis. In order to study the performance of the edge algorithm, Table 1 displays the probability of removing impulse interference from each kind of defect data processing by using the edge algorithm. Table 2 displays the comparison between mean shift algorithm and traditional K-
means with elbow method. The effective rate in the table means the probability of effectively removing impulse interference from 1000 pairs of data. The algorithm processing time means the average time to process each pair of the data edge platform, which includes the processing time of clustering and the processing time of the white noise with wavelet threshold value.

**Table 1.** List of Defect Data Processing Based on Mean Shift Algorithm

| Defect type                  | Aggregate data | Effective data | Effective rate /% |
|------------------------------|----------------|----------------|-------------------|
| Point discharge              | 200            | 189            | 94.5              |
| Particle discharge           | 200            | 121            | 60.5              |
| Suspended discharge          | 200            | 174            | 87.0              |
| Internal insulation discharge| 200            | 169            | 84.5              |
| Discharge along dielectric surface | 200        | 161            | 80.5              |
|                              | 1000           | 814            | 81.4              |

**Table 2.** Mean Shift Algorithm vs K-means Algorithm

| Algorithm     | Effective rate /% | Average processing time /s |
|---------------|-------------------|----------------------------|
| Mean shift    | 81.4              | 7.15                       |
| K-means       | 60.5              | 5.53                       |

Based on the experimental result, we can see that the edge algorithm using mean shift clustering can effectively remove the impulse interference from the partial discharge signals. Besides, the mean shift clustering algorithm is based on the clustering algorithm ascended with the density gradient, which has a better performance in the edge algorithm comparing with the traditional clustering algorithm, but needing a long processing time.

On the other hand, it is obvious in Table 2 that K-means algorithm takes less time than the mean shift algorithm for processing because the maximum number of classifications for K-means algorithm is set as a low value after comprehensively considering the performance of the algorithm. In such method, it obtains the proper number of classifications in combination with the elbow method, the processing time would increase greatly as the maximum number of classifications increases. In other words, it takes more time for processing with K-means algorithm versus the mean clustering algorithm for obtaining a good result.

4. Conclusion

In the paper, it, with the mean shift clustering algorithm, proposes an edge computing method for partial discharge to remove the impulse interference, while uses the Atlas 200 DK as the edge platform to verify in experiment, which shows:

1) The edge algorithm for data processing in the paper has the accuracy rate of 81.4%, which can effectively recognize the impulse interference from the partial discharge data. Comparing with the traditional clustering algorithm, the effective rate of such algorithm is effectively improved by 20%.

2) The edge algorithm in the paper has the characteristics of low cost and easy to application, with the worth of popularizing.

3) In the paper, it adopts PRPD data with the cumulative time of 1s for testing the algorithm, which is suitable for the low-power dissipation of the wireless Internet of Things. The cumulative time of PRPD could be extended properly and the edge computing platform with more powerful computing power could be also used to improve the accuracy rate and comprehensive performance of the data process if being applied in wired Internet of Things and having no need to considering power dissipation.
Acknowledgments
This work is supported by Science and Technology Project of State Grid Shanghai Municipal Electric Power Company (Research on Intelligent Performance Evaluation System of Power Sensing Device, No. SGSHDK00SPJS2000236).

References
[1] Jiang Xiuchen, Luo Lingen, Yu Zhongmin, Fu Xiaofei, Sheng Gehao, Liu Yadong, Qian Yong. Key technologies and solutions for the application of blockchain in the ubiquitous Internet of Things in power equipment [J]. High Voltage Technology, 2019, 45(11): 3393-3400 (in Chinese).
[2] Gunasekaran Manogaran, Naveen Chilamkurti, Ching-Hsien Hsu. Machine learning algorithms towards merging of mobile edge computing and Internet of Things[J]. Computer Networks, 2019, 161.
[3] Jiang Xiuchen, Liu Yadong, Fu Xiaofei, et al. Ideas and development trends of ubiquitous power Internet of Things for transmission and distribution equipment [J]. High Voltage Technology, 2019, 45(5): 1345-1351 (in Chinese).
[4] Song Hui. GIS risk assessment method based on partial discharge deep learning [D]. Shanghai Jiaotong University, 2018 (in Chinese).
[5] Zhao Yangyang. Partial discharge signal analysis and interference suppression of large power transformers [D]. Chongqing University of Technology, 2012 (in Chinese).
[6] Suganya Govindarajan, Jayalalitha Subbaiah, Andrea Caval-lini, et al. Development of Hankel-SVD hybrid technique for multiple noise removal from PD signature. 2019, 13(8):1075-1084.
[7] Muñoz-Muñoz Fabio, Rodrigo-Mor Armando. Partial Discharges and Noise Discrimination Using Magnetic Antennas, the Cross Wavelet Transform and Support Vector Machines.[J]. Sensors (Basel, Switzerland), 2020, 20(11).
[8] N. Morette, L. C. Castro Heredia, Thierry Ditchi, A. Rodrigo Mor, Y. Oussar. Partial discharges and noise classification under HVDC using unsupervised and semi-supervised learning[J]. International Journal of Electrical Power and Energy Systems, 2020, 121.
[9] Wei Wang, Weiwen Peng, Muqin Tian, et al. Partial discharge of white noise suppression method based on EEMD and higher order statistics. 2017, 2017(13): 2043-2047.
[10] Shen Hao, Zhuang Jianjun, Zheng Qianying, Wu Jianyao. Design and implementation of a target tracking system based on MeanShift algorithm[J]. Electronic Measurement Technology, 2018, 41(14): 11-15 (in Chinese).
[11] Li Shiyin, Du Zhongxiang, Zhu Yuan, Li Zongyan, Wang Xiaoming. A UWB fingerprint location method based on MeanShift and weighted k-nearest neighbor algorithm [P]. CN109511085A, 2019-03-22 (in Chinese).
[12] X. Wang, H. Liu, W.L. Ma. Sparse least squares support vector machines based on Meanshift clustering method[J]. IFAC PapersOnLine, 2018, 51(18).