IET Science, Measurement & Technology

Research Article

Online monitoring of transformer through stream clustering of partial discharge signals

ISSN 1751-8822
Received on 10th July 2018
Revised 16th October 2018
Accepted on 3rd December 2018
E-First on 28th January 2019
doi: 10.1049/iet-smt.2018.5389
www.ietdl.org

Keyvan Firuzi[1], Mehdi Vakilian[1] ✉, B.Toan Phung[2], Trevor Blackburn[2]

[1]Electrical Engineering Department and Centre of Excellence in Power System Management and Control, Sharif University of Technology, Tehran, Iran

[2]School of Electrical Engineering & Telecommunications, University of New South Wales, Sydney, Australia

✉ *E-mail: vakilian@sharif.ir*

Abstract: The general method for identifying the partial discharge type in a power transformer is based on their fingerprints in the form of phase-resolved discharge patterns. In the case of multiple defects, traditional clustering methods can be applied for separation of active sources. However, such an approach is impractical for online real-time monitoring due to the very large data size. In this paper a new method using stream clustering is introduced. The method separates the active sources by processing the signal once it is captured, then only a synopsis of the discharge data is stored. Two stream clustering algorithms: Density Grids and DenStream are employed. Through measurements obtained from laboratory experimental setups (corona, surface discharge, transformer defect model) performance of the proposed algorithms are evaluated. It is shown that stream clustering method is able to separate the constituent components involved in the stream of a multi-source discharge signal without the need to store a large amount of information. The performance of the Density Grids method depends on a limited number of features that it can accommodate. In comparison, the DenStream method can capture more features which enable better separation of active sources at the expense of longer processing time.

1 Introduction

Condition monitoring of power system equipment improves plant economy, increases availability and service life [1]. Being able to perform condition monitoring online provides the opportunity for continuous operation service, early detection of problems and possible remedial action and so increases the lifetime expectancy of power transformer [2].

Partial discharge (PD) measurement is an effective method for insulation condition assessment and ensure the reliability of power transformer. The advantage of online monitoring of in-service power transformer is that insulation condition is assessed with the transformer operating under both normal working conditions and abnormal variations related to changes in electrical, thermal, and mechanical operational stresses.

In general, PD does not cause immediate failure of in-service power apparatus. However, if the PD is not properly detected and the source identified, it might eventually lead to complete insulation breakdown. Knowledge of the PD characteristic (e.g. corona, surface, internal discharges) can be obtained from the detected PD signal and this can help to estimate its damaging impact [3].

The general method for identifying the PD type is based on the phase-resolved PD (PRPD) patterns which plot the PD events in terms of their magnitude and phase position. However, more than one PD source may exist due to multiple defects in the test object. The general method for separation of multi-source PD is based on characteristic features extracted from PD pulse waveforms and applied to some signal separation process [4–8]. Traditional clustering methods like k-means [9, 10], fuzzy c-means [11], hierarchical clustering [12], density-based spatial clustering of applications with noise (DBSCAN) [13, 14], affinity propagation [15] and stochastic neighbour embedding (SNE) [16] have been used for PD source separation. However, these standard algorithms need access to all the data points and typically iterate over the data set multiple times. This requirement makes these algorithms unsuitable for online monitoring.

Under online PD monitoring, the recorded data is no longer viewed as a static collection. Furthermore, it is potentially a very large set of dynamic data or stream of captured PD pulses that requires the development of stream clustering algorithms for its analysis.

This paper presents a method for clustering the stream of PD pulses. Traditional clustering methods store the PD data at the first step and the data is then analysed. Since the data size under online measurement can be extremely high, its storage and processing can become a very difficult task. In contrast, stream clustering methods access and process the data once (data will not be saved) and only a synopsis of the data is stored (consequently it will not face any storage shortage).

The remaining of the paper is organised as follows. Section 2 presents the experimental setup used for verification of the proposed method on stream clustering of PD pulses. In Section 3, practical issues of online PD signal processing are discussed. In Section 4, the general concept and structure of stream clustering methods are introduced and two methods, namely density grids and DenStream, are described. Their advantages and drawbacks are examined in Section 5 through a case study. Section 6 concludes and summarises the findings.

2 Experimental test setup

The experimental setup is shown in Fig. 1. The AC test voltage is generated by a 50 Hz, 5 kVA, 100 kV transformer. A 10 MΩ resistor is used as a current limiter and a 300 pF capacitor as a coupling capacitor. PD current pulses and the related phase angle of their occurrences are captured via measuring impedance in series with the coupling capacitor.

Three artificial PD models used in this study are shown in Fig. 1. The corona model is constructed by a needle-plane electrode system. Surface discharge is simulated using a sphere-plane electrode setup with one layer of insulating paper in between. These two external sources are usually observed in power transformers and always a disturbing factor in the PD measurement. The internal discharge is simulated from a void created in a single-phase transformer model. More detail about this transformer model can be found in [17].

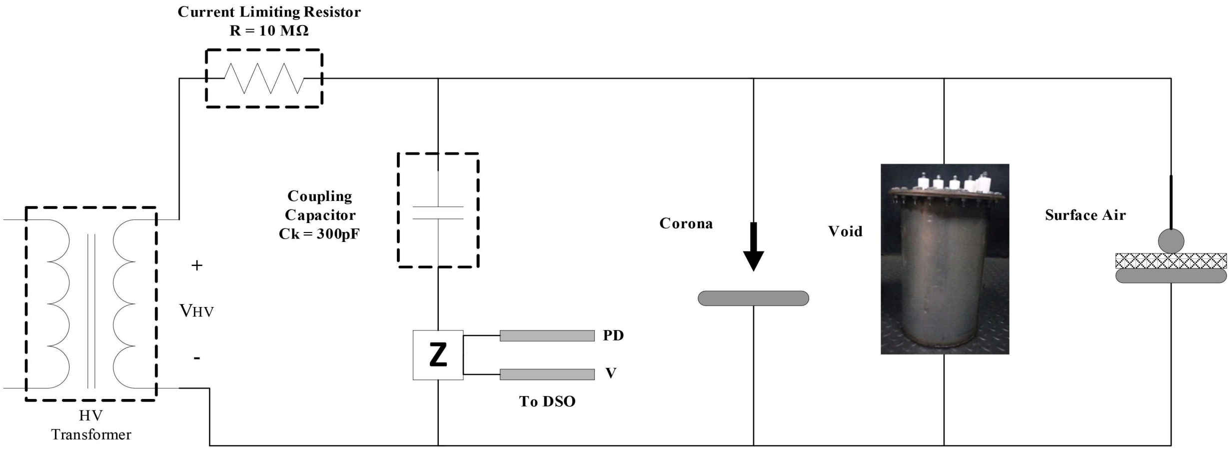

Fig. 1 *Experimental setup*

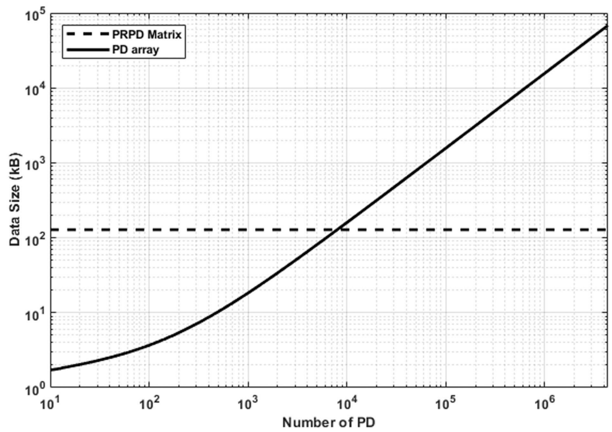

Fig. 2 *Data size versus number of PD pulses (over time) for the two storage methods*

The three PD models are combined together in parallel to simulate multi-source PD. Measurements are performed under the application of an AC voltage which is increased slowly in small steps. When the applied voltage exceeds the highest PD inception voltage of the defects involved in the combined model, PD will occur in all three sources.

The PD signals and voltages were recorded using a 600 MHz bandwidth digital oscilloscope. Each PD signal was captured over duration of 10 µs at a sampling frequency of 100 MHz, and the relevant phase angle (φ) data are acquired through the oscilloscope.

3 Online PD signal processing

Due to the following reasons the PD data size can increase dramatically:

- *Noise:* all pulses present must be first recorded with the assumption that they are PD pulses. During online measurement, the noise level is high which raises the number of saved pulses significantly.
- *PD generation rate:* different PD sources have a different PD generation rate. The measurement time should be long enough to obtain sufficient information about the PD with the lowest generation rate which makes the PD recording time high.
- *Feature extraction:* under multi-source condition (the most common situation is online PD measurement), the appropriate number of features must be extracted for PD source clustering. This causes the size of data multiplied.

The streaming nature of online PD data (stream data) has posed challenges to both database management and data mining methods. One method to address the problem when facing a large amount of data under online PD monitoring is to use the PRPD matrix instead of storing each φ–q data point. To do this, PRPD matrix is first initialised as a 128×128 zero matrix and subsequently updated for each new PD pulse as below:

$$i = \left\lceil \frac{p}{360} \times 128 \right\rceil \tag{1}$$

$$j = \left\lceil \frac{q - q_{\min}}{q_{\max} - q_{\min}} \times 128 \right\rceil$$

$$\mathrm{PRPD}(i, j) = \mathrm{PRPD}(i, j) + 1$$

where p and q are the PD phase angle and charge, respectively. $q_{\min}$ and $q_{\max}$ denote the minimum and maximum charges, respectively. $\lceil . \rceil$ represents the ceiling function for rounding to the nearest integer in the direction of positive infinity. Since the PRPD data is recorded in terms of the above matrix elements, its size does not change with time, and its data size does not increase. Fig. 2 shows how the data size changes when one of the two storage methods is employed. As demonstrated in Fig. 2, if every PD pulse (φ–q) is stored, the size of data is linearly increasing with the number of PD pulses. However, when PRPD matrix is used, the size of data stored is limited and independent of the number of PD pulses.

If the PRPD matrix method is used to store PD data in the presence of multi-source PDs, the information interpretation becomes extremely difficult. Consequently, the multi-source recognition process involves the following steps. Firstly, the PD sources are separated within the PD pulse feature space, using a clustering method. Next, using an identification algorithm, the sub-PRPD patterns of step one are identified. To apply this general method to the online mode process, the clustering algorithm must satisfy the following requirements [18–21]:

- *One-pass constraints:* each PD pulse must be analysed only once.
- *Limited memory:* due to memory size limitations, all the raw data cannot be stored. Thus, clustering methods that require the entire data set cannot be used.
- *Real-time:* the algorithm has to process PD data points on average at least as fast as the PD pulse is arriving.
- *Number of clusters:* the number of active sources (clusters) is unknown. Furthermore, in an evolving PD pulse stream, the number of active PD sources is often changing. So no assumption can be made on the number of clusters.
- *Clusters shape:* the shape of the cluster in the feature space is unknown and can have an arbitrary shape that depends on the PD source and feature type.
- *Ability to handle outliers:* in the PD stream scenario, due to various factors, some random noise may appear occasionally.

A general method for clustering of stream data, which can meet the above requirements, is called 'stream clustering'.

4 Stream clustering of PD signals

A stream clustering algorithm consists of two basic phases [22–24]:

1. *Online phase:* which processes the incoming stream data points to summarise them into microclusters. Microclusters should include sufficient temporal and spatial information such as cluster centres and additional statistics such as weight (density) and dispersion (variance), to facilitate cluster formation.
2. *Offline phase:* which generates clusters from the microclusters, either periodically or on demand by the user.

The cluster partitions, on evolving data streams, are basically computed over a set of certain time intervals (or windows). There are three well-known window models: landmark window, sliding window and damped window. In the landmark window model, all arrived data until now are used for clustering. In the sliding window model, a specified amount of recently arrived data is considered for processing. In the damped window model, the recent data are given higher importance than older ones. Fading is usually implemented by assigning weights to the instances so that most recent data possesses higher weights [25].

4.1 Online phase

Online phase maintains statistical information about the data locality in terms of microclusters. These microclusters are obtained using extracted features from PD pulses. The online microclustering component requires a very efficient process for storage of appropriate summary statistics in a fast PD data stream [22].

The aim of online phase is to maintain statistics at a sufficiently high level of granularity so that it can be effectively used by the offline phase. In the following, two density-based stream data clustering methods for online phase are introduced. The Density Grids method is a simple scheme that can be very useful for understanding stream clustering of PD pulses, and DenStream is an advanced method for discovering clusters of arbitrary shape.

4.1.1 Density grids: This method is based on partitioning the multi-dimensional data space into many density grids, and assigning a PRPD matrix to each grid (i.e. micro PRPD). It continuously reads new data record, places the multi-dimensional data into the corresponding discretised density grid in the multi-dimensional space, and updates the micro-PRPD according to (1). Thus, we do not need to retain the raw data and only need to update and analyse the micro PRPD. This concept is schematically illustrated in Fig. 3.

Assuming the existence of two features such as f_1, f_2, then four parameters: f_1, f_2, φ and q, are calculated for each PD pulse. Using these two features, the PRPD matrix to be updated is computed and updated using φ and q:

$$k_1 = \left\lceil \frac{f_1 - f_{1\min}}{f_{1\max} - f_{1\min}} \times n \right\rceil \quad (2)$$

$$k_2 = \left\lceil \frac{f_2 - f_{2\min}}{f_{2\max} - f_{2\min}} \times m \right\rceil$$

$$k = k_1 + (k_2 - 1) \times n$$

$$\text{PRPD}_k(i, j) = \text{PRPD}_k(i, j) + 1$$

where $f_{1\min}$ and $f_{1\max}$ are minimum and maximum values of the first feature, and $f_{2\min}$ and $f_{2\max}$ are minimum and maximum values of the second feature. The first and second features are divided into n and m parts. k is the number of grids and PRPD_k denotes the corresponding PRPD matrix. i and j are obtained from (1).

As the number of features increases, the number of micro-PRPD can be significantly increased. For example, if each feature is divided into 10 grids then there would be 100 PRPD matrices for 2 features and 1000 PRPD matrices for 3 features. For this reason, this method can only be implemented for 2 or 3 features at most. Each feature is divided into several sections, depending on their importance. The more important the feature, the more divisions are needed. To use this method, there should be proper information about the features, including minimum and maximum values and their importance. In reality, most micro PRPDs are empty or only contain few PD pulses.

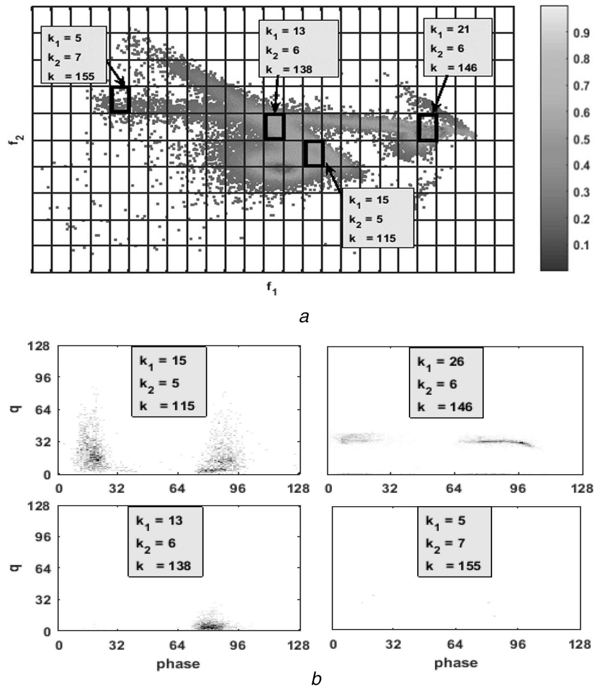

Fig. 3 *Density grids method concept*
(a) Density grid of two feature (colour indicates the density of points), ***(b)*** Micro-PRPD of each grid

4.1.2 DenStream: The second method used in the online phase is DenStream. This is a density-based stream clustering method [26]. DenStream uses the damped window model for stream data clustering in which the weight of each data point decreases exponentially with time t via a fading function $f(t)$:

$$f(t) = e^{-\lambda t} \quad (3)$$

where $\lambda > 0$. The most recent data will have higher weights. Note that with increasing value of λ, the importance of the historical data compared to the more recent data will decrease.

A core microcluster or c-micro-cluster, at time t for a group of close points $p_{i1}, p_{i2}, \ldots, p_{in}$ with time stamps $T_{i1}, T_{i2}, \ldots, T_{in}$ is defined as $\{\overline{\text{CF}^1}, \overline{\text{CF}^2}, w\}$, where

$$w = \sum_{j=1}^{n} f(t - T_{ij}) \quad (4)$$

$$\overline{\text{CF}^1} = \sum_{j=1}^{n} f(t - T_{ij}) p_{ij}$$

$$\overline{\text{CF}^2} = \sum_{j=1}^{n} f(t - T_{ij}) p_{ij}^2$$

These three parameters are calculated and updated at each time step. If after δt seconds, a new point (p) is merged to a microcluster then its parameter will be $\{\overline{\text{CF}^1} + p, \overline{\text{CF}^2} + p^2, w + 1\}$. Otherwise, it is given by $\{2^{-\lambda\delta t} \cdot \overline{\text{CF}^1}, 2^{-\lambda\delta t} \cdot \overline{\text{CF}^2}, 2^{-\lambda\delta t} \cdot w\}$. Two

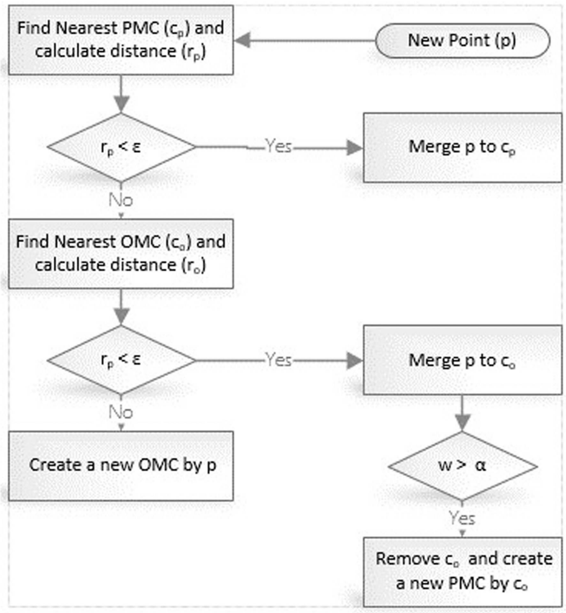

Fig. 4 *Flowchart of DenStream algorithm*

important parameters of microclusters, centre (c) and radius (r), are calculated as below:

$$c = \frac{\overline{\mathrm{CF}^1}}{w} \tag{5}$$

$$r = \sqrt{\frac{\overline{\mathrm{CF}^2}}{w} - \left(\frac{\overline{\mathrm{CF}^1}}{w}\right)^2}$$

where $r \le \epsilon$. This microcluster is divided into two groups for handling the outliers. If $w \ge \alpha$ the microcluster is a potential microcluster (PMC), and if $w < \alpha$ the microcluster is considered an outlier microcluster (OMC). The procedure for merging each new point to the nearest PMC or OMC is shown in Fig. 4.

To avoid increasing the number of microclusters, the weight of each PMC and OMC is checked at every T_p time period

$$T_p = \left\lceil \frac{1}{\lambda} \log\left(\frac{\alpha}{\alpha - 1}\right) \right\rceil \tag{6}$$

If the weight of each PMC is less than α and the weight of each OMC is less than $\beta = (2^{-\lambda(t - t_0 + T_p)} - 1)/(2^{-\lambda T_p} - 1)$ (t_0 is the creation time of the corresponding OMC), this microcluster will be deleted. By doing this, the number of PMC and OMC will always be limited [26].

This method has three unknown parameters that need to be specified:

- ϵ: radius of each microcluster. If selected too small, the number of microcluster will increase; if selected too large, clusters will merge.
- α: weight for discriminating between PMC and OMC (handling outlier). If selected too small, every microcluster will be PMC; if selected too large, every microcluster will be OMC.
- λ: extent of importance of historical data.

As mentioned, each microcluster is determined by three parameters $\left\{\overline{\mathrm{CF}^1}, \overline{\mathrm{CF}^2}, w\right\}$. These parameters are calculated and updated with features obtained from the current PD signals. In addition to current signal features, which are used to separate active sources, the charge magnitude and phase position of each PD should be known for further analysis. Because of this, the PRPD matrix as the fourth parameter is added to each microcluster and updated according to (1).

The number of PRPD matrices is equal to the number of microclusters. In a similar manner, for microclusters there are two types of micro PRPD: p-PRPD and o-PRPD associated with PMC and OMC, respectively.

Each time a call is made, p-PRPD is given to the user as output, and these matrices are then converted to zero but other information about PMC is not changed.

4.2 Offline phase

The PRPD matrices (or micro-PRPD images) generated in the online phase can be called either periodically or on demand by the user. The obtained PRPD matrices are usually much larger than the number of active sources. For this reason, the number of PRPD matrices is reduced to the number of active sources in the offline phase using clustering method. Since the offline part is usually not 'time-critical', traditional clustering algorithms can be used.

In this phase, those PRPD matrices whose number of PD pulses is <2000 are removed. Such a number of PDs is considered inadequate to form the PRPD matrix and does not have enough information about the source of PD. This may result in the removal of PD sources with low rate of occurrence, and the solution to this shortcoming is to increase the call time.

The remaining matrices are clustered based on their correlation. To do this, PRPD matrices are first considered as grayscale images and proper features are then extracted from these images. The histogram of oriented gradient method has been shown to retain excellent performance for extracting features from PRPD images. Each image divided into 4×4 cells and 9 orientation histogram bins results in a feature vector with a length of 324 [27]. The correlation coefficient between two images is calculated using the obtained features:

$$r = \frac{v \sum \boldsymbol{xy} - (\sum \boldsymbol{x})(\sum \boldsymbol{y})}{\sqrt{v(\sum x^2) - (\sum x)^2}\sqrt{v(\sum y^2) - (\sum y)^2}} \tag{7}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the feature vectors of PRPD greyscale images, v is the length of the feature vector and r is the correlation coefficient.

Agglomerative hierarchical clustering is used to cluster the PRPD images based on their correlation coefficient [28]. In this clustering method, a cluster is first assigned to each image, then the similarity parameter (r) between different clusters is computed and two most similar clusters are joined. This process is repeated until only one single cluster is left.

5 Case study

Experiments are carried out to evaluate the effectiveness of the proposed stream clustering methods. The setup is shown in Fig. 1 which includes the high voltage source, PD measurement circuit, and the transformer model with internal PD connected in parallel with external PD sources (corona and surface discharge). The PD measurements are performed under application of a stepwise varying ac voltage.

To perform the feature extraction, six bandpass filters with centre frequencies of 0.5, 2, 5, 8, 12 and 15 MHz are used. The bandwidth of each filter is 650 kHz. Using these bandpass filters, six features are extracted for each PD pulse

$$\boldsymbol{f} = [\boldsymbol{q}_{f1}\ \boldsymbol{q}_{f2}\ \boldsymbol{q}_{f3}\ \boldsymbol{q}_{f4}\ \boldsymbol{q}_{f5}\ \boldsymbol{q}_{f6}] \tag{8}$$

The obtained features are normalised against the feature obtained through the first filter

$$\boldsymbol{F} = \frac{[\boldsymbol{q}_{f2}\ \boldsymbol{q}_{f3}\ \boldsymbol{q}_{f4}\ \boldsymbol{q}_{f5}\ \boldsymbol{q}_{f6}]}{q_{f1}} \tag{9}$$

As previously stated, the presence of noise, the difference in PD rates, and the extraction of features for the separation of multi-sources will increase the amount of data to be stored in online monitoring. In this particular case study, the three PD sources and the test voltage level are such that the discharge rate of the corona

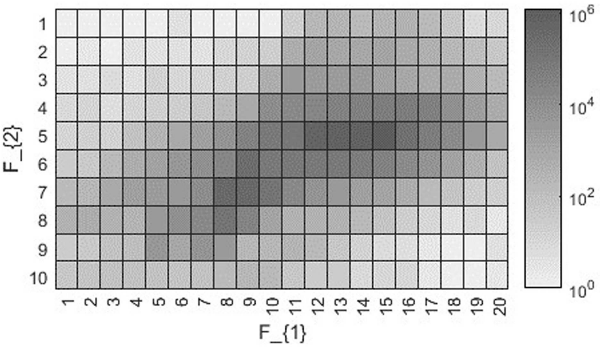

Fig. 5 *Online phase: density grids map of the three active sources (the colour in each grid represents the number of PD pulses)*

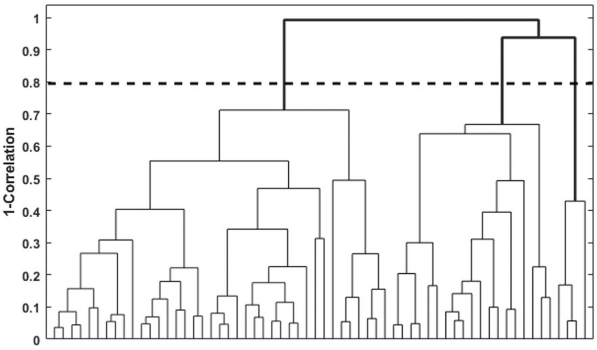

Fig. 6 *Offline phase: clustering of micro PRPD using agglomerative hierarchical clustering method*

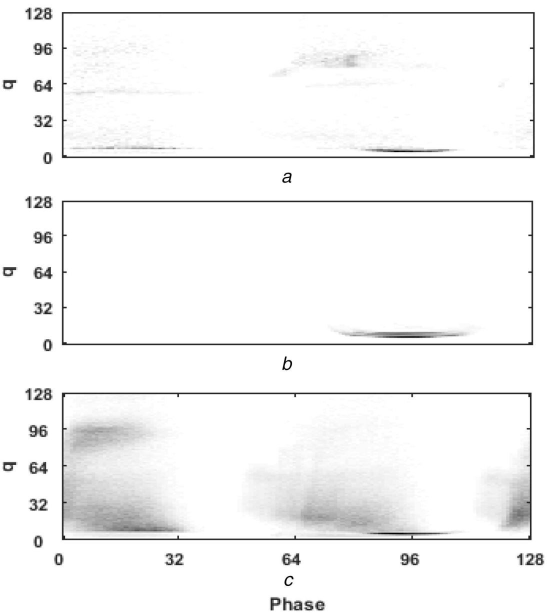

Fig. 7 *Three sub-PRPD patterns that obtained in offline phase*

and surface discharge in air (these two sources also can be considered as noise) is far more than the discharge rate in the cavity. In order to obtain an adequate number of pulses from the cavity discharge for the analysis, a very large number of discharge pulses generated by these three sources should be stored. Storing and analysing this large amount of data is very difficult. Therefore, the two stream clustering methods (density grids and DenStream) are used to address this issue.

5.1 Density grids clustering method

The first two features, namely F_1 and F_2, are used to implement stream clustering on the PD pulses. In the first step, the minimum and the maximum of these features are determined. Realising that the first feature preserves more information for the separation of sources than the second feature; the first feature was divided into 20 grids and the second feature was divided into 10 grids. Hence, the total number of microclusters or micro PRPD is 200. In the online phase, for each PD pulse, the respective micro PRPD is selected using features and updated using φ–q data.

The corresponding density grid map related to this case study is shown in Fig. 5. As can be seen in the figure, many microclusters are either empty or have very few PD pulses. Also, the sources are not completely separated which makes it difficult to categorise them in the offline phase.

In the offline phase, by first inspecting the micro-PRPDs and realising that the number of PD pulses is <2000, they are discarded and the remaining 62 micro-PRPDs are clustered using the agglomerative hierarchical clustering method. Fig. 6 shows the clustering result, where the vertical axis is $(1 - r)$. As can be seen, three clusters can be identified from which the three sub-PRPDs are shown in Fig. 7, obtained based on the correlation value of 0.2. As seen in Fig. 7, the patterns are not completely separate; the corona pattern can be seen in all three sub-PRPD patterns, and the void and surface patterns are overlapped (Figs. 7*a* and *c*).

This method is simple and very fast. It can process PD pulse data at least as fast as the PD pulse is arriving and can be easily implemented for practical application. However, the performance of this method is dependent on the effectiveness of a limited number of features (2 or 3 features) in the separation of PD sources.

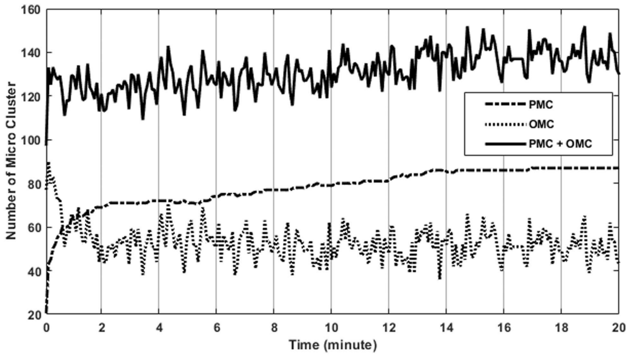

Fig. 8 *Number of PMC and OMC over time*

5.2 DenStream clustering method

To use this method, only three parameters $(\varepsilon, \alpha, \lambda)$ need to be determined. In this study, according to the features, the values of these parameters are $\varepsilon = 0.2$, $\alpha = 15$ and $\lambda = 0.25$. According to (6), the weight of PMC and OMC is checked at each T_p time period of 5 s. Fig. 8 shows the number of PMCs and OMCs over duration of 20 min. Initially, the number of PMCs is increasing rapidly, but it is gradually stabilised. However, the number of OMCs goes up and down over time. The total number of microclusters (PMC + OMC) is always <160. Fig. 9 shows the location of the PMC and OMC, in the space of the first two features, over the durations of 1 and 20 min.

Fig. 10 shows the clustering result of p-PRPD, where the vertical axis is $(1 - r)$. As can be seen, three distinct clusters can be identified in this figure. Their sub-PRPDs are shown in Fig. 11 which demonstrates the excellent performance of this algorithm in clustering active sources without storing a large amount of information.

Since this method can use a large number of features without increasing volume, it can perform well in the separation of active sources. However, the problem with this method is that the computation time increases with increasing number of PMC and OMC. This method may fail to process PD data fast enough to keep up with the new incoming PD data stream.

Fig. 12 shows how the data size changes for the three possible methods. If all PD data are saved and then analysed, the data volume increases linearly. If the density grids method is used, the data volume is constant over time. For the DenStream method, the

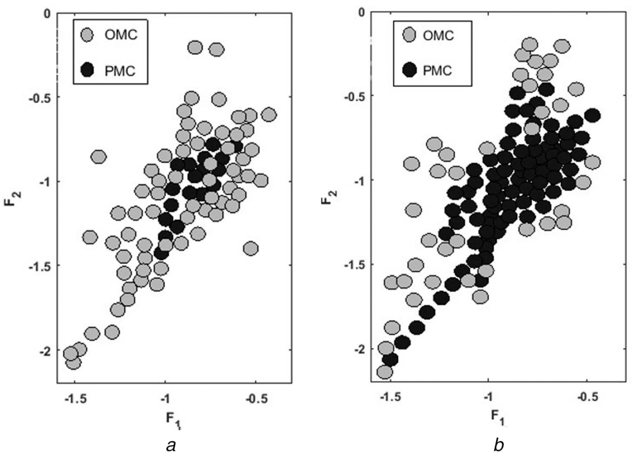

Fig. 9 *Online phase: distribution of PMC and OMC*
(a) At 1 min, ***(b)*** At 20 min

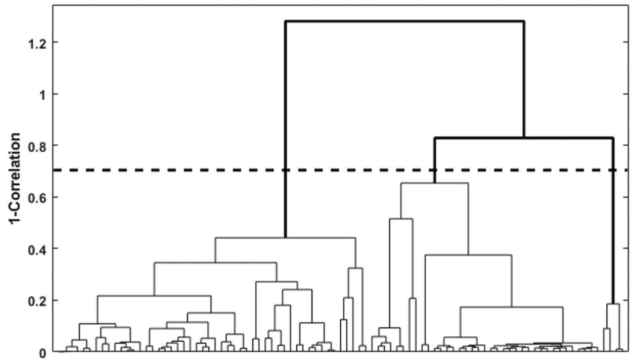

Fig. 10 *Offline phase: clustering of p-PRPDs using agglomerative hierarchical clustering method*

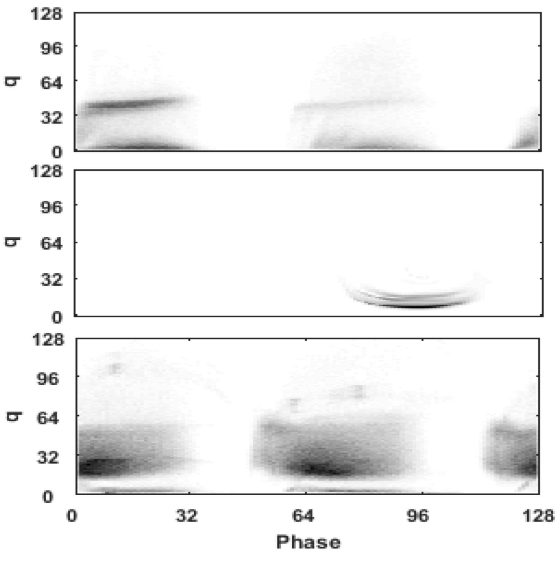

Fig. 11 *Three sub-PRPD patterns that obtained in offline phase*

volume of data shows a small change over time, but its volume is always smaller than the density grids method. In this case study, after 20 min, the volume of stored raw data is 225,000 times greater than the volume of density grids method and 350,000 times the volume of DenStream method. Also, it is easier to separate and identify multi-sources in the DenStream method as compared to the density grids method.

6 Conclusion

A new method for clustering of online PD signals of transformer is presented in this paper. For evaluating the effectiveness of the proposed stream clustering methods, a void defect in a transformer model, a corona source and a surface discharge source in air (as external noise) are connected in parallel to form a multi-source of PD with very different discharging rates.

The stream clustering algorithm consists of an online and an offline phase. The online phase maintains statistical information about the data locality in terms of micro-clusters whilst the offline phase generates clusters from the microclusters. Two stream clustering methods are used for separating the active sources without saving raw PD data and only with one access to PD signals.

The density grids method is very efficient, which is able to process PD pulse data at least as fast as the PD pulse is arriving. However, the performance of this method in the separation of PD internal sources from external sources is highly dependent on the ability of a limited number of features. On the other hand, the DenStream method can handle a large number of features and can perform very well in the separation of active sources. However, its computation time depends on the number of PMC and OMC and it may fail to process the PD signal fast enough.

In the case study presented in this paper, it is shown that the size of stored raw data is 350,000 times the size of data stored when the DenStream method employed. And this is after only 20 min recording PD data. The stream clustering method proves very useful for separating active sources (separating internal PD source from external noises) under online PD measurement without requirement of saving a large amount of data. This will make online PD diagnosis feasible.

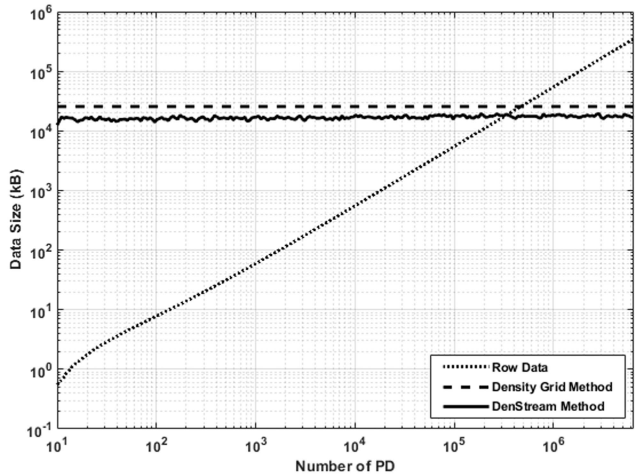

Fig. 12 *Data size changes over time for two storage methods*

7 References

[1] Jain, M.B., Srinivas, M.B., Jain, A.: 'A novel web based expert system architecture for on-line and off-line fault diagnosis and control (FDC) of transformers'. TENCON 2008–2008 IEEE Region 10 Conf., Hyderabad, India, 2008, pp. 1–5

[2] Rusov, V., Zhivodernikov, S.: 'Transformer condition monitoring'. Int. Conf. Condition Monitoring and Diagnosis, 2008, CMD 2008, Beijing, China, 2008, pp. 1012–1014

[3] Naidu, M.S.: '*High voltage engineering*' (Tata McGraw-Hill Education, New York, UK, 2013)

[4] Cavallini, A., Conti, M., Contin, A., *et al.*: 'Advanced PD inference in on-field measurements. II. Identification of defects in solid insulation systems', *IEEE Trans. Dielectr. Electr. Insul.*, 2003, **10**, (3), pp. 528–538

[5] Allahbakhshi, M., Akbari, A.: 'A method for discriminating original pulses in online partial discharge measurement', *Measurement*, 2011, **44**, (1), pp. 148–158

[6] Ardila-Rey, J.A., Martínez-Tarifa, J.M., Robles, G., *et al.*: 'Partial discharge and noise separation by means of spectral-power clustering techniques', *IEEE Trans. Dielectr. Electr. Insul.*, 2013, **20**, (4), pp. 1436–1443

[7] Darabad, V.P., Vakilian, M., Phung, B.T., *et al.*: 'An efficient diagnosis method for data mining on single PD pulses of transformer insulation defect models', *IEEE Trans. Dielectr. Electr. Insul.*, 2013, **20**, (6), pp. 2061–2072

[8] Firuzi, K., Vakilian, M.: 'A novel method for differentiating and clustering multiple partial discharge sources using S transform and bag of words feature'

[9] Lin, Y.-H.: 'Using k-means clustering and parameter weighting for partial-discharge noise suppression', *IEEE Trans. Power Deliv.*, 2011, **26**, (4), pp. 2380–2390

[10] Peng, X., Zhou, C., Hepburn, D.M., *et al.*: 'Application of K-means method to pattern recognition in on-line cable partial discharge monitoring', *IEEE Trans. Dielectr. Electr. Insul.*, 2013, **20**, (3), pp. 754–761

[11] Li, J., Liao, R., Grzybowski, S., *et al.*: 'Oil-paper aging evaluation by fuzzy clustering and factor analysis to statistical parameters of partial discharges', *IEEE Trans. Dielectr. Electr. Insul.*, 2010, **17**, (3), pp. 756–763

[12] Babnik, T., Aggarwal, R.K., Moore, P.J.: 'Principal component and hierarchical cluster analyses as applied to transformer partial discharge data with particular reference to transformer condition monitoring', *IEEE Trans. Power Deliv.*, 2008, **23**, (4), pp. 2008–2016

[13] Zhu, M.-X., Zhang, J.N., Li, Y., *et al.*: 'Partial discharge signals separation using cumulative energy function and mathematical morphology gradient', *IEEE Trans. Dielectr. Electr. Insul.*, 2016, **23**, (1), pp. 482–493

[14] Chan, J.C., Ma, H., Saha, T.K.: 'Time-frequency sparsity map on automatic partial discharge sources separation for power transformer condition assessment', *IEEE Trans. Dielectr. Electr. Insul.*, 2015, **22**, (4), pp. 2271–2283

[15] Wang, K., Li, J., Zhang, S., *et al.*: 'A hybrid algorithm based on s transform and affinity propagation clustering for separation of two simultaneously artificial partial discharge sources', *IEEE Trans. Dielectr. Electr. Insul.*, 2015, **22**, (2), pp. 1042–1060

[16] Ma, H., Chan, J.C., Saha, T.K., *et al.*: 'Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources', *IEEE Trans. Dielectr. Electr. Insul.*, 2013, **20**, (2), pp. 468–478

[17] Rostaminia, R., Saniei, M., Vakilian, M., *et al.*: 'Accurate power transformer PD pattern recognition via its model', *IET Sci. Meas. Technol.*, 2016, **10**, (7), pp. 745–753

[18] Guha, S., Meyerson, A., Mishra, N., *et al.*: 'Clustering data streams: theory and practice', *IEEE Trans. Knowl. Data Eng.*, 2003, **15**, (3), pp. 515–528

[19] Muthukrishnan, S.: 'Data streams: algorithms and applications', *Found. Trends® Theor. Comput. Sci.*, 2005, **1**, (2), pp. 117–236

[20] Andrade, H.C.M., Gedik, B., Turaga, D.S.: '*Fundamentals of stream processing: application design, systems, and analytics*' (Cambridge University Press, Cambridge, UK, 2014)

[21] Garofalakis, M., Gehrke, J., Rastogi, R.: '*Data stream management: processing high-speed data streams*' (Springer, New York, USA, 2016)

[22] Guha, S., Mishra, N.: 'Clustering data streams', in Garofalakis, M. (Ed.): '*Data stream management*', (Springer, New York, USA, 2016), pp. 169–187

[23] Gaber, M.M., Zaslavsky, A., Krishnaswamy, S.: 'Mining data streams: a review', *ACM Sigmod Rec.*, 2005, **34**, (2), pp. 18–26

[24] Aggarwal, C.C.: '*Data streams: models and algorithms*', vol. **31** (Springer Science & Business Media, New York, USA, 2007)

[25] Kim, H.G., Kim, M.H.: 'A review of window query processing for data streams', *J. Comput. Sci. Eng.*, 2013, **7**, (4), pp. 220–230

[26] Cao, F., Estert, M., Qian, W., *et al.*: 'Density-based clustering over an evolving data stream with noise'. Proc. 2006 SIAM Int. Conf. Data Mining, Hong Kong, 2006, pp. 328–339

[27] Firuzi, K., Vakilian, M.: 'Partial discharges pattern recognition of power transformer defect model by LBP & HOG features'

[28] Rokach, L., Maimon, O. (Eds.): 'Clustering methods', in '*Data mining and knowledge discovery handbook*', (Springer, New York, USA, 2005), pp. 321–352