



ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form 8 January 2015
Accepted 9 February 2015
Available online 21 February 2015

Keywords:
Partial discharge measurement
Pattern recognition
Feature extraction
Neural networks

ABSTRACT

It is well known that a correlation exist between the pattern of partial discharge (PD) behavior and the insulation quality. Since different sources of partial discharge have their own unique effects on the degradation of insulation material, it is vital to investigate the relationship between the defect type and the PD to determine the insulation quality. Numerous work had been done to classify partial discharge patterns with variable success. Past research work in partial discharge classification varies greatly in terms of classification techniques used, choice of feature extraction, denoising method, training process, artificial defects created for training purposes and performance assessment. Therefore it is necessary for a literature survey to access the state of the art development in partial discharge classification.

© 2015 Elsevier Ltd. All rights reserved.

Contents

| | |
|---|-----|
| 1. Introduction | 165 |
| 2. Background of PD | 165 |
| 2.1. PD pattern representations | 165 |
| 2.2. Online vs offline pd measurement | 166 |
| 2.3. Electrical pulse detection vs acoustic detection | 166 |
| 3. PD signal denoising | 166 |
| 4. PD feature extraction | 167 |
| 4.1. Statistical parameters | 167 |
| 4.2. Image processing | 167 |
| 4.3. Frequency and time features | 167 |
| 4.4. Weibull features of pha curves | 168 |
| 4.5. Fractal features | 169 |
| 4.6. Cross wavelet spectrum | 170 |
| 4.7. Wavelet coefficients | 170 |
| 4.8. Two pass split window (TPSW) scheme | 170 |
| 4.9. Octave frequency Cepstral coefficients (OFCC) | 171 |
| 4.10. Auto-correlation technique | 171 |
| 4.11. Chaos theory | 171 |

* Corresponding author.

| | | |
|-------|---|-----|
| 5. | PD classification methods | 172 |
| 5.1. | Back propagation neural network (bpnn) | 172 |
| 5.2. | Counter propagation networks (CPN) | 172 |
| 5.3. | Cascaded neural network (CNN) | 172 |
| 5.4. | Radial basis function neural network (RBFNN) | 173 |
| 5.5. | Extension neural network (EXNN) | 173 |
| 5.6. | Ensemble neural network | 173 |
| 5.7. | Probabilistic neural network | 174 |
| 5.8. | Hidden markov models (HMM) | 175 |
| 5.9. | Fuzzy logic based classifier | 175 |
| 5.10. | Self organizing maps (SOM) | 175 |
| 5.11. | Inductive inference algorithm | 176 |
| 5.12. | Adaptive resonance theory (ART) | 176 |
| 5.13. | Support vector machine (SVM) | 176 |
| 5.14. | Rough set theory (RST) | 177 |
| 5.15. | Cerebellar model articulation controller (CMAC) | 177 |
| 5.16. | Sparse representation classifier | 178 |
| 6. | Conclusions | 178 |
| | Acknowledgments | 179 |
| | References | 179 |

1. Introduction

Electrical insulation is a significant part in all high voltage power equipment. Failure analysis reveals that insulation failure is the root cause for more than 60% of high voltage equipment damage [1]. Therefore it is crucial to ensure that the insulation is in good condition. PD measurement has received global acceptance as an effective diagnostic tool with the capability to assess and monitor insulation systems for its integrity during manufacture and while in service [2].

PD is a type of breakdown that do not fully connect the electrodes. This can lead to serious insulation damage and considerably reduce the life span of high voltage equipment [3]. PD occurs if the local electric field is greater than the threshold value causing a partial breakdown of the surrounding medium [4]. PD has a transient nature and is characterized by pulsating currents with a duration of several nanoseconds to few microseconds [5]. PD discharge magnitude is not always proportional to the damage caused since PD of a tiny magnitude may rapidly lead to electrical tree growth, especially for high voltage cables [6]. Therefore it is very cost effective if PD activity can be detected and quantified in its early stage in order for replacement to be scheduled at a suitable time [7].

The IEC has a specific limit of PD for all power equipment. When insulation failure occurs, it is replaced with no information of the type of PD since current commercial PD detector does not provide any information on the PD source [5]. Every defect has its own unique degradation characteristic. It is feasible to utilize this special attribute to correlate the PD patterns with the defect type to determine the insulation quality [8]. PD patterns classification is a crucial criteria to assess the insulation system condition using the provided significant index of the PD severity [9,10]. Apart from that, PD classification can be used to forecast any imminent insulation failure and determine whether there is a need to replace the current insulation system [11].

After data acquisition and pre-processing, which captures, digitizes, and purifies PD signal (usually denoising), there are basically two main steps, which are feature extraction and classification [12]. Feature extraction is the process that extracts the representative attributes from raw PD data. PD contains a set of unique discriminatory features which enables them to be recognized. Hence, the first procedure in the classification process is to ponder which discriminatory features can be used and extraction method of the selected features [13,14].

The following sections of this paper are arranged as such. Section 2 describes related background information about how PD is represented and measured during practical usage. The importance of PD denoising and denoising procedure are presented in Section 3. A brief compilation of all known feature extraction methods are shown in Section 4 followed by a brief summary of all the known PD classification techniques in Section 5. Conclusion about the current state of art development of PD classification is presented in Section 6.

2. Background of PD

2.1. PD pattern representations

Time-resolved and phase-resolved data are the two major types of PD pattern representations currently used for PD related research. Phase-resolved PD data are acquired based on the AC test voltage waveform. The phase angle of the voltage is split into a specific amount of sections while the AC test voltage is maintained at a constant level. A PD detector is required to capture the individual PD signal and quantify all the pulses based on the phase angle occurrence (φ), charge magnitude (q) and the number of PD (n) over a predetermined time duration [13]. The relevant phase and amplitude pulse numbers of PRPD patterns are usually stored in matrix format for computational purposes [15]. These data are commonly known as φ - q - n or PRPD patterns.

Time-resolved data pattern has interesting benefits because individual pulse shape can be observed and there are correlation between the PD signal shape and the nature of the insulation defect, which provides aging information of the insulation system [16]. The measurement procedure of time-resolved patterns typically requires less expensive measurement system compared to phase-resolved measurements.

Phase-resolved data are more widely used in PD classification research because it is able to represent the physical process at the PD location, since individual PD pulse has a solid relation with the PRPD patterns [9].

2.2. Online vs offline pd measurement

PD measurement consists of two main approaches which are off-line and on-line PD detections. Off-line methods refer to tests where the equipment under test is de-energized from normal operation and energized by an external voltage source [17]. On-line tests are performed at operating voltage, thus the exact behavior of PD can be obtained and evaluated. The off-line PD testing has advantage such as able to determine PD extinction voltage (PDEV) and PD inception voltage (PDIV) since measurement voltage can be controlled. The advantages of on-line PD testing are PD characteristics measurable at different load conditions and tests can be carried out without causing power outage.

Currently, only off-line methods have been standardized in IEC 60270. On-line methods remain unconventional with no benchmark that one can make comparison against. However, online tests have increasingly become popular in cable PD detection in recent years. Both on-line and offline PD tests are complimentary of each other. By combining both the methods, a more valid result of cable condition can be obtained. Majority of the research related to PD classification uses offline detection method since it is convenient to conduct in a lab environment.

2.3. Electrical pulse detection vs acoustic detection

The standard PD detection systems rely on electrical voltage or current pulse detection. Electrical pulse detection equipment is commercially available and can be installed in HV labs [16]. Pulse detection method has its own advantages and disadvantages [18]. The advantages of this method are high demand for commercial applications, able to detect PD level, auto calibration supported when double sensors are used, able to analyze PD direction for cable accessories and can work together with PD localization systems. However, the disadvantages of this method are measurement equipment is not cheap and challenging to install in the field, susceptible to radio frequency (RF) interference when lack shielding, shield construction limits inductive methods and low coupling capacity reduces capacitive method's sensitivity.

Acoustic detection uses acoustic sensors to measure pressure fluctuations on the insulation surface. This offers an interesting measurement technique for PD detection [18]. The advantages for this method are electromagnetic noise immunity, non-destructive and non-intrusive, high

sensitivity sensor, frequency spectrum has high range, sensor installation not affected by shielding construction, robust mechanical strength, excellent electrical resistivity and more cost effective compared to other sensors. However, the disadvantages of this method are signal attenuation, measurement sensitivity affected by temperature, cannot detect PD level, highly complex calibration required, and limited capability when handling equipment with air insulation.

Electrical pulse detection has become the common standard in PD detection, however, in recent years, the acoustic detection has gained some popularity and quite a number of research had been done using acoustic methods [18–21].

3. PD signal denoising

Ideally, by analyzing the specific combination of PD phase distribution, pulse magnitude and changes with time, PD patterns can be classified [22]. However, during PD measurement, a difficulty encountered is caused by external noise interference, which degrades the PD measurement detection sensitivity. Noise was such a real problem in PD detection that researchers even add artificially generated noise into their PD data to evaluate their PD classification model to better represent practical situation [2]. Major interferences faced during PD measurements are caused by discrete spectral interferences, stochastic pulse shaped interferences and periodic pulse shaped interferences [23].

One way to easily disregard the presence of noise is to set a threshold to ignore signals that are less than 10% of the maximum discharge amplitude [24]. However this method is not suitable for application which requires good accuracy. This is because big threshold level ignores actual PD pulses that have small magnitudes while small threshold level mistakenly detects huge amount of noise as PD pulse [25,26].

The Mean Square Error method was used by [27] to compare the performance of 28 different types of denoising methods. It was found that wavelet based denoising has the best results. The fundamentals workings of wavelet transform can be found in [28,29]. Steps to select an optimal mother wavelet and setting automated thresholding rule are documented in [7,22]. The J criterion [30] can also be used to determine the optimum mother wavelet. For PD denoising, Daubechies (dB) was the most popular mother wavelet choice since the Daubechies wavelet and a single PD pulse shape have high similarities. After an optimal mother wavelet is determined, the maximum number of decomposition can be calculated using the J_{\max} formula [31]. Discrete wavelet transform (DWT) is preferred over continuous wavelet transform (CWT) since CWT is much more difficult to compute and produces a lot of irrelevant data [23]. Successful denoising is achieved if it has low amplitude reduction, minimum pulse shape distortion and high signal to noise ratio (SNR) [23].

PD denoising are usually done off line because online PD denoising is much more challenging due to nonzero wavelet coefficients that are higher than the PD

coefficients. A new method for online PD denoising is to raise the voltage to slightly below PDIV to record noise level of the measurement system [32]. A threshold value is calculated using the recorded noise level. Lastly wavelet transform de-noises the PD signal using the calculated threshold level.

Notable recent advancement in wavelet denoising is the introduction of second generation wavelet transform (SGWT) [33] and complex wavelet transform [34]. SGWT differs from DWT by providing interpretation of a fully spatial domain of the transform compared to the original frequency domain-based constructions. Complex wavelet transform is shift invariant and possesses greater directional selectivity while filtering multidimensional signals.

4. PD feature extraction

The purpose of feature extraction is to obtain relevant input feature from PD data to represent PD characteristics associated with a particular defect.

As mentioned in Section 2.2 phase-resolved data are more commonly used for PD classification. The PRPD or φ - q - n pattern gathered from a digital PD detector is popularly referred to as PD fingerprint [35]. Apart from collecting relevant input features, feature extraction also tries to decrease the dimension of the original data for easier processing [36,37]. Raw data contained in each of the stored patterns might be too enormous to allow its easy direct handling. A form of data reduction such as reducing the matrix size of the data is normally required [2].

This section provides a brief review on the extracted input feature used by researches for PD classification for the past few decades. Due to allowable length limitation, the mathematical derivations are only briefly mentioned.

4.1. Statistical parameters

Statistical parameters was first used by [24] in 1993 and consist of skewness, kurtosis, mean, variance, and cross correlation factor.

The mean pulse height distribution $H_{qn}(\varphi)$ is used to represent the average PD magnitude vs the phase angle φ . The number of PD vs phase angle φ is represented by the number of PD distribution $H_n(\varphi)$. Both of them can be split into two different distributions, the $H_{qn}^+(\varphi)$, $H_n^+(\varphi)$ (from the positive half cycle of the voltage) and $H_{qn}^-(\varphi)$ and $H_n^-(\varphi)$ (from the negative half cycle of the voltage).

Skewness describes the asymmetry of the $H_{qn}^+(\varphi)$, $H_n^+(\varphi)$, $H_{qn}^-(\varphi)$ and $H_n^-(\varphi)$ distribution with respect to the normal distribution. Positive skewness represents asymmetric with larger left side, zero skewness represents symmetric and negative skewness shows asymmetric with larger right side [38].

Kurtosis describes the sharpness of the $H_{qn}^+(\varphi)$, $H_n^+(\varphi)$, $H_{qn}^-(\varphi)$, and $H_n^-(\varphi)$ distribution with respect to a normal distribution. Zero kurtosis represents a normal distribution, positive kurtosis represents a sharp distribution and negative kurtosis represents a flat distribution [39].

The cross correlation factor denotes the dissimilarity in shape of $H_{qn}^+(\varphi)$ from $H_{qn}^-(\varphi)$ distributions of the positive and negative half cycle. Zero cross correlation factor means total asymmetry while cross correlation factor of 1 indicates complete shape symmetry.

Variance describes how much a group of numbers is spread out. Zero variance means all values are identical. A very detailed mathematical description of skewness, kurtosis, and cross correlation can be found in [16]. The formulas of all mentioned statistical parameters are shown in Eqs. (1)–(5). Where $f(x_i)$ is the function of interest.

$$\text{Average (mean)} \mu = \frac{\sum_{i=1}^N x_i f(x_i)}{\sum_{i=1}^N f(x_i)} \quad (1)$$

$$\text{Variance} : \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2 f(x_i)}{\sum_{i=1}^N f(x_i)} \quad (2)$$

$$\text{Skewness} : S_k = \frac{\sum_{i=1}^N (x_i - \mu)^3 f(x_i)}{\sigma^3 \sum_{i=1}^N f(x_i)} \quad (3)$$

$$\text{Kurtosis} : K_u = \frac{\sum_{i=1}^N (x_i - \mu)^4 f(x_i)}{\sigma^4 \sum_{i=1}^N f(x_i)} - 3 \quad (4)$$

$$\begin{aligned} \text{Cross correlation factor} : cc \\ = \frac{\sum_{i=1}^N x_i^+ x_i^- - \sum_{i=1}^N x_i^+ \sum_{i=1}^N x_i^- / N}{\sqrt{\left[\sum_{i=1}^N (x_i^+)^2 - \frac{(\sum_{i=1}^N x_i^+)^2}{N} \right] \left[\sum_{i=1}^N (x_i^-)^2 - \frac{(\sum_{i=1}^N x_i^-)^2}{N} \right]}} \quad (5) \end{aligned}$$

4.2. Image processing

Relevant input features can be obtained from a PRPD image with the correct image processing technique. The multi-resolution signal decomposition (MSD) can be performed on a two dimensional PRPD image. It is recommended to use 3 level MSD using dB wavelet with 56 coefficients [35]. After decomposition, a reconstruction at the third level produces 4 images as shown in Fig. 1 which are the approximate (A), horizontal (H), vertical (V), and diagonal (D) images. The H and V images represent salient features of the individual sources in a separable form while preserving the phase and magnitude positions of the individual PD pulse. Formulation of a suitable feature vector was carried out by averaging the H and V images, both in the magnitude and phase directions. Averaging the phase direction retains the magnitude information and averaging in the magnitude direction retains information in the phase direction. The final feature vector was determined by concatenating the averaged phase and magnitude vectors.

4.3. Frequency and time features

Frequency and time features of a PD signal can also be used for feature extraction [40]. If a PD signal has K samples and $s_i(t_i)$ indicates the detected sample at time t_i where each buffer of memory is normalized to zero reference, the barycenter time t_o is,

$$t_o = \frac{\sum_{i=0}^K t_i * s_i(t_i)^2}{\sum_{i=0}^K s_i(t_i)^2} \quad (6)$$

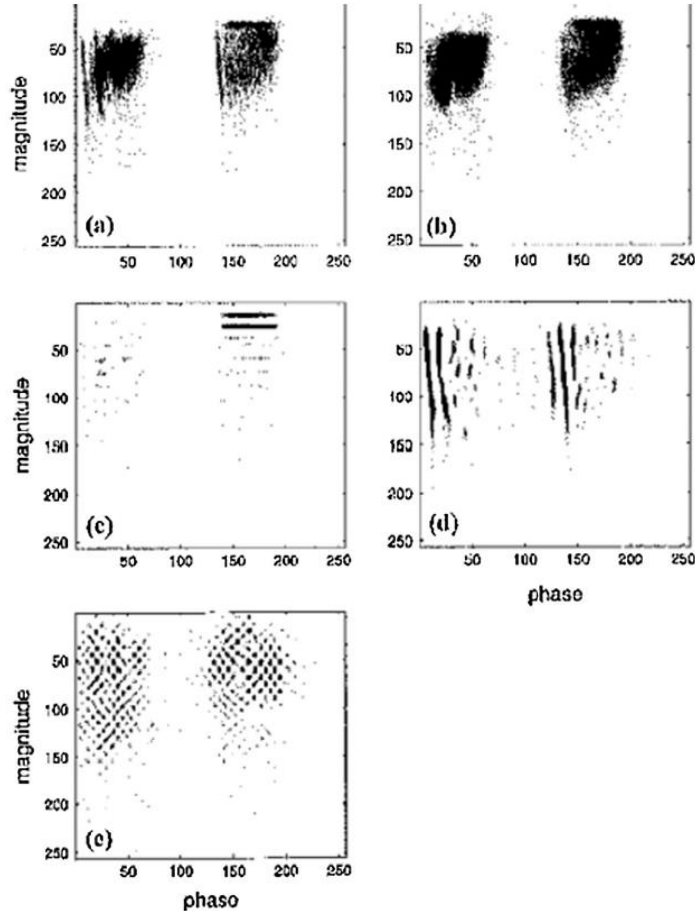


Fig. 1. Reconstruction at the third level; (a) original, (b) approximate, (c) horizontal, (d) vertical, and (e) diagonal [35].

The PD signal equivalent time length is given by

$$T^2 = \frac{\sum_0^K (t_i - t_0)^2 s_i(t_i)^2}{\sum_0^K s_i(t_i)^2} \quad (7)$$

Assuming $X_i(f_i)$ is the PD signal frequency components [40] computed using FFT, the equivalent bandwidth is

$$W^2 = \frac{\sum_0^K f_i^2 |X_i(f_i)|^2}{\sum_0^K |X_i(f_i)|^2} \quad (8)$$

These values gives a simple and concise depiction of the measured signals in a T^2 , W^2 plane. The PD pulse signals with the same shape can be amalgamated in a carefully defined areas of the T^2 , W^2 plane. Similar to other techniques of data compression, this procedure causes the PD shape information to be lost, but the frequency and time features are optimum compromise between calculation difficulty and practical requirements. Suitable time pointers permit the association of every PD pulse with its transformation in T^2 , W^2 plane. Using these pointers, once the

classification is performed, a detailed PD pulse in the detected PD pulse sequence can be recognized and a 3D figure for every PD signal cluster can be created.

4.4. Weibull features of pha curves

Weibull analysis offers a mathematical approach for pulse height analysis pattern. Suppose the probability distribution of PD pulse rate (F) vs PD magnitude (q) could be expressed by the Weibull function [15,36],

$$F(q; \alpha; \beta) = 1 - \exp \left[- \left(\frac{q}{\alpha} \right)^\beta \right]$$

Each pulse height analysis (PHA) curve is represented by the features α and β , and the PD pulse amplitude is represented by q . The features $\alpha+$, $\beta+$, $\alpha-$, $\beta-$ are extracted from the PHA curves of negative and positive. The PHA pattern is compressed using the Weibull method for digital analysis while retaining its relevant information. An easy method to calculate the capability of α and β to identify the

different types of PD. Scatter plots are created by using discharge fingerprints of every type tested at particular voltage level. For example, from Fig. 2, PD fingerprints types are labeled by markers. The left plot is the scatter of α . The summation of $\alpha+$ and $\alpha-$ represents the entire quantity of discharge and the rate of $\alpha+$ and $\alpha-$ indicates the relationship of negative and positive magnitude pulse. The right plot represents the β scatter that shows the PHA curve shape character. The three main clusters which are the slot discharge (SL) and end winding discharge (EW) are in the scatter plot. The marker of EW and SL is removed in the plot of β . Typical PD might be removed from void PD using the summation of $\alpha+$ and $\alpha-$.

4.5. Fractal features

Fractals are good for modeling complex shapes and natural phenomena, where existing mathematical

techniques are found to be insufficient. Since PD can be considered a natural phenomenon that has complex surfaces and shapes, fractals can be used to model it. The usage of fractal features in PD classification is fascinating because it represents the 3D PRPD pattern directly [41].

A 3D PRPD pattern can be represented using two fractal features, lacunarity (A) and fractal dimension (D) that are extracted using the box counting method. Since D is invariant to variation in scale, it can be utilized to gauge the roughness of the surface. A is the denseness of the fractal surface. Both D and A are functions of the box size L . The number of boxes N , of side L needed to cover a fractal set is shown in Eq. (9) where D is the fractal dimension set and K is a constant [42]:

$$N(L) = KL^{-D} \quad (9)$$

The lacunarity $A(L)$ relies on the 2nd order statistics of $p(m, L)$. It can be defined after calculating $M(L)$ and $M_2(L)$.

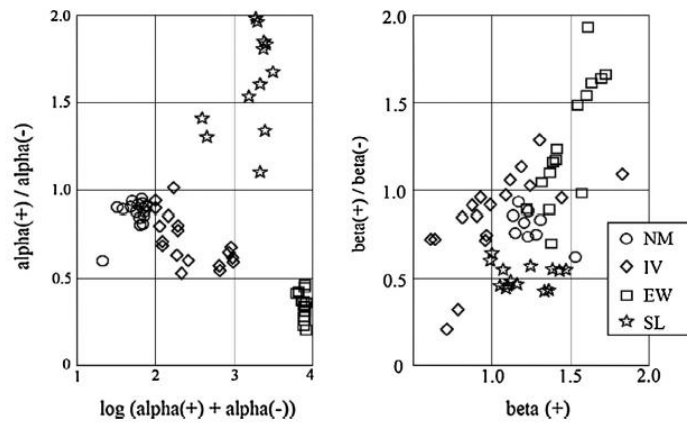


Fig. 2. Weibull features scatter plots of α and β extracted from PD fingerprints [36].

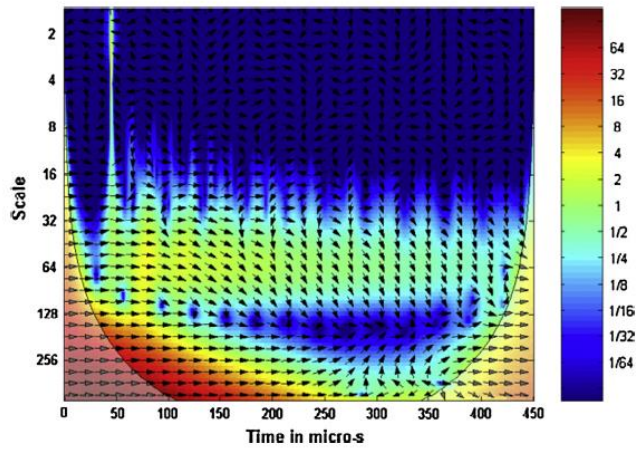


Fig. 3. Cross wavelet spectrum [50].

The formulas of $A(L)$, $M(L)$ and $M_2(L)$ is shown in Eqs. (10)–(12). Mathematical derivation of D and A can be found in [19]

$$A(L) = \frac{M^2(L) - [M(L)]^2}{[M(L)]^2} \quad (10)$$

$$M(L) = \sum_{m=1}^N mp(m, L) \quad (11)$$

$$M^2(L) = \sum_{m=1}^N m^2 p(m, L) \quad (12)$$

Fractal features was used in [43–48] as input features to classify PD patterns. In 2006, for fractal dimension approximation of PD images, an enhanced differential box-counting technique was introduced [49].

4.6. Cross wavelet spectrum

Cross wavelet transform originates from the wavelet analysis extension. The cross-wavelet spectrum shows the correlation degree among two signals in frequency–time domain. In other words, cross wavelet spectrum identify the regions in frequency–time dimension where two signals have high common power.

Fig. 3 shows a sample of cross wavelet transform spectrum [50]. In the cross wavelet spectrum, the values of $|W^{xy}|$ at different scale and time are plotted. The y-axis represents the scale, which is associated to the inverse of frequency and x-axis represents the time. The value of $|W^{xy}|$ at that time–frequency space is represented by colors based on the color-bar given beside the plot where higher power at that frequency–time point is represented by higher color values. The phase angle is represented by black arrow. Arrows pointing left means anti phase while right pointing arrow means in-phase.

The two signals cross-wavelet transform, $x(t)$ and $y(t)$ is shown in Eq. (13). The mathematical background is detailed in [51].

$$W^{xy}(s, \tau) = \frac{1}{k_\varphi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W^x(a, b) W^y\left(\frac{a}{s}, \frac{b-\tau}{s}\right) \frac{dadb}{a^2} \quad (13)$$

The 7 input features that could be extracted from the cross wavelet transform are shown in Eqs. (14)–(20) [50]. These input features are unique because they can be extracted directly from noisy signal without denoising and achieve good classification results [52].

$$F_1 = \frac{\sum_s \sum_\tau s \tau |W^{xy}(s, \tau)|}{\sum_s \sum_\tau |W^{xy}(s, \tau)|} \quad (14)$$

$$F_2 = \sqrt{\frac{\sum_s \sum_\tau s^2 \tau^2 |W^{xy}(s, \tau)|}{\sum_s \sum_\tau |W^{xy}(s, \tau)|}} \quad (15)$$

$$F_3 = \frac{\sum_s \sum_\tau |W^{xy}(s, \tau)|}{|W^{xy}(s, \tau)|_{\text{peak}}} \quad (16)$$

$$F_4 = \frac{\sum_s \sum_\tau |W^{xy}(s, \tau)|}{(s_{\max} - s_{\min}) - (\tau_{\max} - \tau_{\min})} \quad (17)$$

$$F_5 = \sqrt{\frac{\sum_s \sum_\tau (F_4 - |W^{xy}(s, \tau)|)^2}{(s_{\max} - s_{\min}) - (\tau_{\max} - \tau_{\min})}} \quad (18)$$

$$F_6 = |W^{xy}(s, \tau)|_{\text{peak}} \quad (19)$$

$$F_7 = |W^{xy}(s, \tau)|_{\tau_{\text{peak}}} \quad (20)$$

4.7. Wavelet coefficients

This is a feature extraction method based on the computation of the wavelet coefficients first four cumulants at different scales [53]. Wavelet packet transform was used to transform the PD signal into wavelet coefficients [54]. Then the k th cumulant function can be used to calculate any stochastic process for statistics of higher order. The k th dimensional function faces exponential increase in calculation difficulty when k is higher. Therefore just the first four cumulants are considered and one-dimensional slice was used in a cumulant of multidimensional by freezing the $k - 1$ indexes. In PD feature extraction, the cumulants at zero indexes were used to decrease the feature space dimensionality using statistical parameters. Each node's wavelet coefficients were treated as a stochastic event, hence the first 4 cumulants at zero time lags were computed and the nodes with the similar scale were summed to produce statistical parameters at each scale. These values can then be used as input to the classifier.

4.8. Two pass split window (TPSW) scheme

The TPSW scheme was successfully used as a feature extraction and pre-processing method in various fields such as sonar signal processing and speech recognition [55]. This method is widely used for audio processing and underwater passive sonar signals classification [56]. The spectrum of the emitted signal comprises of two spectral types which are narrowband and broadband. Tonal features extraction from mixed spectra is an important task, which is similar to the discontinuous pulsating components extraction in PD patterns. This is due to it is analogous to the radiated spectrum signal tones of sonar waves. TPSW filtering can produce local-mean estimates which is smooth along the signal although the analyzed signal has spurious spikes. This includes having a moving-average filtering throughout a long pulse of specific segment. The continuous spectrum is first estimated, before the tonal components are extracted. TPSW feature extraction scheme is proven to be able to achieved higher classification rate compared to using statistical features. The steps of the TPSW scheme are shown below [8]:

Step 1: A window centered on k bins is selected for signal $f(x)$:

$$R_k = [k - M, k - M + 1, \dots, k, k + M - 1, k + M] \quad (21)$$

The number of bins in the bus windows is $2M + 1$.

Step 2: First pass: The local mean is computed:

$$\hat{f}(k) = \frac{1}{2M+1} \sum_{i=k-M}^{k+M} f(i) \quad (22)$$

Step 3: A clipped sequence $g(k)$ is formed to avoid of bias on the local mean estimation by the tonal presence, where α is a constant and set as 0.5 and serves as a priori estimate.

$$g(k) = \begin{cases} f(k) & f(k) \leq \alpha \hat{f}(k) \\ \alpha \hat{f}(k) & f(k) > \alpha \hat{f}(k) \end{cases} \quad (23)$$

Step 4: Second pass: The continuous spectrum is obtained by determining the local mean using $g(k)$:

$$\hat{m}(k) = \frac{1}{2M+1} \sum_{i=k-M}^{k+M} g(i) \quad (24)$$

Step 5: Narrowband component $h(k)$ is computed:

$$h(k) = f(k) - \hat{m}(k) \quad (25)$$

4.9. Octave frequency Cepstral coefficients (OFCC)

OFCC is a feature comparable to the Mel frequency Cepstral coefficients [57], which is computed using the average spectral magnitude, E_{ave} , in the k th octave frequency sub-band given by

$$E_{ave}^k = \frac{1}{P} \sum_{i=l_b}^{u_b} m_{ik} \quad (26)$$

where P is the frequency bins number contain in every sub-band. It is possible to use the average energy as a substitute to the average magnitude if required. The OFCC is defined as DCT in the sub-band of the log average energy, which is computed using

$$OFCC_i = \frac{1}{\sqrt{L}} \sum_{k=0}^{L-1} \log(E_{ave}^k) \cos(\pi ik/L) \quad (27)$$

After denoising, the PD signal spectral energy is spread across multiple sub-bands of low frequency band containing the most energy. Majority of the signal energy is contained in the first some coefficients depicting the distinctive DCT packing properties. These coefficients can be treated as input features to the classifier such as Support Vector Machines (SVM) to achieve great accuracy.

4.10. Auto-correlation technique

Auto-correlation technique computational effort is lesser compared to other classification methods since it only needs to accumulate the products of corresponding sample values and time shifting [58]. Auto-correlation is a measure of resemblance of two matching waveforms when a function of a time-lag was applied on one of them. Feature extraction based on auto-correlation is a mathematical tool used to detect concealed patterns in a signal. Measured PD signal for a source location auto-correlates to itself to gain a sequence of auto-correlation. For this method to succeed, 7 features (F_1 - F_7) are computed from the discharge signals that were detected. The PD signal is denoted as $X(n)$ while the same PD signal correlates with

itself. The resultant auto-correlation sequence enables features to be extracted. The signatures of single or multiple PD source type is contained in A_{xx} along with the corresponding locations as shown by

$$A_{xx}(m) = \begin{cases} \sum_{n=0}^{N-M-1} X_{n+m} X_n & m \geq 0 \\ A_{xx}(-m) & m < 0 \end{cases} \quad (28)$$

where $m = -N \dots, 2, -1, 0, 1, 2, \dots, +N$, indicates the time shift while the subscript 'XX' signifies the auto correlated PD signal size for a specific position of the source. If N is the samples number in each of the detected discharge signals, the auto-correlation sequence has $(2N - 1)$ samples. In short, A_n is considered the n th coefficient of auto-correlation of the two discharge signals. F_1 - F_7 are extracted where F_1 is the sequence maximum, F_2 illustrates the sequence equivalent width, F_3 represents sequence centroid, F_4 describes the centroid of absolute, F_5 is RMS width, F_6 represents the sequence of mean value and F_7 is the standard deviation of sequence [59].

$$F_1 = \text{Maximum value of sequence } (A_{n \max}) \quad (29)$$

$$F_2 = \frac{\sum_{n=-N}^N n A_n}{A_{n \max}} \quad (30)$$

$$F_3 = \frac{\sum_{n=-N}^N n^2 A_n}{\sum_{n=-N}^N A_n} \quad (31)$$

$$F_4 = \frac{\sum_{n=-N}^N |n| A_n}{\sum_{n=-N}^N A_n} \quad (32)$$

$$F_5 = \sqrt{\frac{\sum_{n=-N}^N n^2 A_n}{\sum_{n=-N}^N A_n}} \quad (33)$$

$$F_6 = \frac{A_n}{(2N+1)} \quad (34)$$

$$F_7 = \sqrt{\frac{\sum_{n=-N}^N (A_n - F_6)^2}{(2N+1)}} \quad (35)$$

4.11. Chaos theory

In order to obtain chaotic characteristics from a chaos signal, it is necessary to reconstruct the phase space before calculating the reconstruction trajectory. Phase space reconstruction is a technique of analyzing system dynamic behavior by rebuilding attractors based on inadequate information. The key concept of this technique is based on the evolution of any component in a system that relies on other components which interacts with it [60]. The data of the interaction component is concealed in the change of other component. Thus, the original system can be recreated based on single component observation of the system.

Chaos exist when the biggest Lyapunov exponent is exceeds 0 [61]. The largest Lyapunov exponent λ_1 is obtained as follows [62]:

$$\lambda_1 = \frac{1}{M-1} \sum_{i=1}^{M-1} \lambda_1(i) \quad (36)$$

where $\lambda_1(i)$ is defined as

$$\lambda_1(i) = \frac{1}{jh \cdot \Delta t} \log \frac{\|X_{ip,jh} + X_{i,jh}\|}{\|X_{ip} - X_i\|} \quad (37)$$

where terms inside the modulus are the Euclidean distance. After PRPD patterns are mapped into a matrix, the λ_1 can be computed and used as input to the classifier.

5. PD classification methods

PD patterns can be classified into several different categories. Some research work attempts to classify the size of cavities based on PD patterns [12], while others are interested in the phase determination of PD in three-phase transmission lines [63]. However, majority of research are aiming to identifying the type of insulation fault that causes PD to occur. According to [39], there are 4 main types of defect that can be classify based on PD patterns are cavities or voids, surface discharges, corona and treeing. This section provides a summary of all the methods used for PD classification to date.

5.1. Back propagation neural network (bpnn)

BPNN consist of one layer of input (IL), a minimum of one hidden layer (HL) and one layer of output (OL). It had been confirmed that with two hidden layers any complex decision region can be generated [2]. A typical model of BPNN is shown in Fig. 4. Every layer is completely joined to the next layer. The primary function of the HL is to obtain PD features from different sources and pass the information to the OL. The amount of processing elements (PE) in the IL rely on the amount of PD fingerprint data. The amount of PE within the OL is dependent on the number of defects to be classified [39]. Details about the mathematical models and learning algorithm can be found in [12,39,64,65]. For PD classification purposes, at least two input features are required to ensure convergence of the BPNN during training [30].

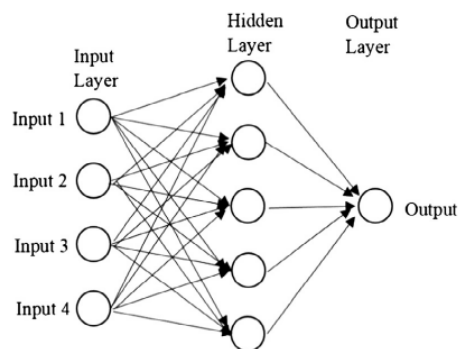


Fig. 4. Typical model of BP NN with one hidden layer.

Link to Full-Text Articles :

<http://www.sciencedirect.com/science/article/pii/S0263224115000901>

<https://www.infona.pl/resource/bwmeta1.element.elsevier-a19d514d-ae46-3879-81dc-5da4523ae1d>

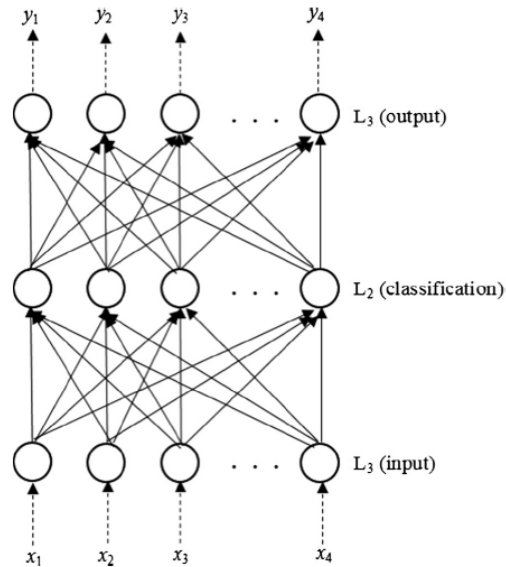


Fig. 5. Architecture of the standard FCPN.

5.2. Counter propagation networks (CPN)

CPN is used for estimating a function that is characterized by a group of desired pairs of input output, and their function of inverse. For PD classification, the forward variant only of the CPN (FCPN) is utilized. An FCPN is depicted in Fig. 5 [11]. In this figure, part of the Kohonen self-organizing map is merged with the outstar structure [66]. It works as a lookup table which can compare a pattern with the prototypes encoded in input-to-hidden weights, and chooses the most identical one. Then, the results are encoded in the hidden-to-output weights. The previous weights are trained by unsupervised competitive learning while the latter weights are trained by supervised learning. To increase the classification efficiency, the network structure is dynamically modified by using a vigilant structure. The complete working of CPN can be found in [11,67].

5.3. Cascaded neural network (CNN)

For classification of different cavity size using PD patterns, CNN arrangement has higher classification accuracy compared to single stage neural network (NN) [68]. This is because CNN uses an indexed feature that set up the highest vital input to the next stage of CNN that is influential in producing a classified attribute output from the CNN.

Fig. 6 shows an example of CNN with double outputs [68]. For the CNN system, the output of the first stage that is indexed I_d , is created to represent unity output for a 1.5 mm void and zero for a 1.0 mm void. The indexed signal I_d is passed to the second stage along with ΔV and ΔQ_m inputs as one of the inputs. After training, CNN can successfully distinguish different cavity sizes efficiently using three inputs (ΔV_1 , ΔQ_m , I_d) with 6 neurons in the HL. The