HYPERGRAPH BASED PROBABILISTIC NEURAL NETWORK FOR PARTIAL DISCHARGE PATTERN CLASSIFICATION

S. Venkatesh[1*], S. Gopal[2] and K. Kannan[1]
[1]SASTRA University, Thanjavur, Tamil Nadu, India
[2]W.S. Test Systems, 27th km Bellary Road, Bangalore, India
*Email: <venkatsri73in@gmail.com>

Abstract: Researchers have attempted classification of Partial Discharge (PD) patterns utilizing a gamut of Neural Network (NN) paradigms, supervised or unsupervised, in addition to various feature extraction techniques without compromising aspects such as trainable parts of NN, outliers etc. Hence, a trade-off between computational complexity and data compaction is a major aspect in classification of large dataset PD patterns. Recently, salient properties of Hyper Graph (HG) have been exploited by researchers for designing compact algorithms for pre-processing data in various engineering applications due to its inherent strength in representing data, based on topological as well as geometrical aspects. This research presents a novel approach of utilizing a HG model for developing cluster/ center selection of PD patterns for subsequent classification using Probabilistic Neural Network (PNN). PD data represented by a union of hyper edges that describe the relationship between φ-q-n and vertices are called HG representation of patterns. Helly property of HG is invoked for extracting features for classification of PD sources. In order to determine the ability of the proposed scheme, exhaustive studies are carried out on laboratory models. Extensive analysis is carried out to ascertain the classification capability of HG based PNN in comparison with traditional pre-processing techniques.

1 INTRODUCTION

Among various non-pervasive techniques utilized to diagnose the electrical insulation system of a power apparatus, Partial Discharge (PD) measurement has emerged as an indispensable and a vital tool since it forms an integral part of testing and acceptance criteria of most statutory and international standards. Recently, there has been a major spurt in utility of PD measurement and diagnosis among researchers and personnel handling power utilities, due to the advent of compact and advanced digital hardware with increased memory and enhanced computational capability of processors in addition to availability of flexibly designed associated data acquisition systems. The identification of the source of PD is one of the preliminary yet vital steps in the diagnosis of the insulation system, since each source of PD produces a unique mechanism of discharge pattern that affects the reliability of the insulation system. However, the inability to obtain models that provide an index of validity of the recognition/ identification methodology has thus necessitated intensive studies pertaining to identification of source of PD.

Further, the focus of researchers has shifted to identification of defects in insulation system due to multiple PD sources [1-2] since it is most often encountered during practical PD measurements. Identification of multiple PD sources becomes increasingly difficult with the degree of overlap. Notwithstanding, problems associated with classification of complex overlapped patterns practical insulation systems, ill-conditioned data and complex variations in patterns due to varying applied voltages continue to pose considerable difficulty [3]. A variety of divergent yet important tools such as Mixed Weibull Function, Contour Mapping, Neural Networks (NN), Wavelet Transformation etc have been attempted in identifying partially and fully overlapped patterns with reasonable level of success. Incidentally, a gamut of NN paradigms, falling under the category of supervised or unsupervised learning techniques, in addition to various feature extraction techniques have been attempted by researchers for PD pattern classification. It is obvious that this needs to be achieved without compromise on essential aspects of a NN, its trainable parts, effect of outliers, over-learning etc. Hence, a trade-off between computational complexity and data compaction is a major task leading to difficulties in classification of large database patterns obtained from modern digital PD acquisition systems.

Hyper Graph (HG) theory and its salient properties have been exploited by researchers for designing computationally compact algorithm for pre-processing data in various engineering applications such as image processing, bio-informatics etc [4] due to its inherent strength in representing data based on topological as well as geometrical aspects while most other algorithms are topology based only. This research work attempts a novel approach of utilizing a HG model for developing feature extraction technique which in addition serves as a center selection technique for obtaining a frugal set of representative features that describe the source of PD for subsequent training by a Probabilistic Neural Network (PNN) for single source PD classification. Vertices of HG represent φ-q-n, while a hyper edge represents a group of vertices of patterns pertaining to a PD source. Thus data are represented by a union of hyper edges. Vertices together with hyper edges are called HG representations of PD data. Pair-wise intersecting hyper edges are generated by exploiting the relationship between φ-q-n. Helly property of HG is invoked for extracting features of PD data for classification of PD sources by PNN. In order to determine the ability of the proposed scheme in identifying the sources of PD, extensive analysis is carried out to ascertain classification capability of the novel HG based PNN.

2 PD MEASUREMENT, PATTERN REPRESENTATION & PREPROCESSING SCHEMES

Though PD pulses are characterized by their magnitude, rise time, recurrence rate, phase relationship of occurrence, time interval between successive pulses, discharge inception, extinction voltage etc., from the view point of pattern recognition, the most important representation of features of PD which are also displayed by most modern digital systems is the phase angle of occurrence of PD pulses (φ), magnitude of the apparent charge during discharge (q) and the discharge rate (n). The depiction of PD pulse patterns as a three dimensional representation (φ–q-n patterns) has made it possible to carry out analysis based on two broad categories namely the phase resolved and time resolved analysis. Phase resolved PD studies provide information regarding pulse distribution [5] for a fixed time base and hence augurs well for pattern classification and discrimination of PD sources while information regarding pulse shapes, relationship between the nature of defect and the shape of pulses is facilitated by the time resolved approach [5].

The raw PD data is pre-processed in order to ensure compactness without compromising on conserving the details of the characteristic input feature vector. The goal of utilizing a range of methods of pre-processing is to ascertain the performance of the proposed NN so that important decisions may be taken on the role played by the various key parameters of a NN such as smoothing parameter, curse of dimensionality, mean value of the probability density estimate etc. In this context it is pertinent to note that a wide variety of divergent pre-processing approaches have been utilized for the task of PD pattern classification such as statistical measures, pulse characteristic parameters, image processing parameters, signal processing based transformation parameters etc [6-7]. Statistical measures include computation of statistical moments (skewness and kurtosis), measures based on dispersion (range, standard deviation, variance, quartile deviation etc), measures of central tendency (arithmetic mean, median, moving average etc), cross-correlation, discharge asymmetry etc. Pulse characteristic tools are usually used in time resolved PD studies which include parameters such as pulse rise time, decay time, pulse width, repetition rate, quadratic rate, peak discharge magnitude etc. In the case of signal processing based tools since the applied test voltage is periodic, the feature vectors comprise average values of the spectral components in the frequency domain. Since the nature of the PD database acquired is large and possibly ill-conditioned, statistical measures have been utilized for pre-processing raw PD database for subsequent feature extraction.

The input data presented to PNN is based on phase window technique wherein statistical operators namely: 1. Measures based on Maximum values of q (10º and 30º); 2. Measures based on Minimum values of q (10º and 30º); 3. Measures based on Central Tendency (10º and 30º); and 4. Measures based on types of mean values i.e. arithmetic mean (AM), harmonic mean (HM), geometric mean (GM) and root mean (RM) using the inequality relationship HM) ≤ Geometric Mean GM ≤ Arithmetic Mean AM ≤ RM are utilized to ascertain the capability of the proposed HG-PNN classifier in identifying sources of PD patterns.

3 HYPER GRAPH (HG) FOR FEATURE EXTRACTION CENTER SELECTION IN PD PATTERN CLASSIFICATION

The primary role of a feature selection is to distinguish features from a set of candidates, while that of feature extraction is to utilize transformation techniques to generate novel features from raw data. It is hence evident that both are vital for effective clustering and center selection of a large dataset. A wide range of clustering and center selection algorithms have been utilized by researchers in diverse engineering applications which fall under eight major categories. These are based on similarity and sequence similarity measures, hierarchy, square error measures, mixture density estimation, combinatorial search, kernel and graph theory. Hence, it is apparent that the choice of the appropriate type of clustering/ center selection technique would play a vital role in handling the classification of large dataset PD.

3.1 HG Preliminaries

Graph theoretic representation of data utilizes binary relations which may not comprehensively represent structural properties of temporal data, the nature association being binary neighbourhood. HG on the other hand deals with finite combinatorial sets and has the ability to capture both topology and geometrical relationship among data. A HG [8] 'H' is a pair (X, ξ) consisting of a non- empty set X together with a family $\bigcup_{i \in I} E_i = X$, I= {1, 2, … n}, $n \in \mathbb{N}$. Figure 1 shows a generic HG representation.

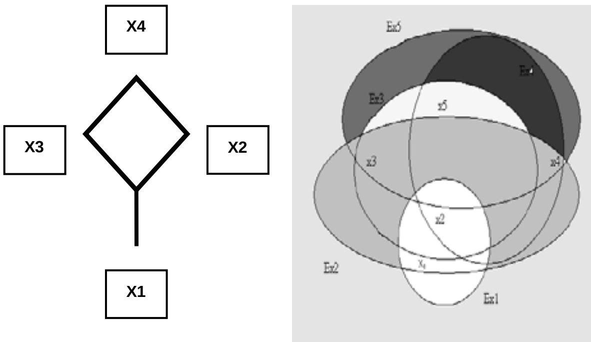

Figure 1: Generic representation of Graph and HG

An important structure that can be studied in a HG is the notion of an intersecting family. An intersecting family of hyper edges of a HG 'H' is a family of edges of H which have pair-wise non empty intersections. There are two types of interesting families: 1. intersecting families with an empty intersection and 2. intersecting families with a non-empty intersection. A HG has the Helly property if each family of pair- wise intersecting hyper-edges has a non-empty intersection (that is they belong to a star). Figure 2, represents two types of intersecting hyper-edges.

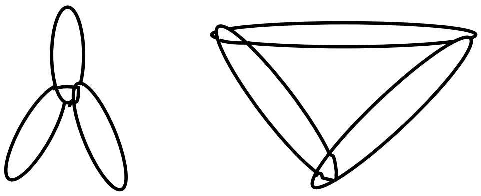

Figure 2A and 2B: Representation of Pair-wise intersecting hyper-edges with (A) non-empty and (B) empty intersection

Several researchers in allied fields [9-10] of engineering have utilized various properties of HG such as Helly, transversal, mosaic, conformal etc. for obtaining segmentation and clustering algorithms pertaining to a diverse set of applications. The neighbourhood HG representation utilizes the Helly which play a vital role in identifying homogeneous regions in the data. These homogeneous regions form the main aspect for developing segmentation and clustering algorithms.

3.2 HG based center-selection algorithm for PD pattern classification

To carry out an exhaustive analysis to identify various sources of PD in the insulation system, the pre-processed data obtained as discussed in Section 2 and 3.1 is represented as $V_i = (\varphi_i, q_i, n_i)$, i = 1,2, 3,4...m where 'm' is the number of vertices of the data per cycle. The data is grouped in terms of feature vectors which act as the best representatives of entire data base. Hence if pair-wise intersecting edges are created from the entire data base, Helly property of HG can be invoked to find the common intersection which in turn provides the feature vectors that represent the centers of a particular set of data pertaining to the source of PD. Hence, a minimum distance metric scheme (Euclidean) is developed to obtain the nearest among various intersections of the intra-cluster and inter-cluster dataset so as to obtain the optimal set of common intersection vectors that serve as the centers representing the dataset. These feature vectors are taken as training vectors of the PNN.

4 PROBABILISTIC NEURAL NETWORK (PNN) FOR PD PATTERN RECOGNITION

PNN [11] is a model based on competitive learning with a 'winner takes all attitude. The standard version of PNN does not have feedback paths and hence do not essentially comprise a separate training phase. Hence, all the input vectors that are given as input to this NN forms a part of the patterns/ exemplars that themselves represent the weight vectors. The development of PNN is based on the concept of obtaining multivariate probabilities using the Parzen window. It is a classifier version, which merges the Bayesian strategy for decision with a non-parametric estimator for obtaining the probability density function (PDF). The PNN network architecture comprises an input layer, two hidden layers (one each for exemplar/ pattern and a class/ summation layers) and an output layer. The first hidden layer forms a product of the weight vector and the sample for classification, where the weights entering a node are from the input sample. The product passes through the activation function $exp\ [(x^T w_{ki}-1) / \sigma^2]$. The second hidden layer consists of one summation unit for each class. Each summation unit (node) receives the output from the pattern nodes associated with a given class given by $\Sigma_{i=I}^{Nk} exp\ [(x^T w_{ki}-1)/ \sigma^2]$. The output layer has as many neurons as the number of data classes considered. The output nodes are binary neurons that produce the classification decision based on $\Sigma_{i=I}^{Nk}\ exp\ [(x^T w_{ki}-1)/ \sigma^2] > \Sigma_{i=I}^{Nj}\ exp\ [(x^T w_{ki}-1)/ \sigma^2]$. The pattern unit requires normalization of the input and exemplar vectors to unit length. Figure 3 shows the architecture of the standard version of PNN.

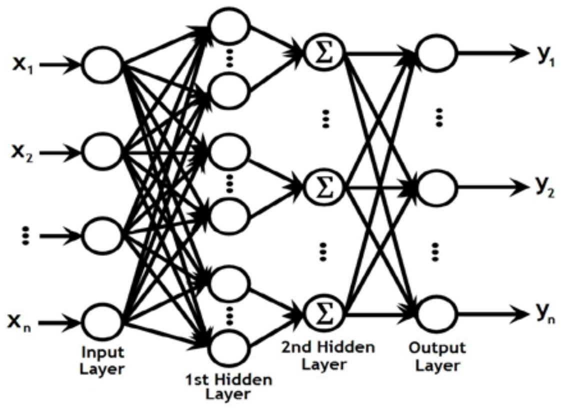

Figure 3: PNN Architecture

5 PD TESTING AND DATA ACQUISITION

5.1 PD Test Setup

PD pattern recognition has been carried out using a W.S Test Systems make (Model No.: DTM-D) Digital PD Measurement System suitable for measuring PD in the range 2-5000pC which comprises a built-in oscilloscope (TDS 2002B) and a narrow-band tuneable adjustable filter-insert (Model: DFT-1) whose center frequency is variable from 600 kHz- 2400 kHz at a bandwidth of 9 kHz. The pulses are tapped from the analogue output terminal and displayed on the built-in oscilloscope. The measured partial discharge intensity is displayed as magnitude of discharge (apparent charge) in accordance with IEC 60270. In addition, the PD measurement and acquisition system is equipped with window gating facility to enable masking of unwanted background noise during measurement. Hence, the direct detection test setup as recommended in IEC is utilized. To enhance the transfer characteristics of the test circuit a 1nF coupling capacitor is also added to the test setup in line with the recommendations of IEC 60270.

PD Gold is interfaced with the PD measurement system to acquire the PD patterns wherein the software detects PD pulse of power cycle so as to display the pulses in sinusoidal or elliptical time base selectable in auto or manual mode. In manual mode, the user may record the data for a period of 5-10 minutes which is acquired from a minimum of 240 to a maximum of 750 waveforms per channel. Figure 4 depicts a typical screenshot of PD in elliptical time base.

5.2 Artificially Simulated Laboratory Models for PD Pattern Classification

Four types of laboratory models have been fabricated to simulate distinct classes of single and multiple PD sources namely electrode bounded cavity, air-corona, oil-corona and electrode bounded cavity with air-corona which in addition would serve as a validation method to replicate the representative reference patterns recommended in [12]. Internal discharge is simulated by an electrode bounded cavity of dimension 1mm diameter and 1.5mm depth on 12mm thick Poly Methyl Metha Acrylite (PMMA) commonly referred to as perspex of diameter 80mm. Air-corona discharge is replicated by an electrode of apex angle 85º attached to the high voltage (HV) terminal. Corona discharge in oil is produced with a sharp point immersed in transformer oil. Electrode bounded cavity with air-corona is produced by inserting a needle configuration at the HV electrode. Table 1 shows the patterns acquired for large dataset from various sources of PD.

Table 1: Multiple source PD pattern recognition studies for varying applied voltages

Type of PD	Discharge Inception Voltage (kV)	Applied Voltage (kV)	Total Patterns
Electrode Bounded Cavity	8.01	8.63 10.4 12	120
Air-Corona	6.4	7.35 9.48 11.49	120
Oil-Corona	8.9	15.04 17.36 20.08	120
Electrode Bounded Cavity with Air-Corona	5.8	6.9 8 10	120

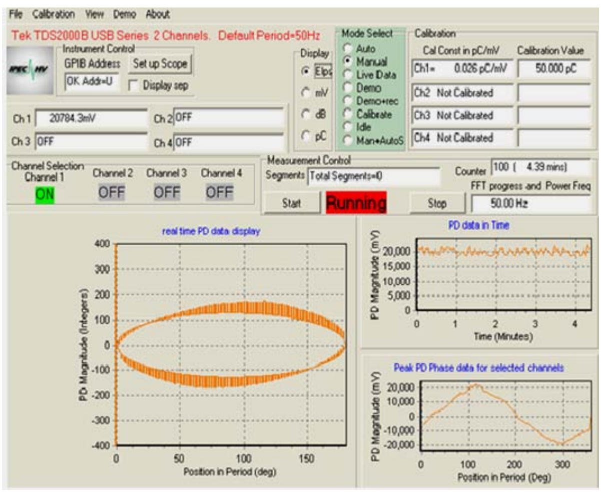

Figure 4: Typical snapshot of patterns acquired for Multiple Source PD

6 OBSERVATIONS AND ANALYSIS

Based on the exhaustive studies conducted as indicated in Table 1, 30 sets each of varying applied voltage pertaining to the source of PD was taken up for center selection using the HG algorithm. The selected optimal centers obtained from HG based clustering/ center selection algorithm was used as exemplars that represent the training/ exemplar data of the PNN.

6.1 Role of HG based centers in classifying PD sources

It is evident from Table 2 and Table 3 that though the HG-PNN performed exceptionally, the number of centers obtained by the HG algorithm is considerably high as compared to previous studies carried out by the authors of this research using density estimation based clustering/ center selection algorithm [13]. Yet, this aspect may be attributed to the utility of only one of the properties of HG (Helly properties). However, since the objective of the research is to ascertain primarily the ability of the HG algorithm in performing as a center selection technique, other more salient properties of HG such as transversal, conformal, mosaic etc have not been attempted.

Table 2: Number of Optimal Centers obtained from HG algorithm for PD Pattern Classification

Type of Feature	Number of Optimal Centers from HG Algorithm			
	Electrode bounded cavity	**Air Corona**	**Oil Corona**	**Multiple Source**
Φ-q_{max}-n (30º)	8kV-26 10kV-13 12kV-7	7kV-17 9kV-15 11kV-8	15.04kV-18 17.36kV- 17 20.08kV- 14	6.9kV- 3 8kV- 10 10kV- 17
Φ-q_{max}-n (10º)	8kV- 6 10kV-16 12kV-10	7kV-8 9kV-12 11kV-10	15.04kV-15 17.36kV- 16 20.08kV- 14	6.9kV- 7 8kV- 12 10kV- 17
Φ-q_{min}-n (10º)	8kV-12 10kV-18 12kV-8	7kV-16 9kV-18 11kV-17	15.04kV-12 17.36kV- 13 20.08kV- 15	6.9kV- 18 8kV- 21 10kV- 17
AM-GM-HM-RM (10º)	8kV-9 10kV-26 12kV-9	7kV-15 9kV-18 11kV-17	15.04kV-17 17.36kV- 17 20.08kV- 19	6.9kV- 18 8kV- 19 10kV- 15

Table 3: Classification Capability for multiple source PD patterns for varying applied voltages

Feature Extraction Scheme	Phase Window	No. of Tuples	Training Patterns	Classification (%)
Measures based on Maximum Value	Φ-q_{max}-n (30º)	36	165	98
Measures based on Maximum Value	Φ-q_{max}-n (10º)	36	170	96.67
Measures based on Minimum Value	Φ-q_{min}-n (10º)	36	188	94.3
Measures based on Mean	AM-GM-HM-RM (10º)	36	190	91.3

It is also obvious from Table 2 that for measures based on φ-q_{max}-n (30º) the number of centers obtained were much higher than the number of centers achieved by HG algorithm for measures based on φ-q_{min}-n (10º). It is interesting to note that the classification capability of measures with 30º window was much improved than that of classification based on 10º. However, it is also worth noting that this was achieved at the cost of more number of centers as observed in Table 1.

6.2 Capability of HG-PNN in classifying multiple-source PD patterns

It is also observed that the HG-PNN classifier serves as a substantially good center selection/ clustering algorithm though only a modest/ less frugal set of centers were obtained for classification. Table 2 clearly elucidates the novel method of utilizing HG as a clustering algorithm in PD pattern recognition.

It is also essential to note that in the entire studies the best classification was obtained for values of the smoothing parameter ranging from 20-30. This aspect clearly delineates the fact that the separation of class boundaries is much wider (less sharper) than the previous studies carried out by the authors [14] of this research on similar set of testing of multiple source PD thus providing an index of good set of centers that represent the class of PD.

6.3 Role played by Dimensionality of the selected HG algorithm based centers

It is also observed from studies and as indicated in Table 1 and Table 2 that the number of clusters/ centers that essentially describe the source of PD is dependent on the dimensionality of the HG centers. It is evident that the classification capability is enhanced with more number of representative centers while a slightly inferior classification rate is obtained for a larger dimensionality (tuple) though with substantially larger number of centers. Though 'curse of dimensionality' is a vital aspect in designing computationally effective clustering algorithms, the nature of centers obtained as indicated in Section 6.1 provides a much broader value of the smoothing parameter, thus circumventing the stated aspect.

7 INFERENCES AND CONCLUSIONS

It is evident from the studies that HG based center selection/ clustering algorithm provides an exciting and a viable option for obtaining reasonably parsimonious set centers that describe the class of PD. Though the properties of the HG algorithm was utilized only to cluster and classify the PD patterns in this research, this scheme provides an exciting opportunity to correlate the relationship/ association of PD pulses in terms of geometric aspects also. This research aspect is presently ongoing.

Since much larger sets of representative centers are observed during this study, more appropriate properties of HG such as transversal, conformal, mosaic etc can be attempted to further validate the approach.

Though validation has been carried out to ascertain the veracity of the database taken up for study using the 'Partial Test Set' and 'One-Hold-One-Out' methodologies, it is pertinent to note that further research and cross- validation is essential for obtaining tangible conclusions.

8 ACKNOWLEDGMENTS

This research was supported by the Research and Modernization Fund (RMF) Grant, constituted by SASTRA University. The first author is extremely grateful to Prof. R. Sethuraman, Vice-Chancellor, SASTRA University, Dr. S. Swaminathan, Dean-Sponsored Research and Dr. S. Vaidhyasubramaniam, Dean- P & D., SASTRA University for awarding the grant and for the unstinted support and motivation extended to him during the course of the project.

9 REFERENCES

[1] E. M. Lalitha, L. Satish: "Wavelet analysis for classification of multi-source PD patterns", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 7, No. 1, pp. 40–47,2000

[2] June-Ho Lee, Tatsuki Okamoto, Chin Woo Ti: "Classification of PD Patterns from Multiple Defects", Proceeding of the 6th International Conference on Properties and Applications of Dielectric Materials, June 21-26, China, Vol.1, pp. 463- 465, 2000

[3] M.G. Danikas, A.D. Karlis: "On the Use of Neural Networks in Recognizing Sources of Partial Discharges in Electrical Machine Insulation: A Short Review", International Review of Electrical Engineering (I.R.E.E.), Vol.1, No.2, pp. 277-285, 2006

[4] Ze Tian, TaeHyun Hwang and Rui Kuang: "A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge", Bioinformatics, Vol.25, No. 21, pp. 2831-2838, 2009

[5] E. Gulski, F.H. Kreuger: "Computer-aided recognition of discharge sources", IEEE Trans. on Electrical Insulation, Vol. 2, No.1, pp.82–92,1992

[6] R. Candela, G. Mirelli, R. Schifani: "PD Recognition by means of Statistical and Fractal Parameters and a Neural Network", IEEE Trans. on Dielectrics and Electrical Insulation. Vol.7, No. 1, pp. 87-94, 2000

[7] N.C. Sahoo, M.M.A. Salama, R. Bartinkas: "Trends in Partial Discharge Pattern Classification: A Survey", IEEE Trans. on Dielectrics and Electrical Insulation, Vol.12, No.2, pp. 248-264, 2005

[8] A. Bretto: "Introduction to Hypergraph Theory and Its Use in Engineering and Image Processing", Advances in Imaging and Electron Physics, Vol. 131, Chapter 1, Elsevier Academic Press, 2004

[9] D. Zhou, et al.:"Learning with hypergraphs: "clustering, classification and embedding", Advances in Neural Information Processing Systems (NIPS), MIT Press, Cambridge, MA, pp. 1601–1608, 2006

[10] Jong-Sheng Cherng, Mei-Jung Lo: "A hypergraph based clustering algorithm for spatial data sets", Proc. of IEEE International Conference on Data Mining (ICDM 2001), San Jose, USA, pp.83-90, 2001

[11] D.F. Specht;"Probabilistic neural networks for classification, mapping or associative memory", IEEE International Conference on Neural Networks, No.1, Vol.1,pp.525–532, 1988

[12] CIGRE Working Group Report 21.03:"Recognition of discharges", Electra No. 11, pp. 61–98, 1969

[13] S.Venkatesh, S.Gopal: "Robust Heteroscedastic Probabilistic Neural Network for multiple source partial discharge pattern recognition- Significance of outliers on classification capability", International Journal on Expert Systems with Applications, 38, 11501- 11514, 2011

[14] S. Venkatesh, S. Gopal, P.S. Srinivasan, B. Karthikeyan: "Identification of multiple source partial discharge patterns using Bayesian classifier based neural networks - A comparison of supervised and unsupervised learning techniques", Proc. XV International Symposium of High Voltage Engineering (ISH 2007), Ljubljana, Slovenia, T5-407 pp.198, 2007