Competitive Learning for Self Organizing Maps used in Classification of Partial Discharge

Rubén Jaramillo-Vacio[1, 2], Alberto Ochoa-Zezzatti[3], Julio Ponce[4]

[1]Comisión Federal de Electricidad-Laboratorio de Pruebas a Equipos y Materiales (LAPEM).
[2]Centro de Innovación Aplicada en Tecnologías Competitivas (CIATEC).
[3]Universidad Autónoma de Ciudad Juárez.
[4]Universidad Autónoma de Aguascalientes
ruben.jaramillo@cfe.gob.mx

Abstract. This paper different competitive learning algorithms for Self Organizing Map (SOM) are experimentally examined, the characterization of the obtainable results in terms of quality of SOM. The competitive learning algorithms showed to SOM algorithm are Winner-takes-all, Frequency Sensitive Competitive Learning and Rival Penalized Competitive Learning. As a case study: the performance in classification of partial discharge on power cables.

1. Introduction

Competitive learning is an efficient tool for Self Organizing Maps, widely applied in variety of signal processing problems such as classification, data compression, etc.

In the field of data analysis two terms frequently encountered are supervised and unsupervised clustering methodologies. While supervised methods mostly deal with training classifiers for known symptoms, unsupervised clustering provides exploratory techniques for finding hidden patterns in data. With the huge volumes of data being generated from the different systems everyday, what makes a system intelligent is its ability to analyze the data for efficient decision-making based on known or new cluster discovery. The partial discharge (PD) is a common phenomenon which occurs in insulation of high voltage, this definition is given in [1]. In general, the partial discharges are in consequence of local stress in the insulation or on the surface of the insulation.

The typical competitive learning algorithm k-means (or called hard c-means) clustering is a batch algorithm for designing a vector quantizer, which is a mapping of input vectors to one of c predetermined codevectors (also called codebooks) [2]. Fuzzy c-means (FCM) clustering is a fuzzy extension of hard c-means clustering. The FCM and its varieties have been widely studied and applied in various areas [3-5]. During the last fifteen years there have been developed new advanced algorithms that eliminate the "dead units" problem and perform clustering without predicting the exact cluster number, as for example: the frequency competitive algorithm (FSCL) [6], the

M.A. Cruz-Chávez (Ed): CICos 2011, ISBN. 978-607-00-5091-6. pp. 349 – 358, 2011

incremental k-means algorithm, the rival penalizing competitive algorithm (RPCL) [7].

We evaluate the performance of algorithms in which competitive learning is applied of partial discharge dataset, quantization error, topological error and time in seconds per training epoch. The result from classification of PD shows that Winner-takes-all (WTA) has better performance than Frequency Sensitive Competitive Learning (FSCL) and Rival Penalized Competitive Learning (RPCL).

Table 1 show a concentration of researchers who worked on the feature extraction, recognition and classification of PD, as well as the different artificial intelligent tools and the constraint to utilize these methods.

Table 1. Classification and diagnosis in pd using data mining tools.

Authors	Tool and Objective	Constraints
Mozroua *et al* [8] Kravida [9]	Tool: Supervised Neural Networks. Objective: Recognition between different sources formed of cylindrical cavities.	Recognition of different sources in the same sample
Kim *et al* [10]	Tool: Fuzzy-Neural Networks. Objective: Comparison between Back Propagation Neural Network and Fuzzy-Neural Networks	Performance in the case of multiple discharges and including defects and noises.
Ri-Cheng *et al* [11]	Tool: Particle Swarm Optimization Objective: Localization of PD in the power transformer.	On site application should improve performance.
Chang *et al* [12]	Tool: Self Organizing Map (SOM). Objective: PD pattern recognition and classification.	Quality and optimization structure of SOM.
Fadilah Ab Aziz *et al*[13]	Tool: Support Vector Machine (SVM). Objective: Feature Selection and PD classification.	SVM is not reliable for small dataset.
Hirose *et al* [14]	Tool: Decision Tree Objective: Feature Extraction and PD classification	The allocation rules are sensitive to small perturbations in the dataset (Instability)

This discovered knowledge then forms the basis of two separate decision support systems for the condition assessment/defect classification of these respective plant items. In this paper is shown a comparative of competitive learning algorithms to classify measured PD activities into underlaying insulation defects or source that generate PD's using Self Organizing Maps (SOM). Multidimensional scaling (MDS) is a nonlinear feature extraction technique [15]; it aims to represent a multidimensional dataset in two or three dimensions such that the distance matrix in the original k-dimensional feature space is preserved as faithfully as possible in the projected space. The SOM, or Kohonen Map [16], can also be used for nonlinear feature extraction. It should be emphasized that the goal here is not to find an optimal clustering for the data but to get good insight into the cluster structure of the data for data mining purposes. Therefore, the clustering method must be fast, robust, and visually efficient.

2. Partial discharge: concepts

Partial discharges occur wherever the electrical field is higher than the breakdown field of an insulating medium: Air: 27 kV/cm (1 bar), SF6: 360 kV/cm (4 bar), Polymers: 4000 kV/cm

They are generally divided into three different groups because of their different origins:

- **Corona Discharges** – Occurs in gases or liquids caused by concentrated electric fields at any sharp points on the electrodes.
- **Internal Discharges** – Occurs inside a cavity that is surrounded completely by insulation material; might be in the form of voids (e.g. dried out regions in oil impregnated paper-cables).
- **Surface Discharges** – Occurs on the surface of an electrical insulation where the tangential field is high e.g. end windings of stator windings.

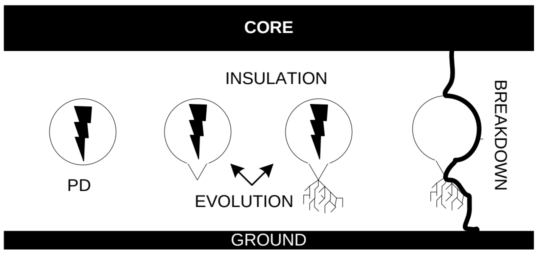

Fig. 1. Example of damage in polymeric power cable from the PD in a cavity to breakdown.

In general, the partial discharges are in consequence of local stress in the insulation or on the surface of the insulation. This phenomenon has a damaging effect on the equipments, for example transformers, power cables, switchgears, and others (see Figure 1). The first approach in a diagnosis is selecting the different features to classify measured PD activities into underlying insulation defects or source that generate PD's. The partial discharge measurement is a typical nondestructive test and it can be used to judge the insulation performance at the beginning of the service time taking into account the reduction of the performance during the service time by ageing, whereby the ageing depends on numerous parameters like electrical stress, thermal stress and mechanical stress. In particular for solid insulation like XLPE on power cables where a complete breakdown seriously damages the test object the partial discharge measurement is a tool for quality assessment. The charge that a PD generates in a cavity is called the physical charge and the portion of the cavity surface that the PD affects is called the discharge area. $E_{applied}$ is the applied electric field and $q_{physical}$ is the physical charge [17].

The pulse repetition rate n is given by the number of partial discharge pulses recorded in a selected time interval and the duration of this time interval. The recorded pulses should be above a certain limit, depending on the measuring system as well as on the noise level during the measurement. The pulse repetition frequency N is the number of partial discharge pulses per second in the case of equidistant pulses. Furthermore, the phase angle ϕ and the time of occurrence t_i are information on the partial discharge pulse in relation to the phase angle or time of the applied voltage with period *T*. For PD diagnosis test, is very important to classify measured PD activities, since PD is a stochastic process, namely, the occurrence of PD depends on many factors, such as temperature, pressure, applied voltage and test duration; moreover PD signals contain noise and interference [18]. Therefore, the test engineer is responsible for choosing proper methods to diagnosis for the given problem. In order to choose the features, it is important to know the different source of PD; an alternative is though pattern recognition. This task can be challenging, nevertheless, features selection has been widely used in other field, such as data mining [19] and pattern recognition using neural networks [8-10]. This research only shows test on laboratory without environment noise source, and it is a condition that does not represent the conditions on site, Markalous [20] presented the noise levels on site based on previous experiences.

The phase resolved analysis investigates the PD pattern in relation to the variable frequency AC cycle. The voltage phase angle is divided into small equal windows. The analysis aims to calculate the integrated parameters for each phase window and to plot them against the phase position (ϕ).

- ($q_m - \phi$): the peak discharge magnitude for each phase window plotted against ϕ, where q_m is peak discharge magnitude.

3. Self Organizing Map (SOM)

3.1 Winner takes all

The Self Organizing Map developed by Kohonen, is the most popular neural network models. The SOM algorithm [15,21] is based on unsupervised competitive learning called winner – takes – all, which means that the training in entirely data-driven and that the neurons of the map compete with each other.

Supervised algorithms [8, 9] like multi-layered perceptron, required that the target values for each data vector are known, but the SOM does not have this limitation. The SOM is a neural network model that implements a characteristics non-linear mapping from the high-dimensional space of input signal onto a typically 2-dimensional grid of neurons. The SOM is a two-layer neural network that consists of an input layer in a line and an output layer constructed of neurons in a two-dimensional grid.

The neighborhood relation of neuron *i*, an n-dimensional weight vector w is associated; n is the dimension of input vector. At each training step, an input vector **x** is randomly selected and the Euclidean distances between **x** and **w** are computed. The image of the input vector and the SOM grid is thus defined as the nearest unit w_{ik} and best-matching unit (BMU) whose weight vector is closest to the **x** [21]:

$$D(x, w_i) = \sqrt{\sum_k (w_{ik} - x_k)^2} \qquad (2)$$

The weight vectors in the best-matching unit and its neighbors on the grid are moved towards the input vector according the following rule:

$$\begin{aligned} \Delta w_{ij} &= \delta(c,i)\alpha(x_j - w_{ij}) \\ \Delta w_{ij} &= \alpha(x_j - w_{ij}) \ \ to \ \ i = c \\ \Delta w_{ij} & \quad to \quad i \neq c \end{aligned} \qquad (3)$$

where **c** denote the neighborhood kernel around the best-matching unit and α is the learning rated and δ is the neighborhood function.

The number of panels in the SOM is according the A x B neurons, the U-matrix representation is a matrix U ((2A-1) x (2B-1)) dimensional [22]. The selection of the distance criterion depends on application. In this paper, Euclidean distance is used because it is widely worn with SOM [23].

It is complicated to measure the quality of an SOM. Resolution and topology preservation are generally used to measure SOM quality [24]. There are many ways to measure them. The quantization error (qe) is calculated to measure the quality of the map. The quantization error qe is the average distance between each data vector and its BMU, measuring map resolution. The topological error te is the proportion of all data vectors for which first and second BMUs are adjacent units, otherwise this is regarded as violation of topology and thus penalized by increasing the error value.

3.2 The frequency sensitive competitive algorithm

The k-means algorithm has also the “dead units” problem, which means that if a centre is inappropriately chosen, it may never be updated, thus it may never represent a class.

To solve the “dead units” problem it has been introduced the so called “frequency sensitive competitive learning” algorithm (FSCL) [25] or competitive algorithm “with conscience”. Each centre counts the number of times when it has won the competition and reduces its learning rate consequently. If a center has won too often “it feels guilty” and it pulls itself out of the competition. The FSCL algorithm is an extension of k-means algorithm, obtained by modifying relation (2) according to the following one:

$$j = \arg\min \gamma_i \left\| x(n) - c_i(n) \right\| \qquad i = 1, \ldots, N \qquad (4)$$

where n is the inputs, N represents the centres numbers, the relative winning frequency γ_{i} of the centre c_i defined as:

$$\gamma_i = \frac{s_i}{\sum_{i=1}^{n} s_i} \qquad (6)$$

where s_i is the number of times when the centre c_i was declared winner in the past. So the centers that have won the competition during the past have a reduced chance to win again, proportional with their frequency termγ. After selecting out the winner, the FSCL algorithm updates the winner with next equation:

$$c_i(n+1) = c_i(n) - \eta\left[x(n) - c_i(n)\right] \qquad (7)$$

η is the learning rate, in the same way as the *k*-means algorithm, and meanwhile adjusting the corresponding s_i with the following relation:

$$s_i(n+1) = s_i(n) + 1 \qquad (8)$$

3.3 The rival penalized competitive learning algorithm

The rival penalized competitive algorithm (RPCL) [25] performs appropriate clustering without knowing the clusters number. It determines not only the winning centre j but also the second winning center r, named rival

$$r = \arg\min \gamma_i \left\| x(n) - c_i(n) \right\|, \quad i = 1, \ldots, N \;\; i \neq j \qquad (9)$$

The second winning centre will move away its centre from the input with a ratio □□, called the de-learning rate. All the other centres vectors will not change. So the learning law can be synthesized in the following relation:

$$c_i(n+1) = \begin{cases} c_i(n) + \eta\left[x(n) - c_i(n)\right] & if \;\; i = j \\ c_i(n) - \beta\left[x(n) - c_i(n)\right] & if \;\; i = j \\ c_i(n) \;\; if \;\; i \neq j \;\; and \;\; i \neq r \end{cases} \qquad (10)$$

If the learning speed η is chosen much greater than β, with at least one order of magnitude, the number of the output data classes will be automatically found. In other words, suppose that the number of classes is unknown and the centres number N is greater than the clusters number, than the centres vectors will converge towards the centroids of the input data classes. The RPCL will move away the rival, in each iteration, converging much faster than the k-means and the FSCL algorithms.

4. Analysis of PD Data

PD measurements for power cables are generated and recorded through laboratory tests. Corona was produced with a point to hemisphere configuration: needle at high voltage and hemispherical cup at ground. Surface discharge XLPE cable with no stress relief termination applied to the two ends. High voltage was applied to the cable inner conductor and the cable sheath was grounded, this produces discharges along the outer insulation surface at the cable ends. Internal discharge was used a power cable with a fault due to electrical treeing. Were considered the pattern characteristic of univariate phase-resolved distributions as inputs, the magnitude of PD is the most important input as it shows the level of danger, for this reason the input in the SOM the raw data is the peak discharge magnitude for each phase window plotted against ($qm - \phi$). Figure 2 shows the conceptual diagram training. In the cases analyzed, the original dataset is 1 million of items, was used a neurons array of 10×10 cells to extract features. As it is well known, in fact, a too small number of neurons per class could be not sufficient to represent the variability of the samples to be classified, while a too large number in general makes the net too much specialized on the samples belonging to the training set and consequently reduces its generalization capability. Moreover a too large number of neuron per class implies a long training time and a possible underutilization of some of the neural units.

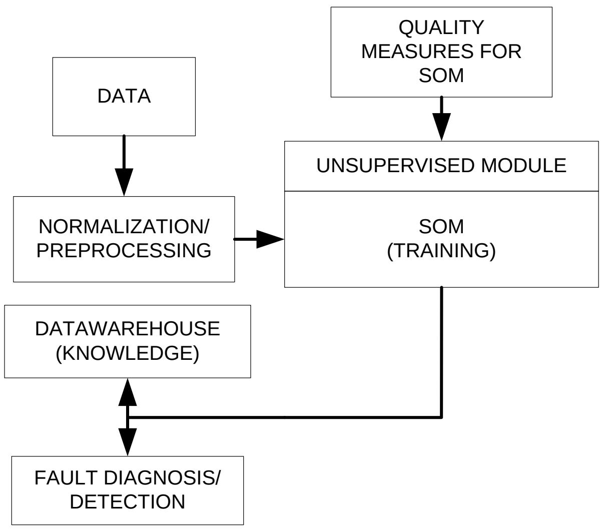

Fig. 2. The component interaction between SOM

In table 2 are shower the parameters for training to each competitive learning algorithm. The coefficient γ has been dynamically changed during the training.

Table 2 Parameter for training

	WTA	**FSCL**	**RPCL**
Epoch	100	100	100
η	0.1	0.1	0.05
β	0.01	0.01	0.01

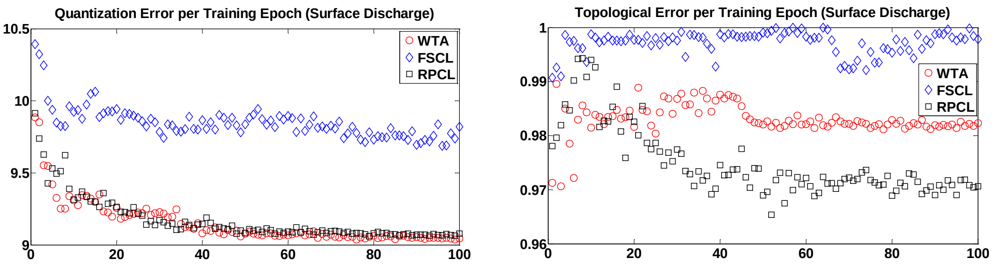

Fig. 3. Quantization and Topological error per Training Epoch (Surface Discharge)

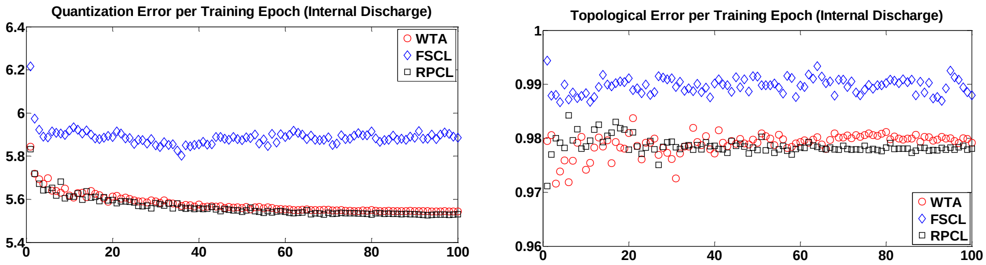

Fig. 4. Quantization and Topological error per Training Epoch (Internal Discharge)

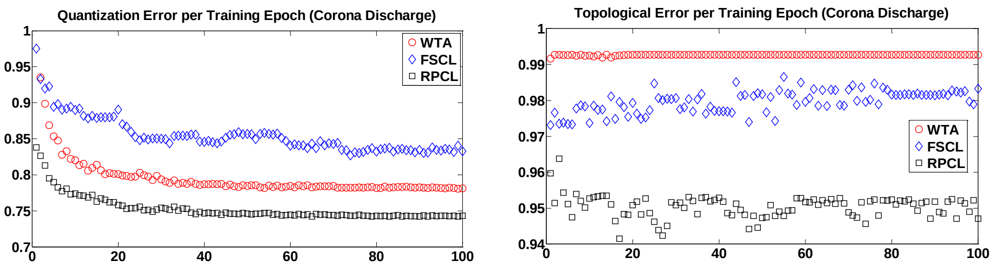

Fig. 5. Quantization and Topological error per Training Epoch (Corona Discharge)

In figure 3, 4 and 5 is showed the performance of the competitive learning algorithms to different PD source.

We evaluate the performance of algorithms in which competitive learning is applied of partial discharge dataset, quantization error, topological error and time in seconds per training epoch are showed in table 3, and it is observer that the WTA works with less error and less time of training, but the FSCL is not always satisfactory because the training time is very long. The RCPL have longest training time but is the algorithm with more error.

Table 3 Performance of training in SOM

	WTA	**FSCL**	**RPCL**
Surface Discharge			
$\mathbf{q_e}$	9.0	9.8	9.1
$\mathbf{t_e}$	0.985	1	0.97
time	849 seconds	1160 seconds	1226 seconds
Internal Discharge			
$\mathbf{q_e}$	5.5	5.9	5.4
$\mathbf{t_e}$	0.98	0.99	0.98
time	173 seconds	222 seconds	241 seconds
Corona Discharge			
$\mathbf{q_e}$	0.78	0.85	0.75
$\mathbf{t_e}$	0.99	0.98	0.95
time	889 seconds	1191 seconds	1362 seconds

Conclusion

PD patterns recognition and classification require an understanding of the traits commonly associated with the different source and relationship between observed PD activity and responsible defect sources. This paper shows the performance of SOM using different competitive learning algorithms to classify measured PD activities into underlaying insulation defects or source that generate PD's, its showed that WTA is the better algorithm with less error and training time, but its overall performance are not always satisfactory, being alternative in accord at the performance FSCL or RPCL algorithms.

References

[1] IEC 60270 Ed. 2. High-voltage test techniques - Partial discharge measurements. (2000). 15 – 16.
[2] Pollard D. Quantization and the method of k-means. IEEE Trans. Information Theory (1982). 28 – 199
[3] Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press; 1981.
[4] Yang MS. A survey of fuzzy clustering, Math Comput Model 1993;18:1.16.
[5] Höppner F, Klawonn F, Kruse R, Runkler T. Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition. New York: Wiley; 1999.

[6] X. Wang, H. Liu, J. Lu, and T. Yohagy, “Combining recurrent neural networks with self-organizing maps for channel equalization,” IEEE Trans. on Communications, vol. E85-E, pp. 2227–2235, 2002.
[7] L. Xu, A. Krzyzak, and A. E. Oja, “Rival penalized competitive learning for clustering analysis. rbf net and curve detection” IEEE Trans. on Neural Networks, no. 4, pp. 636–64, 1993.
[8] Mazroua, A. PD pattern recognition with neural netwoks using the multilayer perception technique. IEEE Transactions on Electrical Insulation, 28, (1993). 1082-1089.
[9] Krivda, A. Automated Recognition of Partial Discharge. IEEE Transactions on Dielectrics and Electrical Insulation, 28, (1995). 796-821.
[10] 10 Kim, J. Choi, W. Oh, S. Park, K. Grzybowski, S. Partial Discharge Pattern Recognition Using Fuzzy-Neural Networks (FNNs) Algorithm. IEEE International Power Modulators and High Voltage Conference. (2008). 272-275.
[11] Ri-Cheng, L. Kai, B. Chun, D. Shao-Yu, L. Gou-Zheng, X. Study on Partial Discharge Localization by Ultrasonic Measuring in Power Transformer Based on Particle Swarm Optimization. International Conference on High Voltage Engineering and Application. (2008). 600-603.
[12] Chang, W. Yang, H. Application of Self Organizing Map Approach to Partial Discharge Pattern Recognition of Cast-Resin Current Transformers. WSEAS Transaction on Computer Research. Issue 3. Volume 3. (2008). 142-151.
[13] Fadilah Ab Aziz, N. Hao, L. Lewin, P. Analysis of Partial Discharge Measurement Data Using a Support Vector Machine. 5th Student Conference on Research and Development. (2007). 1-6.
[14] Hirose, H. Hikita, M. Ohtsuka, S. Tsuru, S. Ichimaru, J. Diagnosis of Electric Power Apparatus using the Decision Tree Method. IEEE Transactions on Dielectrics and Electrical Insulation, 15, (2008). 1252-1260.
[15] M. Kantardzic, Data Clustering, Theory, Algorithms and Methods. ASA-SIAM, 2007. 53 – 57.
[16] Vesanto, J. Data Exploration Process Based on the Self Organizing Map. Doctoral Thesis in Computer Science. Helsinki University of Technology. (2002).
[17] Forssén, C. Modelling of cavity partial discharges at variable applied frequency. Sweden: Doctoral Thesis in Electrical Systems. KTH Electrical Engineering. (2008).
[18] Edin, H. Partial discharge studies with variable frequency of the applied voltage. Sweden: Doctoral Thesis in Electrical Systems. KTH Electrical Engineering. (2001).
[19] K. Lai, B. P. Descriptive Data Mining of Partial Discharge using Decision Tree with genetic algorithms. AUPEC. (2008).
[20] Markalous, S. Detection and location of Partial Discharges in Power Transformers using acoustic and electromagnetic signals. Stuttgart University: PhD Thesis. (2006).
[21] Kohonen T. Engineering Applications of Self Organizing Map. Proceedings of the IEEE. (1996).
[22] Rubio-Sánchez, M. Nuevos Métodos para Análisis Visual de Mapas Auto-organizativos. PhD Thesis. Madrid Politechnic University. (2004).
[23] Vesanto J., Alhoniemi E., Clustering of the Self Organizing Map. IEEE Transactions on Neural Networks , Vol. 11, Nr. 3, pages 1082-1089. (2000).
[24] Pölzlbauer, G. Survey and Comparison of Quality Measures for Delf-Organizing Maps. *Proceedings of the Fifth Workshop on Data Analysis* (2004). 67-82.
[25] C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton "Competitive learning algorithms for vector quantization", Neural Networks, vol. 3, no. 3, pp. 277-290. (1990).