Competitive learning for self organizing maps used in classification of partial discharges

Aprendizaje competitivo para mapas autoorganizados utilizados en clasificación de descargas parciales

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

[1] Laboratorio de Pruebas a Equipos y Materiales, Comisión Federal de Electricidad. Reforma 164, col. Juárez, México, DF, México
[2] Centro de Innovación Aplicada en Tecnologías Competitivas
[3] Universidad Autónoma de Ciudad Juárez. Omega 201, fracc. Industrial Delta, CP 37545, León, Guanajuato, México.
[4] Universidad Autónoma de Aguascalientes. México
* E-mail: ruben.jaramillo@cfe.gob.mx

KEYWORDS:

competitive learning, self-organizing maps, partial discharge, quality measurements, diagnosis.

ABSTRACT

In this paper different competitive learning algorithms for self-organizing maps (SOM) are experimentally examined. The characterization of the results obtained is presented in terms of quality of SOM. The competitive learning algorithms evaluated through SOM are winner-takes-all, frequency sensitive competitive learning, and rival penalized competitive learning. Case study: their performance in the classification of partial discharges on power cables.

PALABRAS CLAVE:

aprendizaje competitivo, mapas autoorganizados, descargas parciales, métricas de calidad, diagnóstico.

RESUMEN

En este artículo se presentan diferentes algoritmos de aprendizaje competitivo utilizados en mapas autoorganizados (SOM, por sus siglas en inglés). Estos algoritmos son evaluados y comparados experimentalmente, y la caracterización es presentada en términos de la calidad del SOM. Los algoritmos de aprendizaje competitivo evaluados de esta manera son *el ganador toma todo* (WTA, por sus siglas en inglés), aprendizaje competitivo sensible a la frecuencia y rival penalizado. Dominio de aplicación: desempeño en la clasificación de patrones de descargas parciales en cables de potencia.

1 INTRODUCTION

Competitive learning is an efficient tool for self organizing maps (SOM), widely applied in a 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 huge volumes of data being generated from 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. Partial discharge (PD) is a common phenomenon that occurs in insulation of high voltage; this definition is given in IEC 60270 [1]. In general, partial discharges are a consequence of local stress in the insulation or on the surface of the insulation.

The typical competitive learning algorithm *k*-means (also 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* pre-determined codevectors (also called codebooks) [2]. Fuzzy *c*-means (FCM) clustering is a fuzzy extension of hard *c*-means clustering. FCM and its varieties have been widely studied and applied in various areas [3-5]. During the last fifteen years there new advanced algorithms have been developed that eliminate the "dead units" problem and perform clustering without predicting the exact cluster number, as for example: the frequency sensitive competitive learning algorithm (FSCL) [6], the incremental *k*-means algorithm, and the rival penalizing competitive learning algorithm (RPCL) [7].

We evaluated the performance of algorithms in which competitive learning is applied to partial discharge datasets, in terms of quantization error, topological error and time in seconds per training epoch. The result from classification of PD shows that *winner-takes-all* (WTA) algorithm has better performance than FSCL and RPCL.

Table 1 shows a summary of researchers who worked on feature extraction, recognition and classification of PD, next to the artificial intelligence tools they used and the constraints associated with these methods.

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 we compare different competitive learning algorithms to classify measured PD activities into underlying insulation defects or source that generate PDs using self organizing maps

Table 1. Classification and diagnosis in partial discharge 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 networks and fuzzy neural networks	Performance, in the case of multiple discharges, and including defects and noise
Ri-Cheng *et al.*	Tool: Particle swarm optimization Objective: Localization of PD in the power transformer	On site application should improve performance
Chang *et al.*	Tool: Self-organizing map (SOM) Objective: PD pattern recognition and classification	Quality and optimization structure of SOM
Fadilah Ab Aziz *et al.*	Tool: Support vector machine (SVM) Objective: Feature selection and PD classification	SVM is not reliable for a small dataset
Hirose *et al.*	Tool: Decision tree Objective: Feature extraction and PD classification	The allocation rules are sensitive to small perturbations in the dataset (instability)

(SOM). Multidimensional scaling (MDS) is a nonlinear feature extraction technique, whose aim is to represent a multidimensional dataset in two or three dimensions, so 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, 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 its cluster structure 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 occur in gases or liquids caused by concentrated electric fields at any sharp points on the electrodes.
- Internal discharges occur 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 occur on the surface of an electrical insulation where the tangential field is high (e.g. end windings of stator windings).

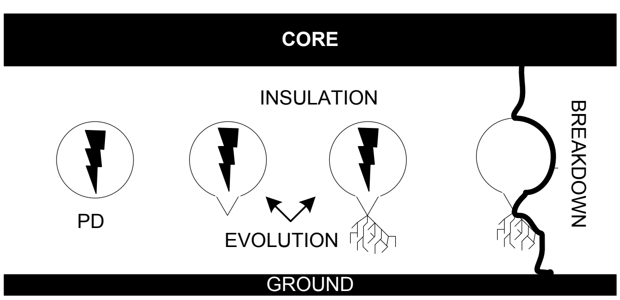

Figure 1. Example of damage in a polymeric power cable, from a partial discharge in a cavity, to its breakdown.

In general, partial discharges are a consequence of local stress in the insulation or on the surface of the insulation. This phenomenon has a damaging effect on equipment, for example on transformers, power cables, switchgears, and others (Figure 1). The first approach in a diagnosis is selecting the different features to classify measured PD activities into underlying insulation defects or sources that generate PDs. The partial discharge measurement is a typical nondestructive test and it can be used to judge the insulation performance at the beginning of service time, taking into account the reduction of the performance during 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 cross-linked polyethylene (XLPE) on power cables, where a complete breakdown seriously damages the test object— 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.

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 a PD diagnosis test, it 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. Therefore, the test engineer is responsible for choosing proper methods to diagnose for the given problem. In order to choose the features, it is important to know the different sources of PD; an alternative is using pattern recognition. This task can be challenging; nevertheless, feature selection has been widely used in other fields, such as data mining and pattern recognition using neural networks [8-10]. This research shows tests under laboratory conditions, without an environmental source of noise, and does not represent the conditions on site. Markalous presented 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 model. The SOM algorithm is based on an unsupervised competitive learning algorithm called *winner takes all*, which means that the training is entirely data-driven and that the neurons of the map compete with each other.

Supervised algorithms [8, 9] like multi-layered perceptron, require 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 non-linear mapping of features from the high-dimensional space of input signals onto a typical 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 are thus defined as the nearest unit w_{ik} and best-matching unit (BMU) whose weight vector is closest to **x**:

$$D\left(x, w_i\right) = \sqrt{\sum_k \left(w_{ik} - x_k\right)^2} \tag{1}$$

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

$$\begin{aligned} \Delta w_{ij} &= \delta\left(c, i\right)\alpha\left(x_j - w_{ij}\right) \\ \Delta w_{ij} &= \alpha\left(x_j - w_{ij}\right) \ \ to \ \ i = c \\ \Delta w_{ij} & \quad to \quad i \neq c \end{aligned} \tag{2}$$

where **c** denotes 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 to the A x B neurons, the U-matrix representation is a U ((2A-1) x (2B-1)) dimensional matrix. The selection of the distance criterion depends on application. In this paper, Euclidean distance is used because it is widely employed with SOMs.

It is complicated to measure the quality of a SOM. Resolution and topology preservation are generally used to measure SOM quality. There are many ways to measure it. The quantization error (*qe*) is calculated to measure the quality of the map. The *qe* is the average distance between each data vector and its BMU, measuring map resolution. The topological error t_e is the proportion of all data vectors for which first and second BMUs are adjacent units, otherwise this is regarded as a 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 center is inappropriately chosen, it may never be updated, thus it may never represent a class.

To solve the "dead units" problem the so called frequency sensitive competitive learning algorithm (FSCL) was introduced, also known as a competitive algorithm "with conscience". In this, each center counts the number of times 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 the *k*-means algorithm, obtained by modifying relation (2) according to the following one:

$$j = \arg\min \gamma_i \left\| x\left(n\right) - c_i\left(n\right) \right\| \qquad i = 1, \ldots, N \tag{3}$$

where n are the inputs, N represents the number of centers, the relative winning frequency γ_i of the center c_i defined as:

$$\gamma_i = \frac{s_i}{\sum_{i=1}^{n} s_i} \tag{4}$$

where s_i is the number of times when the center 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 term γ. After selecting out the winner, the FSCL algorithm updates the winner with the next equation:

$$c_i(n+1) = c_i(n) - \eta\left[x(n) - c_i(n)\right] \quad (5)$$

where 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:

$$s_i(n+1) = s_i(n) + 1 \quad (6)$$

3.3 The rival penalized competitive learning algorithm

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

$$r = \arg\min \gamma_i \left\| x(n) - c_i(n) \right\|, \quad i = 1,\ldots,N \;\; i \neq j \quad (7)$$

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

$$c_i(n+1) = \begin{cases} c_i(n) + \eta\left[x(n) - c_i(n)\right] \;\; if \;\; i = j \\ c_i(n) - \beta\left[x(n) - c_i(n)\right] \;\; if \;\; i = j \\ c_i(n) \;\; if \;\; i \neq j \;\; and \;\; i \neq r \end{cases} \quad (8)$$

If the learning speed η is chosen much bigger 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 number of centers *N* is bigger than the number of clusters, then the centers' 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. A 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 was 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. For internal discharge a power cable with a fault due to electrical treeing was used. The pattern characteristics of univariate phase-resolved distributions were considered as inputs; the magnitude of PD is the most important input as it shows the level of danger, for this reason the raw data input in the SOM was the peak discharge magnitude for each phase window plotted against (qm –ϕ). Figure 2 shows the conceptual training diagram. In the cases analyzed, the original dataset was 1 million items; a neurons array of 10 10 cells was used to extract features. As it is well known, too small a number of neurons per class could not be sufficient to represent the variability of the samples to be classified, while too large a number in general makes the net too specialized on the samples belonging to the training set and consequently reduces its generalization capability. Moreover too large a number of neurons per class implies a long training time and a possible underutilization of some of the neural units.

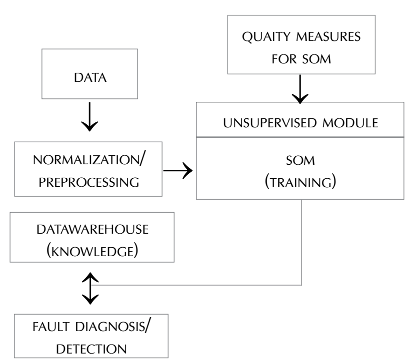

Figure 2. Interaction between components in a self-organizing map

Table 2. Parameters for training

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

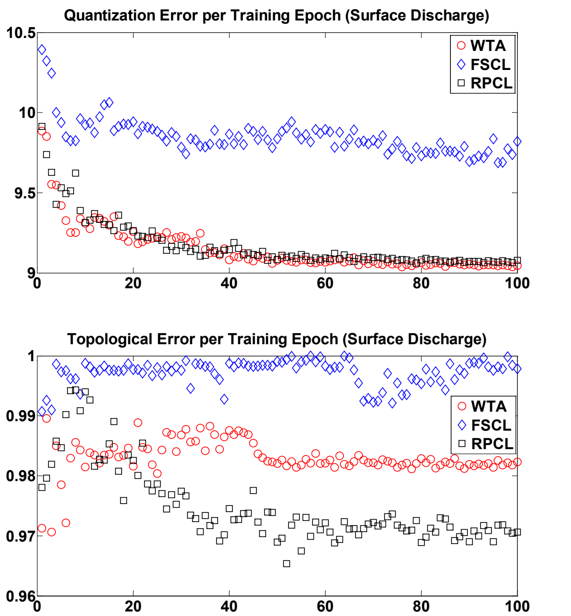

Figure 3. Quantization and topological errors per training epoch (surface discharge)

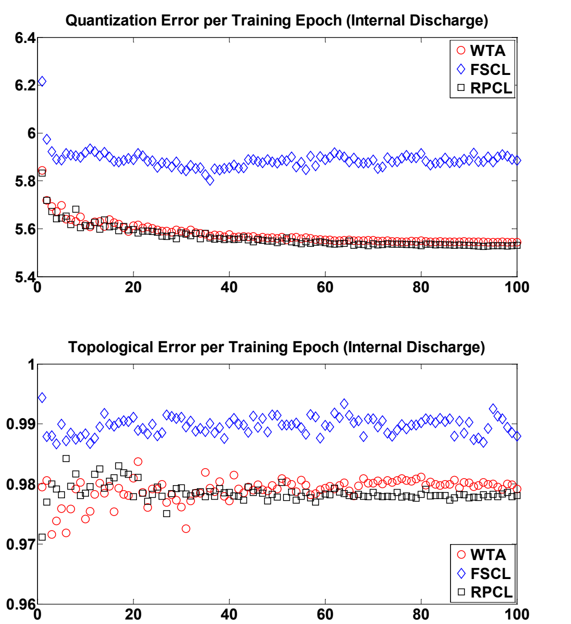

Figure 4. Quantization and topological errors per training epoch (internal discharge)

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

Figures 3, 4 and 5 show the performance of the competitive learning algorithms for different PD sources. We evaluates the performance of algorithms in which competitive learning is applied to a partial discharge dataset, in terms of quantization error, topological error and time in seconds per training

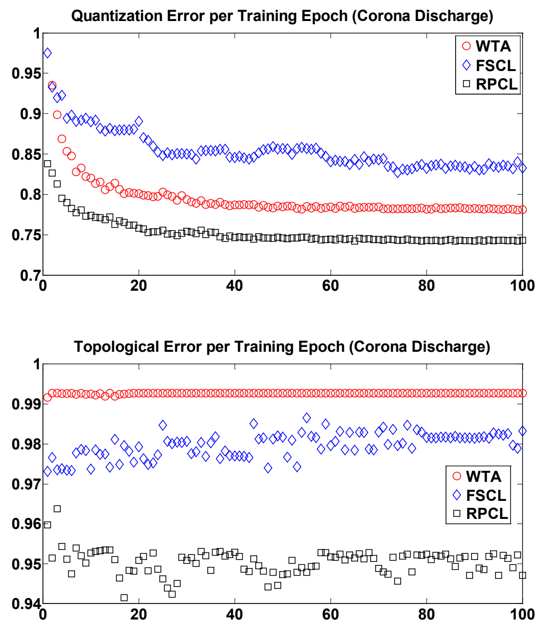

Figure 5. Quantization and topological errors per training epoch (corona discharge)

Table 3. Performance of training in self-organizing maps

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

epoch (Table 3), and we observed that WTA works with less error and less training time, FSCL is not always satisfactory because training time is very long, and RPCL has the longest training time and is the algorithm with more error.

5 CONCLUSION

PD pattern recognition and classification require an understanding of the traits commonly associated with its different sources and of the relationship between observed PD activity and defects in sources responsible for it. This paper shows the performance of SOM using different competitive learning algorithms to classify measured PD activities into underlying insulation defects or sources that generate PDs. It is shown that WTA is the best algorithm with less error and training time, but its overall performance is not always satisfactory, being an alternative to the performance of FSCL or RPCL algorithms.

REFERENCES

1. International Electrotechnical Comission. (2000). IEC 60270 Ed. 2. High-voltage test techniques - Partial discharge measurements. Geneva: IEC, 2000:15-16.
2. Pollard D. (1982). Quantization and the method of k-means. IEEE Trans. *Information Theory* 28 (2), 199-205.
3. Bezdek J. C. (1981). Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press.
4. Yang M. S. (1993). A survey of fuzzy clustering. *Math Comput Model* 18 (11), 1-16.
5. Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy. (1999). Cluster Analysis: Methods for Classification Data Analysis and Image Recognition. New York: Wiley.
6. Wang X., Lin H., Lu J., Yahagi T. (2002). Combining recurrent neural networks with self-organizing maps for channel equalization. *IEEE Trans. on Communications*. 2002, E85-B (10), 2227-2235.
7. Xu L, Krzyzak A, Oja E. (1993). Rival penalized competitive learning for clustering analysis, RBF net, and curve detection. *IEEE Trans Neural Netw*, 4 (4), 636-649.
8. Mazroua A. (1993). PD pattern recognition with neural netwoks using the multilayer perception technique. *IEEE Transactions on Electrical Insulation*, 28, 1082-1089.
9. Krivda A. (1995). Automated Recognition of Partial Discharge. *IEEE Transactions on Dielectrics and Electrical Insulation*, 28, 796-821.
10. Kim J, Choi W, Oh S, Park K, Grzybowski S. (2008) Partial Discharge Pattern Recognition Using Fuzzy-Neural Networks (FNNs) Algorithm. IEEE International Power Modulators and High Voltage Conference. Proceedings of 2008, 272-275

About the authors

Ruben Jaramillo-Vacio received its BSc in Electromechanical Engineering from ITESI (2002), Master in Electrical Engineering from San Luis Potosi University (2005), and Master in Management Engineering from La Salle University (2010). He has joint CIATEC (Conacyt research center) in 2009 to carry out his PhD research in industrial engineering and manufacturing. Since 2005 he is test engineer in CFE-LAPEM in dielectric test, partial discharge diagnosis at power cables. His research interest includes partial discharge diagnosis using intelligence artificial tools.

Carlos Alberto Ochoa Ortiz Zezzatti (BSc 1994; Eng. Master, 2000; PhD, 2004; Postdoctoral researcher, 2006; Industrial postdoctoral research, 2009). He has written three books and eleven chapters in books related to AI. He has supervised ten PhD theses, 21 Master theses and 32 Bachelor theses. He participated in the organization of conferences such as HAIS'07, HAIS'08, ENC'06, ENC'07, ENC'08, MICAI'09, MICAI'10 and MICAI'11. His research interests include evolutionary computation, natural processing language and social data mining.

Julio César Ponce Gallegos received its BSc degree in computer system engineering from the Universidad Autónoma de Aguascalientes (UAA) in 2003, its MSc degree in computer sciences from the UAA in 2007, and its PhD in evolutionary computation from UAA. He is currently professor in UAA. His research interests include evolutionary computation and data mining.