Performance Analysis of Hybrid Non-Supervised & Supervised Learning Techniques Applied to the Classification of Faults in Energy Transport Systems

Jhon Albeiro Calderón[1], Germán Zapata Madrigal[2] and Demetrio A. Ovalle Carranza[3]
[1]Interconexión Eléctrica S.A. E.S.P.,
[2]Electrical and Mechanics Engineering Department,
[3]Computer Science Department, National University of Colombia – Campus Medellin
Colombia

1. Introduction

Most of the power systems protection techniques are related to the definition of system states by means of the identification of patterns from waveform of voltage and associated current. This means that the development of an adaptive protection can essentially be treated as a problem about classification/recognition of patterns (Song et al., 1997). Nevertheless, because the main causes of faults and the operation of nonlinear devices under certain conditions of fault, the methods of recognition of conventional patterns are unsatisfactory in some applications, particularly, in the case of high complexity electrical systems. In this sense, neural networks play an important role due to their unique ability of mapping nonlinear relations.

Some successful applications of neural networks in the area of electrical engineering (Song et al., 1996), (Dillon & Niebur, 1996) have demonstrated that they can be used like an alternative method to solve certain problems of great complexity where the conventional techniques have experienced difficulties. Nevertheless, when giving a glance to the different applications of neural networks to electric power systems, it is clear that almost all the developments that have been carried out are based on the multi-layers perceptron with retro-propagation learning algorithms (BP). Although, BP can provide very compact distributed representations of complex data sets, it has some disadvantages such as the following: it exhibits slow learning, it requires great sets of training, they easily fall in local minimums, and in general it shows little robustness (Song et al., 1997).

Another type of learning is the non-supervised one that surrounds the learning of patterns without a target. A typical non-supervised learning network is the Self-Organized Mapping (SOMs) developed by Teuvo Kohonen. A SOM network has the advantage of fast learning and small sets of training. Nevertheless, due to the absence of an output "truth" layer in the SOM, its use is not recommendable for the classification of patterns. Instead, it is used as an

Open Access Database www.intechweb.org

Source: Machine Learning, Book edited by: Abdelhamid Mellouk and Abdennacer Chebira,
ISBN 978-3-902613-56-1, pp. 450, February 2009, I-Tech, Vienna, Austria

initial procedure ("front-end") to an output layer with a supervised training, that is, combined non-supervised/supervised learning.
The networks that combine non-supervised and supervised learning have the powerful ability to organize any complexity of highly nonlinear patterns recognition problem. This type of neural network is insensitive to noise due to the low internal dimensional representation (Song et al., 1996). Based on this kind of characteristics the present research developed a hybrid model entitled Artificial Intelligence Adaptive Model (AIAM) (Calderón, 2007).
Next, it will be initially described the basic functionality of the several models analyzed in the research, the respective main results, and finally the AIAM model and conclusions.

2. Neural networks models

With the purpose of selecting the most appropriate neural network model to be used for the classification of faults in an Electrical Power System (EPS) an exploration of alternatives on models of neural networks was carried out based on the state-of-the-art of the subject (El-Sharkawi & Niebur, 1996), (Aggarwal & Song, 1997), (Aggarwal & Song, 1998a), (Aggarwal & Song, 1998b), (Kezunovic, 1997), (Dalstein & Kulicke, 1995), (Keerthipala et al, 1997), (Sidhu & Mitai, 2000), (Fernandez & Ghonaim, 2002), (Dalstein et al, 1996), (Zahra et al, 2000), (Ranaweera, 1994), (Oleskovicz et al., 2001), (Song et al, 1997), (Song et al, 1996), (Dillon & Niebur, 1996), (Dillon & Niebur, 1999),(Badrul et al., 1996).
Next, four important classifiers, based on neural networks, will be briefly described. Special emphasis was placed on the basic principles and differences, instead of a detailed description itself.

2.1 Back-Propagation classifier (BP)

BP classifiers are the most popular and widely applied neural networks. They train with supervision using the descending gradient algorithm to diminish the error between the real exits and the wished exits of the network.
In Fig. 1. the general architecture of this type of network is illustrated.

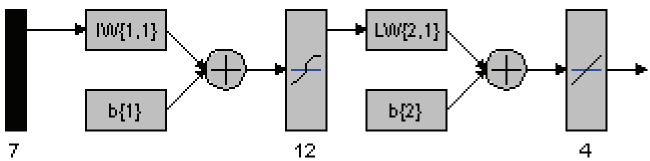


Fig. 1. General architecture used by the model of retro-propagation training. (Matlab educational license).

Many articles provide good introductions to the methods and successful applications of this type of neural networks applied to the power systems. Nevertheless, in general, most of the BP classifiers are (1) of prolonged training time; (2) of difficult selection for the optimal size, and (3) potentially with tendency to be caught in a local minimum (Song et al., 1996).
For this reason, improvements have been developed in recent years, particularly in the aspect concerning the learning process. In this sense, it is valuable to mention the fuzzy algorithms of controlled learning and the training based on genetic algorithms.

2.2 Feature Mapping classifier (FM)

One of the most important algorithms of non-supervised learning is the Self-Organized Feature Mapping (SOFM) proposed by Kohonen shown in Fig. 2. The SOFM is used to map non-supervised input vectors in a bi-dimensional space where the vectors are self-organized in groups that represent the different types.

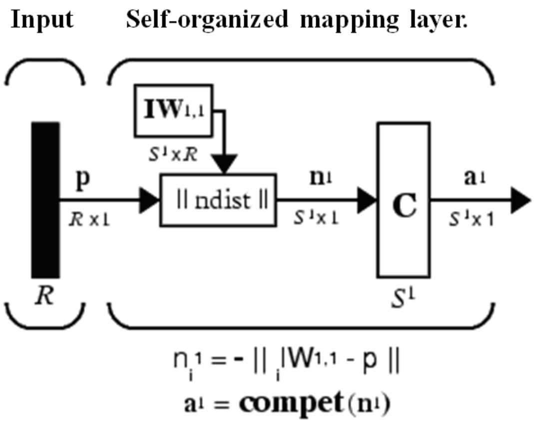


Fig. 2. General architecture used by the Kohonen model of SOMF. (Matlab educational license).

The SOFM learns to classify input vectors according to the form that they are grouped in the input space. This method differs from the competitive method of layers in which the neighboring neurons in the SOFM learn to recognize also adjacent sections in the input space. Thus, the self-organized maps learn so much the distribution (as the competitive method of layers makes it) as well as the topology of the input vectors that train. Neurons in the layer of a SOFM are originally organized in physical positions according to a topological function. The distances between neurons are calculated from their positions with a distance function.

In these networks there is no target for the error evaluation. That is, the learning of the synaptic weights is non-supervised, which means that, under the presentation of new input vectors, the network dynamically determines these weights, in such a way that, input vectors that are closely related will excite neurons that are closely grouped (Badrul et al., 1996). It is able to separate data in a specified number of categories and therefore able to act like a classifier. In the Kohonen network there are only two layers: an input layer where the patterns of the variables are placed and an output layer that has a neuron for each possible category or type.

2.3 Radial Base Function classifier (RBF)

The construction of a RBF in its most basic form considers three layers entirely different, as in Fig. 3. The first layer consists of the input nodes. The second layer is composed by the denominated Kernel nodes (base radial layer) which functions are different from those of a BP network. The Kernel nodes based on the radial base functions calculate symmetrical functions which are a maximum when the input is near the centroid of a node. The output nodes are simple sums.

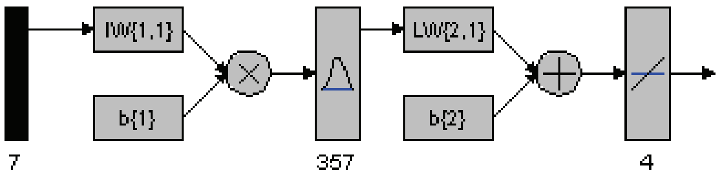


Fig. 3. General architecture used by the RBF model. (Matlab educational license).

This particular architecture of RBF has been proven to improve the training time but at the expense of considering many nodes in the radial base layer and connections of weights (in critical cases the number of neurons of this layer could get to be equal to the number of training samples, that is to say, a neuron per input pattern).

2.4 Vector Quantification Learning classifier (LVQ)

The Vector Quantification Learning network (LVQ) is a form of adaptive classifier which structure is shown in Fig. 4. This classifier requires a final stage of supervised training to improve its performance. LVQ contains an input layer, a Kohonen layer and the output layer. The number of nodes of the entrance layer is equal to the number of entrance parameters. The number of nodes of the Kohonen layer is based on the number of input vectors in the training data. The output layer contains a node for each type.

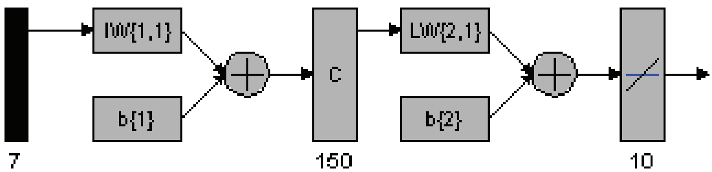


Fig. 4. General architecture used by the LVQ model. (Matlab educational license).

Based on the analysis of the previous neural networks models the research was oriented into two ways:

- To complement the neural model BP with a learning method that allowed improving the generalization and the resulting classification error. In order to do this the Bayesian Regularization methodology (BR) described in (Foresee and Hagan, 1997), (Hagan et al, 2002), (MacKay, 1998), was used.
- The search of an Adaptive model that would take advantage of the kindnesses of the combination of the non-supervised learning with the supervised learning but, as well as looking for to fix the weaknesses found in LVQ and RBF methods. To get this, the doctorate thesis (Vasilic, 2004), consisting in the Adaptive Resonance Theory (ART), was used as the starting point.

2.5 BP neural network model with Bayesian regularization

Taking into account the considerations of the previous concept, it was implemented the performance evaluation of the neural network BP incorporating additional training techniques to improve its performance. With the purpose of obtaining a high capacity of generalization of the network, it was considered (Foresee and Hagan, 1997), (Hagan et al,

2002), (MacKay, 1998), the approach known as the Bayesian regularization in which the weights of the network are assumed as random variables with specific distributions so that the parameters of stabilization get associated to the unknown variances connected to these distributions. In this way, it is possible to consider the parameters using technical statistics. A more a detailed description of this approach and its combination with other techniques can be found in (Foresee & Hagan, 1997).
In MATLAB it is possible to use this methodology by means of the trainbr algorithm that can be established as argument at the time of defining the network by means of the function newff.
For the detection and classification of the fault, a feed-forward network was used with a single hidden layer of s neurons (Foresee & Hagan, 1997), (Hagan et al, 2002), (MacKay, 1998). 7 neurons were considered for the input layer that corresponds to the rms values of the voltages and currents of the 3 phases, plus the sequence zero current. For the output layer, 4 neurons were considered corresponding to the binary values that indicate the failed phase (the 3 first bits) and if it is or not grounded (last bit). In this case, the used value of s was of 12 (value that was obtained after doing different tests of verification trying to diminish the resulting error, but at the same time guaranteeing a suitable level of generalization).
The general model of this network it is shown in Fig. 5. The functions of activation of MATLAB Tansig were used in the hidden layer and in the output layer the linear transference Purelin.

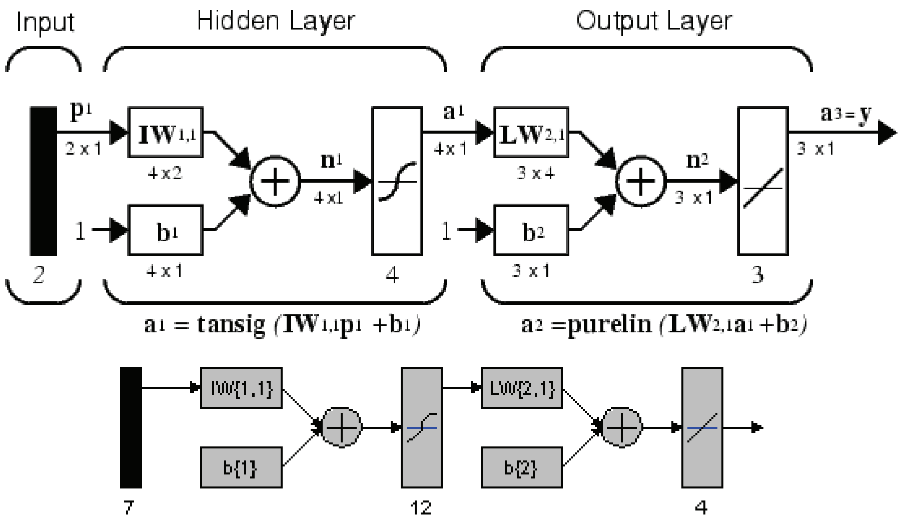


Fig. 5. Diagram of the classifier algorithm and the used neural network architecture BR. (Matlab educational license).

2.6 ART model (adaptive resonance theory)

ART Model does not have a defined typical structure with a specified number of neurons. Instead, it is made up of an adaptive structure with auto-evolving neurons. The structure solely depends on the characteristics and order of presentation of the patterns in the input data set. In Fig. 6. the diagram of the complete procedure used for the training of the neuronal network type ART is explained in (Vasilic, 2004).

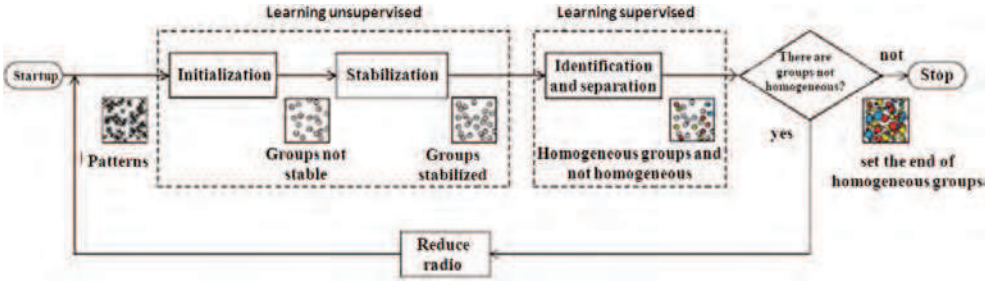


Fig. 6. Combined learning of Supervised and Non- supervised Neural Networks (Vasilic, 2004).

The training consists in numerous iterations in the stages of supervised and non-supervised learning, suitably combined to obtain a maximum efficiency. Groups of similar patterns lay in groups, defined as hyper-spheres in a multidimensional space, where the dimension of the space is determined by means of the length of the input patterns. The neural network initially uses non-supervised learning with input patterns not tagged in order to form unstable fugitive groups. This is an attempt to discover the patterns density by means of getting them in groups to consider prototypes of groups that can serve as prototypes of the typical input patterns. The category tags are assigned later on to the groups during the stage of supervised learning. The tuning parameter called “threshold of monitoring” or “radio”, controls the size of the group and therefore the number of generated groups, and it is consecutively reduced during the iterations. If the monitoring threshold is high, many different patterns within a group can then be incorporated, and this generates a small number of heavy groups. If the monitoring threshold is low, they only activate the same group patterns that are very similar, and this generates a great number of fine groups. Subsequent to the training, the centers of the groups serve as typical neurons of the neural network. The structure of prototypes only depends on the density of the input patterns. Each training pattern has been located in only one group, at the same time as each group contains one or more similar input patterns. A prototype is centrally located in the respective group, and it is either identical to one of the real patterns or identical to a synthesized prototype of the found patterns. A category tag is assigned to each group symbolizing a type of groups with a symbolic characteristic, meaning that each group belongs to one of the existing categories. The number of categories corresponds to the desired number of outputs of the neural network. Finally, during the implementation phase of the trained network, the distance between each new pattern and the established prototypes is calculated, and using a fuzzy classifier of the nearest neighbors, it is assigned the most representative category to the pattern in evaluation.

In Fig. 7 it is shown the steps carried out for the mapping of the input space in categories decision regions using algorithm ART2 proposed by (Vasilic, 2004). Initially, using non-supervised/supervised learning, the space of the training patterns is transferred within a level of initial abstraction that contains a set of groups with the corresponding prototypes, size and category. Later, the groups are fuzzyficated and transformed into an intermediate level of abstraction. Finally, by means of the defuzzyfication, regions of refined decision are established, and a level of final abstraction is obtained.

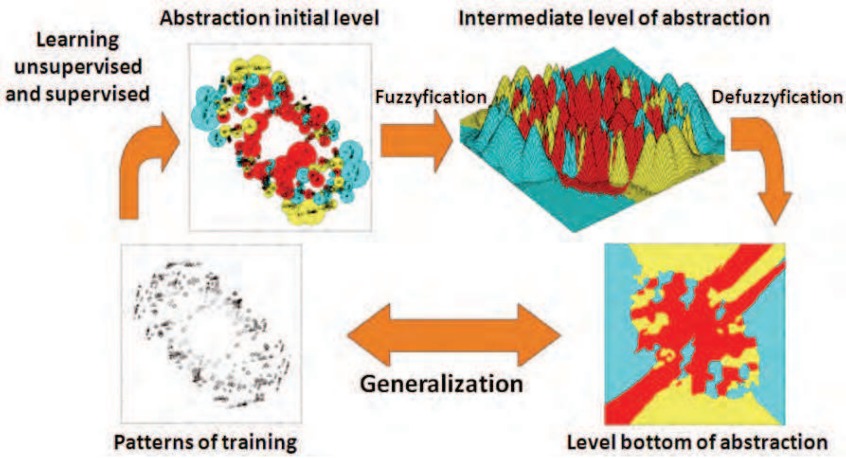


Fig. 7. Mapping of the patterns space using supervise/non-supervised training through the ART2 algorithm. (Vasilic, 2004).

It is observed in Fig. 8, the classification obtained from homogenous groups of the same category using the ART1 methodology (which allows the overlapping of groups and the presence of elements in this zone) and which it is obtained by means of the ART2 methodology (which reduces the radios until no longer elements in the zones of overlaps are present). By means of this modification, the model ART2 tries to improve its performance in relation to the classification error, since it reduces the ambiguity that appears when there are elements in the zones of overlaps that could produce erroneous classification of some pattern. However, it is important to outline that as the radios get more restricted, the network loses capacity of generalization. The final ideal model is a commitment between the needs of precision in the classification with the generalization capacity of the model.

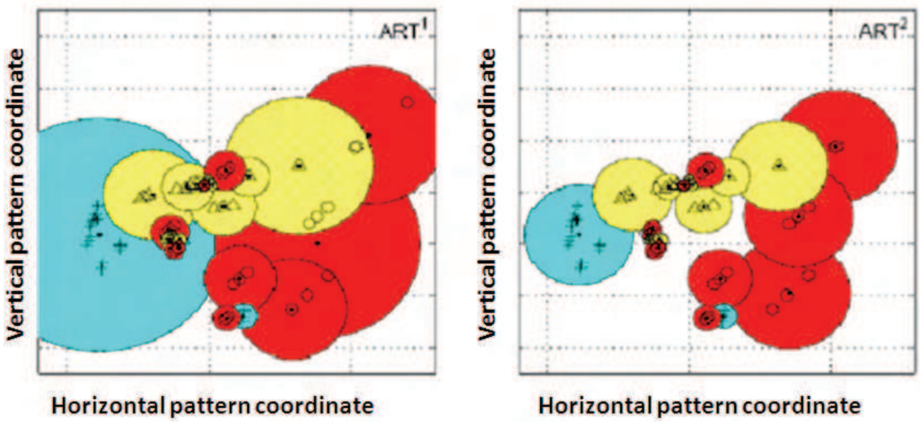


Fig. 8. Comparison between ART1 and ART2 models. (Vasilic, 2004).

3. Art2 model improved

As a contribution of the current research, the model ART2 of (Vasilic, 2004), was improved by introducing a formal methodology for the "reduction of the radios" and by introducing a novel concept denominated "learning on line".

3.1 Formal methodology for reduction of radios

With the purpose of trying to solve the ambiguity that appears when clusters belonging to different categories are present, and with a region of non-zero intersection among them (that is to say, that there are a certain number of training patterns in that region), (Vasilic, 2004), proposes a solution to this problem (ART2) consisting of introducing some rule during the phase of supervised training to construct homogenous clusters that covers solely patterns of exactly one category (valid rule of homogenous intersected clusters), allowing regions of intersection between clusters of different categories as long as patterns do not exist in those regions.

It is outlined in (Vasilic, 2004), the ART2 methodology expressed in natural language, but it is not formally described the algorithm, nor details on its implementation provided. In the current research work, a formal proposal was developed to carry out the implementation of the model classification ART2 and be able to go from the obtained clusters with ART1 to the obtained clusters with ART2, as seen in Fig. 8.

Next, it is presented the procedure used and the formal description of the implemented algorithm. Initially, homogenous intersected valid cluster rule is defined and then the rules to make the reduction of the radios.

3.1.1 Homogenous intersected valid clusters rule

Let to ch be a homogenous cluster of the form [r, P, C, CP], where:

r: is the radius of cluster. r belongs to the real numbers.

P: a vector of dimension n, is the cluster prototype found with the training patterns; each input of this vector belongs to the real numbers set.

C: is the type of cluster, C pertaining to the integer numbers.

CP: is a set of vectors of dimension n where each input of each vector belongs to the real numbers, (training patterns which conform the cluster).

Let to @: [V1 x V2→R] be a function that delivers the Euclidian distance between two vectors, where V1 and V2 are vectors of dimension n.

Let to A be a finite set of homogenous clusters without training patterns in their intersection regions, A = {ch1, ch2, ch3, ..., chn}.

Let to ch1[r1,P1,C1,CP1] and ch2[r2,P2,C2,CP2] be a pair of homogenous clusters of A.

Let to M be the number of patterns in CP1, let to K be the number of patterns in CP2 .

Then:

ch1, ch2 belong to A
IF, AND ONLY IF:
(P1@P2 < r1 + r2) and (C1 ≠ C2) and
FOR EVERYTHING (cpm ϵ CP1) { P2@cpm > r2 } and
FOR EVERYTHING (cpk ϵ CP2) { P1@cpk > r1 } ; k = 1, 2, …, K ; m = 1, 2,.., M

The rule of homogenous intersected valid clusters is fulfilled if the clusters belong to A.
During the supervised training (phase of stabilization), once the homogenous clusters are obtained, it must be verified that no patterns exist inside some region of intersection among the clusters found, that is to say, that the homogenous intersected valid clusters rule is fulfilled; if this rule is not accomplished, the reduction of radios rule should be applied, analyzing the possibility of reducing the radius of any of both clusters in conflict, or of both, if it is necessary, assuring not to exclude any pattern, and having this way the rule fulfilled, to later on add these new optimized clusters to the final set of clusters.

3.1.2 Reduction of radios rule

Let to min (): [Rn → R] be a function that receives a set of real numbers and delivers the minor.
Le to max() : [Rn → R] be a function that receives a set of real numbers and delivers the major.
If ch1, ch2 does not belong to A
It is verified if patterns of ch1 in ch2 exist, and if it is possible the reduction of its radius is done.

```
1   If
2   EXIST (cpm ϵ CP1) { P2@cpm < r2 } ; m = 1, 2, …, M
3   then
4           r2max = min(P2@cpm=1, P2@cpm=2, ..., P2@cpm=M)
5               r2min = max(P2@cpk=1, P2@cpk=2, ..., P2@cpk=K)
6           si r2min > r2max
7           then
8                   r2 = r2max - (r2max - r2min)/L
```

It is verified if patterns of ch1 in ch2 exist, and if it is possible, the reduction of its radius is done.

```
1   If
2   EXIST (cpk ϵ CP2) { P1@cpk < r1 } ; k = 1, 2, …, K
3   then
4               r1max = min(P1@cpk=1, P1@cpk=2, ..., P1@cpk=k)
5         r1min = max(P1@cpm=1, P1@cpm=2, ..., P1@cpm=M)
6         si r1min < r1max
7           then
8                   r1 = r1max - (r1max – r1min)/L
```

Where L is an arbitrary constant inverse to the magnitude in which the radius is reduced. If the given restriction in line 6 is fulfilled the radius can be reduced, and add the cluster to the final set of homogenous clusters. This operation is done for all the homogenous clusters found after the stabilization, and also done against the homogenous clusters that previously have been added in set A.
In Fig. 9 to Fig. 11 it is graphically illustrated what can happen in the intersection of the clusters.
Notice that in Fig. 10 cluster 2 (yellow) cannot reduce its radius since it would exclude the most distant pattern, for this reason these patterns must be part of the following iteration in

the non-supervised & supervised training of ART. Cluster 1 can reduce its radius, and it is included in the set of the final cluster, this is illustrated in Fig. 11.

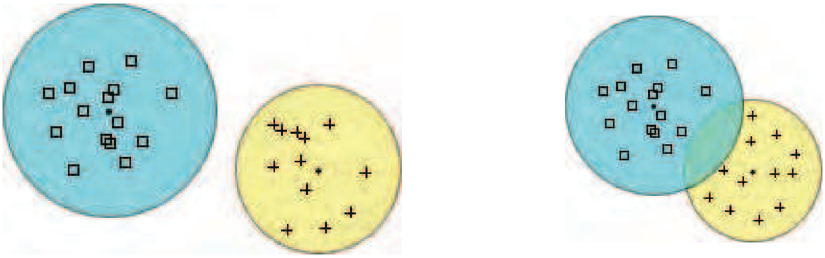


Fig. 9. To the left two homogeneous clusters without intersection. To the right two homogeneous clusters with intersection and without patterns in this region.

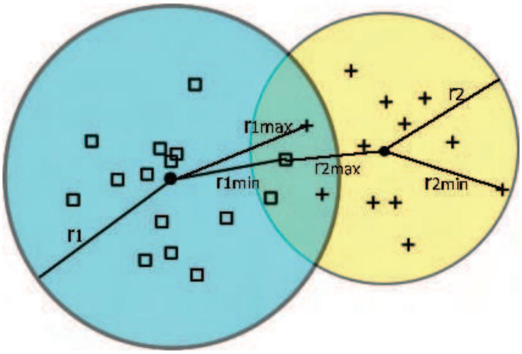


Fig. 10. Two homogeneus clusters with intersection and patterns inside this region.

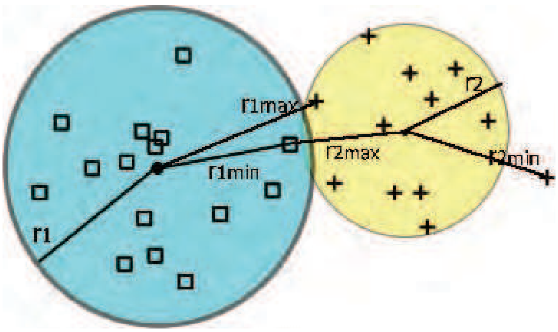


Fig. 11. Reduction of radios is applied to the left cluster and the right cluster is disregarded because it is not possible to do it.

3.2 On-line learning methodology

By means of this technique an adaptive model is obtained, that gradually accommodate its structure to the changes that provide the actual enviroment where this adaptive model develops. That is to say, whenever the algorithm makes an erroneous classification, it will have the opportunity of re-training it and learning on line with the appropriate type provided by the expert. In this way, the algorithm will be learning out more and more from the experiences to improve along its life time. In Fig. 12 to Fig. 15 the used procedure is shown.

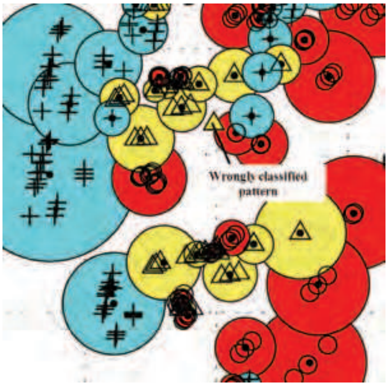


Fig. 12. Erroneous Classification. (Calderón, 2007).

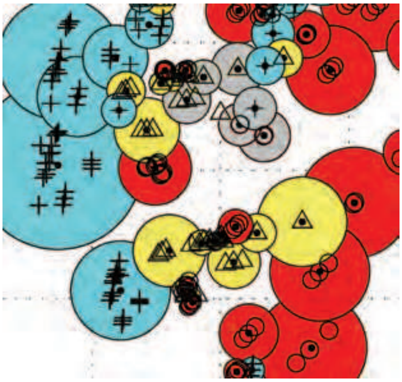


Fig. 13. Selection of the nearest neighboring K-clusters. (Calderón, 2007).

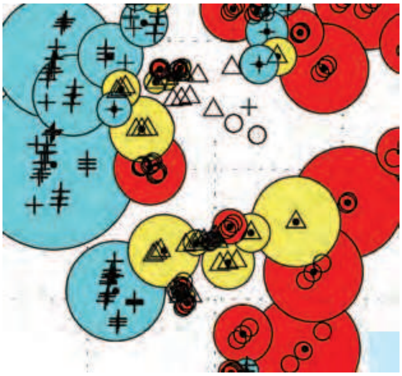


Fig. 14. New set of patterns for re-training. (Calderón, 2007).

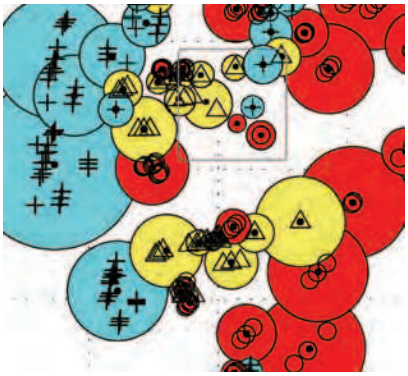


Fig. 15. Resulting set of clusters after on-line re-training. (Calderón, 2007).

Initially, the patterns for the re-training are prepared, extracting the cluster that made the erroneous classification (see Fig. 12) and the nearest neighboring k clusters (see Fig. 13), to form a reduced set of patterns for re-training (Fig. 14). With this set of patterns, the phases of non-supervised learning and supervised learning are implemented again illustrated in Fig. 6. This time the re-training is very efficient now that from the beginning, a subgroup of reduced clusters it is taking into account, all of them homogenous (the time of re-training takes a few seconds).

4. Design of faults classifiers based on neural networks

4.1 Generation of training and validation data

With the purpose of obtaining training samples of the signals of currents of phase and of zero-sequence the tool of simulation ATP was used (Alternative Transient Program) which has been validated at world-wide level as one of the most adapted to analyze electrical power systems (Electric Power Research Institute, 1989), (CanAm EMTP User Group, 1992).
In Fig. 16 the electrical system used for systematic exploration of the considered cases is illustrated.
With the purpose of automatically generating the data file with the ATP cases of variability of conditions of the SEP was developed a module in MATLAB that constructs the ATP format for the sensitivity analysis. Then, this file is run by means of the ATP program to generate the samples of Training, Validation and Checking of the studied models. In Fig. 17 all the flow of information from simulations with ATP and MATLAB until the model of neural network is schematically shown.
Initially, by means of the interface MATLAB-ATP, 508 patterns for training and 246 patterns for validation and checking were simulated. These cases of validation and checking were simulated as intermediate conditions of the training patterns with the purpose of verifying that over-training (validation stage) and the capacity of generalization of the model (checking stage) do not happen. Sensitivity was made on several parameters such as the impedances of source, chargeability of the transmission line, location of the fault, impedance of fault, and the type of fault: mono-phase (A, B, C), two-phase isolated (AB, BC and CA), two-phase to earth (AB-g, BC-g, CA-g) and three-phase (ABC).

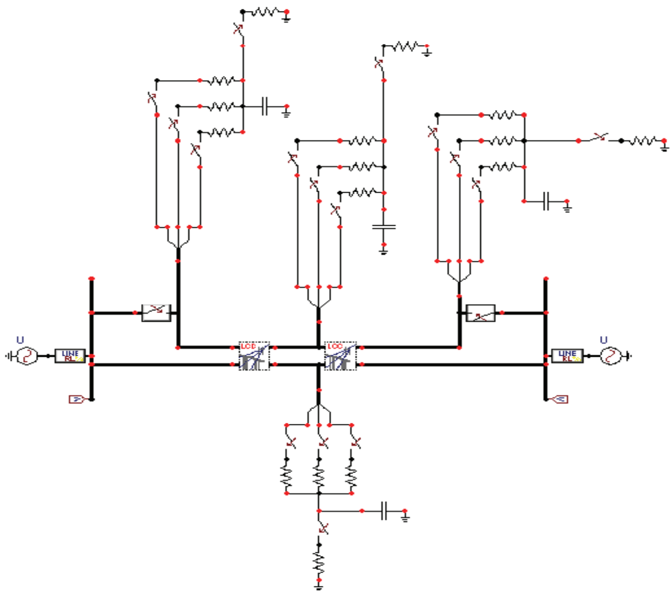


Fig. 16. Typical electrical circuit for analyzing conditions of faults in a SEP. (Alternative Transient Program-ATP).

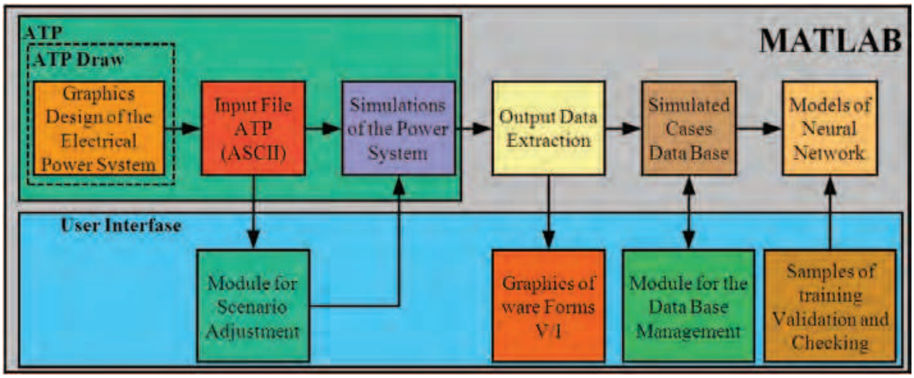


Fig. 17. Integrated software tools for the simulation of electrical power system by means of ATP and MATLAB.

After that, this interface was used to generate 46996 simulated ATP cases of which 36500 were used for training, 5248 for validation, and 5248 for checking. Such the BR model as the ART 2 improved model were verified with these cases.

4.2 Inputs and outputs of the neural networks

The application of a patterns classifier requires first of all the selection of characteristics that contain information necessary to distinguish between the classes, it must be insensitive to the input variability, it must be limited in number to allow efficient calculation, and to limit the amount of required data of training.

As signals of input to the neural network can be selected different parameters from the system. The outputs of the neural network must indicate the type of fault. In general, two types of definition of outputs can be adopted. The first format is that the outputs are compound of A, B, C and G, which indicates that the fault was in the phase a, b or c and there is a connection with earth (G).

A	B	C	G	
0	0	0	0 -	normal condition
1	0	0	1 -	phase to earth fault
1	1	0	0 -	phases a and b without earth fault
1	1	1	0 -	phases a, b and c without earth fault.

The second type has 11 outputs, the first that represents the normal condition and each one of the remain ten is responsible for a type of fault, for example:

1000000000 normal condition.
0100000000 fault phase A.

4.3 Comparison of performance of the classifiers

4.3.1 Size of the neural network

The number of inputs such as the first four neural networks of study as to BR and ART 2 improved model were chosen equal to 7 consisting of three RMS (Root Mean Square) voltages, the three RMS phase currents and RMS zero sequence current. For the number of outputs BP, BR, ART 2 improved and RBF the first type of output was used, for the LVQ the second type and for network MF the output chose like a bi-dimensional Kohonen matrix of dimension 8x8.

As it is well-known, the selection of an optimal number of hidden layers and nodes for a continue BP network being point of research although numerous articles in these areas have been published. For the present study it was only used a hidden layer that turned out to be adapted for this individual application.

For the BP model a hidden layer was used that was to be adapted for this individual application. The number of nodes analyzed was considered from 10 to 16. Finally, with a selection of 12 neurons for the hidden layer a good performance was obtained. This size was used to BR model getting excellent performance too. The size of the matrix for the Kohonen model depends to a great extent on the kind of problem and the availability of training vectors. In this study, a matrix of 8x8 was selected after running a series of simulations and comparing the obtained results.

In order to determine the optimal structure of a RBF, a set of RBF models were trained and validated. In these simulations was carried out an analysis of sensitivity of the number of Kernel nodes varying from 300 to 508 based on the global performance of the network and it was found that with 357 neurons in the hidden layer a suitable performance is obtained.

In the adjustment of LVQ structure, the critical part is the selection of the number of neurons of the layer of Kohonen (competitive layer). In this analysis the total number of training vectors is to be kept in mind and select the number of neurons of the Kohonen layer like a multiple of the number of output nodes. In this study the number of nodes of Kohonen was selected based on the total of training vectors and the number of the eleven outputs. After several simulations were carried out it was found that an optimal number of neurons for the hidden layer of Kohonen for this application are 150.

Finally, the structure of ART 2 model is so different than the previous models and in this case the final number of clusters is automatically determined according to algorithm and is dependent of the threshold assigned by the user.

4.3.2 Learning process

In the study of BP network and BR model were used the functions of MATLAB tansig transference for the neurons of the hidden layer and the function purelin for the output linear layer. The learning factor that controls the rate of convergence and stability was chosen equal to 0.05. Initially, all the weights and bias were put in small random values. The input values were presented to the network and the output variables were specified. The training process took place until the value of error RMS (Root Mean Square) between the real output and the wished output reached an inferior acceptable value of 0.1. For the Kohonen layer was chosen equal to 0,9 for the phase of initial ordering and 0,02 for the final phase of tuning. 1000 steps were simulated in the ordering phase and it was considered a neighboring radius of 1 for tuning phase. A grille form was used for the space distribution of the neurons in the matrix of two dimensions.
The parameters of the units of RBF network were determined by means of the function newrb of MATLAB. First a subtractive grouping of the data by means of the function subclust of MATLAB with the purpose of estimating an average spread that complied most of the considered data was carried out. From this analysis was obtained a spread value of 0.19. Later, the centers of the radial units were determined by means of an algorithm of adaptive group that uses the function dist of MATLAB combined with a parameter of bias=0.833/spread. Once the centers are determined the algorithm newrb of MATLAB makes an iterative procedure of addition of new neurons until obtaining an error adapted between the real outputs of the model and targets of training assigned.
In LVQ network the function newlvq of MATLAB was used. The algorithm newlvq constructs an LVQ neural network like the one presented in Fig. 4. It is used as input criterion to consider the percentage of samples that each class has. For example, in this study 10 classes corresponding to 10 conditions of fault were considered. For the 508 samples each fault condition has associated 50 samples (0,1 p.u). All together, the sum must be equal to 1 p.u (100% of the 508 samples). For the learning process the function learnlv1 of MATLAB was used considering a learning rate of 0.01.
The ART 2 model training was explained in detail above and is depicted in Fig.6.

4.3.3 Training and validation error

The error often is used like a criterion to finish the learning process. It has been possible to find that, for a given set of training data and structures of network, the error of minimum learning that can be reached is similar for all the networks. Nevertheless, the time to reach the value of the error (speed of learning) is entirely different (Song et al., 1997). It is important to notice that obtaining the smaller error during the learning does not necessarily imply the best performance of the network. That is, there must be commitment between learning error and error during the validation phase.

4.3.4 Precision of classification

Initially, the BP, FM, RBF and LVQ neural networks trained were validated with 246 cases generated by means of ATP program under several conditions of the system and intermediate conditions of fault to the 508 cases considered in the training.

From the results obtained from the research for networks analyzed it was possible to observe that the rates of error vary with respect to the type of fault. As it was expected, the classification error is greater for the faults phase-phase without earth, which is the type of fault more difficult to detect.
The network that had the smaller error of classification was the RBF (only a 7% of the validation cases did not have suitable classification). In Fig. 18 the results of the training of RBF network are shown, where it can be noticed that the error between what is wished and what is real is very low.

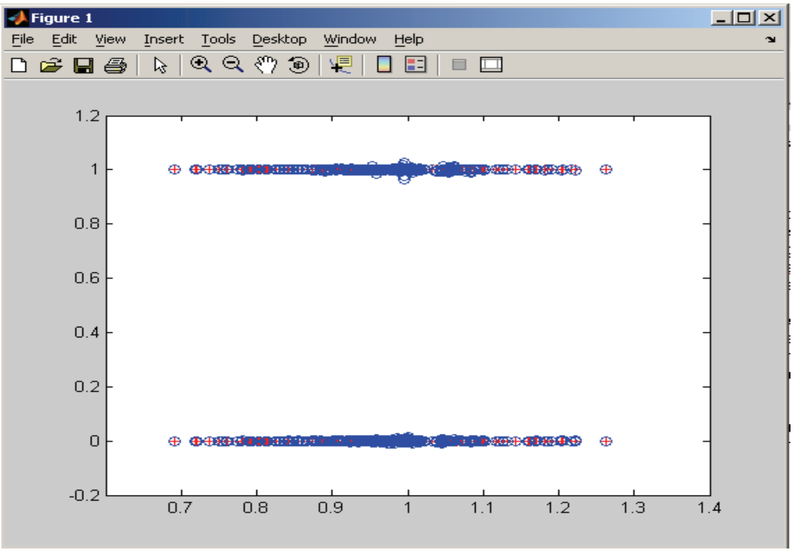


Fig. 18. Vector target (o) vs. Network output (+) during the training of RBF network. (Matlab educational license).

Although the previous models are, in general, good for classification purposes, these had some difficulties when certain conditions of the electrical system were considered (for example, high impedances faults). In some of these cases, the classification error was not suitable.
Due to this, based in the previous results, the research was oriented in the search of a hybrid model that was able to adapt itself to many expected conditions from the electrical power system and at the same time had a low classifcation error, and high level of generalization. The BR and ART 2 improved models were developed and then trained and validated using the methology described and ilustrated in Fig. 6.
By using the BR model the training error was 0% for the 36500 cases considered, 0.74% for the 5248 validation cases and 1,39% for the 5248 checking cases.
By using the ART 2 model the training error was 0.1% for the 36500 cases considered and 3.7% for the 10496 validation and checking cases.

4.3.5 Robustness

For many reasons, it is not possible to assume that the cases presented to the classifier during the phase of application are complete and precisely represented by the training set. It

is particularly certain in the classifiers of fault of transmission lines. The patterns of training are normally limited, and in most cases are generated by means of simulation in computer that does not exactly match the real data of field. In general the data that enter the algorithms will be affected by the transducers and the noise from the atmosphere. Also, the parameters of the power system and the conditions change continuously. Thus, then actually a good robustness of the trained classifier is required. It includes deviations in the measurements and superposed white noise. The rates of undesired classification are considerably greater in BP network for the different cases considered, whereas the error rates for FM, RBF, LVQ and ART 2 improved neural networks increased moderately or very little. This is due to the purely supervised nature of BP network. The surfaces of decision of BP networks can take non intuitive forms because the space regions that are not occupied by the training data are classified arbitrarily. One way to improve this problem could be to combine BP with BR method (Bayesian regularization) in order to reduce the classification errors. Instead, the other networks analyzed are governed by non-supervised learning in which the regions of the input space occupied by the training data are not classified according to the proximity that commonly exists among the training data.
In summary, it is important to underline that a classifier has to be evaluated by its time of training, error rate, calculation, adaptation, and its real time implementation requirements. At the time of making a decision related to the selection of a network in particular it must be taken into account the combination of all these aspects and the possibility of considering new changes in the algorithms that allow improvement of the performance for the specific applications that are considered.

5. Conclusions and future work

The design of power systems protections can be essentially treated like a problem of pattern classification/recognition. The neural networks can be used like an attractive alternative for the development of new protection relays as much as the complexity of the electrical power systems grows. Different strategies of learning have to be explored before adopting a particular structure to a specific application, and establishing a commitment between the off-line training and the real time implementation.
In general, the combined non-supervised/supervised learning techniques offers better performance than the purely supervised training. In the present study it was possible to verify that FM, RBF and LVQ networks have a greater speed of training, similar error rate, better robustness to consider variations of both the system and the environment, and require much less amount of training data compared with BP network (Song et al., 1997). . On the other hand, the BP network is more compact and it is hoped to be faster when it is placed in operation under the real time performance.
This study, additionally showed, that in spite of those models have good performance to classify faults in electrical power systems in some special cases (for example, high impedances faults) the resultant error is not suitable. In order to take this fact into account, it is necessary to consider BP with BR or ART 2 improved models which resolve this kind of conflict.
It is important noticing that the present study focused in the performance of different models of neural networks applied to the classification of faults in electrical power systems.

Nevertheless, for the effects of being considered as protection alternatives of electrical power systems the techniques presented have to be integrally evaluated, considering in addition several practical issues. For example, it has to be combined with real field tests and the implementation of corresponding hardware.

6. Acknowledements

This work presents the results of the researches carried out by the National University of Colombia and the company Interconexión Eléctrica S.A. E.S.P (ISA), as partial advance of the research project co-financed by COLCIENCIAS entitled: "Computer Science Tools for the Automatic Diagnosis of Events in Transmission Lines of Electrical Power".

7. References

Aggarwal, R. and Song, Y. (1997). Artificial neural networks in power systems: Part I - General introduction into neural computing, *Power Engineering Jour.*, vol.112 11, no. 3, pp. June 1997,129–134.

Aggarwal, R. and Song, Y. (1998a). Artificial neural networks in power systems: Part II - Types of artificial neural networks, *Power Engineering Jour.*, vol. 12, no.1, Feb. 1998, pp. 41–47.

Aggarwal, R. and Song, Y. (1998b). Artificial neural networks in power systems: Part III - Examples of applications in power systems, *Power Engineering Jour.*, vol. 12, no. 6, Dec. 1998, pp. 279–287.

Badrul, H. ; Choudhury and Kunyu, Wang. (1996). Fault Classification Using Kohonen Feauture Mapping, *IEEE, Electrical Engineering Department*, University of Wyoming.

Calderón, J. A. (2007). Artificial Intelligence Adaptive Model for the Automatic Faults Diagnosis using oscillate-graphics records. *Thesis - Magister on Computer Science*, Computer Science Department, National University of Colombia – Campus Medellin.

Calderón, J. A. Ovalle ,D and Zapata, G. (2007). Comparative Analysis between Models of Neuronal Networks for the classification of faults in Electrical Systems. *IEEE CERMA México.*

CanAm EMTP User Group. (1992). Alternative Transients Program (ATP) - *Rule Book*, Portland, OR, Sep. 1992.

Dalstein, T. and Kulicke, B. (1995). Neural network approach to fault type classification for high speed protective relaying, IEEE Trans. Power Delivery, vol. 10, no. 2, Apr. 1995, pp. 1002–1011.

Dalstein, T.; Friedrich, T. ; Kulicke, B. and Sobajic, D. (1996). Multi neural network based fault area estimation for high speed protective relaying, *IEEE Trans. Power Delivery*, vol. 11, no. 2, Apr. 1996, pp. 740–747.

Dillon, T.S. & Niebur, D. (1996). Artificial Neural networks Applications in Power Systems. CRN Press.

Dillon. T. and Niebur. D. (1999). *Tutorial on artificial neural networks for power systems*. Engineering intelligent Systems for Electrical Engineering and Communications, v 7, n 1, pp. 3-17.

Electric Power Research Institute. (1989). Electromagnetic Transient Program (EMTP) - Rule Book, EPRI EL 6421-1, CA, June 1989, Palo Alto.

El-Sharkawi, M.A. & Niebur, D. (1996). Application of artificial neural networks to power systems. IEEE Tutorial Course Text 96TP112-0, Piscataway, NJ.

Fernandez, A. L. O. and Ghonaim, N. K. I. (2002). A novel approach using a FIRANN for fault detection and direction estimation for high-voltage transmission lines, *IEEE Trans. Power Delivery*, vol. 17, no. 4, Oct. 2002, pp. 894–900.

Foresee, F.D. and Hagan, M.T. (1997). Gauss-Newton approximation to Bayesian regularization, *Proceedings of the 1997 International Joint Conference on Neural Networks*, pp. 1930-1935.

Hagan. M.T. ; Demuth. H.B.; De Jesus. O. (2002). An introduction to the use of neural networks in control systems. International Journal of Robust and Nonlinear Control, v 12, n 11. pp. 959-85

Keerthipala, W.W. L. ; Wai, C. T. and Huisheng, W. (1997). Neural network based classifier for power system protection, *Electric Power Systems Research*, vol. 42, no. 2, Aug. 1997, pp. 109–114.

Kezunovic, M. (1997). A survey of neural net applications to protective relaying and fault analysis, *Engineering Intelligent Systems for El. Engr. and Comm.*, vol. 5, no. 4, Dec. 1997, pp. 185–192.

MacKav. D.J.C.; Takeuchi. R. (1998). (Cavendish Lab., Cambridge Univ., UK). *Statistics and Computing*, v 8, n 1, .pp. 15-23.

Oleskovicz, M.; Coury, D.V.; Aggarwal, R.K. (2001). A complete scheme for fault detection, classification and locationin transmission lines using neural networks. Developments in Power System Protection, 2001, Seventh International Conference on (IEE). vol , Issue , pp.335 – 338.

Ranaweera, D. K. (1994). Comparison of neural network models for fault diagnosis of power systems, Electric Power Systems Research, vol. 29, no. 2, Mar. 1994, pp. 99–104.

Sidhu. T.S. and Mitai. L. (2000). Rule extraction from an artificial neural network based fault direction discriminator. *2000 Canadian Conference on Electrical and Computer Engineering. Conference Proceedings*. Navigating to a New Era (Cat. No.00TH8492), pp. 692-6 vol.2

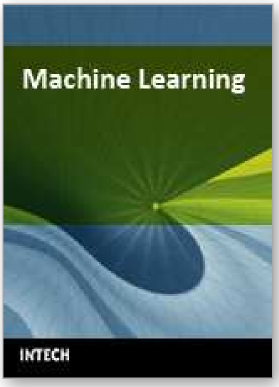
Song, Y. H. ; Xuan, Q. X. ; Johns, A. T. (1997). Comparison studies of five neural network based fault classifiers for complex transmission lines. *Electric Power Systems Research*, vol. 43, no. 2, (Nov. 1997), pp. 125–132.

Song, Y. H. ; Xuan, Q. X. and Johns, A. T. (1997). Comparison studies of five neural network based fault classifiers for complex transmission lines, *Electric Power Systems Research*, vol. 43, no. 2, Nov. 1997, pp. 125–132.

Song, Y.H., Johns, A.T. and Aggarwal, R.K. (1996). Applications in Power Systems, Science Press and Kluwer Academic, *Computational Intelligence*.

Vasilic, S. (2004). Fuzzy neural network pattern recognition algorithm for classification of the events in power system networks. *Doctoral dissertation*, Texas A&M University. Available electronically from http : / /handle .tamu .edu /1969 .1 /436.

Zahra, F. ; Jeyasurya, B. and Quaicoe, J. E. (2000). High-speed transmission line relaying using artificial neural networks, *Electric Power Systems Research*, vol. 53, no. 3, Mar. 2000, pp. 173–179.



Machine Learning

Edited by Abdelhamid Mellouk and Abdennacer Chebira

ISBN 978-953-7619-56-1
Hard cover, 450 pages
Publisher InTech
Published online 01, January, 2009
Published in print edition January, 2009

Machine Learning can be defined in various ways related to a scientific domain concerned with the design and development of theoretical and implementation tools that allow building systems with some Human Like intelligent behavior. Machine learning addresses more specifically the ability to improve automatically through experience.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jhon Albeiro Calderón, Germán Zapata Madrigal and Demetrio A. Ovalle Carranza (2009). Performance Analysis of Hybrid Non-Supervised & Supervised Learning Techniques Applied to the Classification of Faults in Energy Transport Systems, Machine Learning, Abdelhamid Mellouk and Abdennacer Chebira (Ed.), ISBN: 978-953-7619-56-1, InTech, Available from:
http://www.intechopen.com/books/machine_learning/performance_analysis_of_hybrid_non-supervised__amp__supervised_learning_techniques_applied_to_the_cl

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821