Breaker status uncovered by autoencoders under unsupervised maximum mutual information training

Vladimiro Miranda, *Fellow, IEEE*, Jakov Krstulovic, Joana Hora, Vera Palma
and José C. Príncipe, Fellow, IEEE

***Abstract* – This paper introduces a more efficient procedure for the training of autoassociative neural networks, The autoencoder is split along its hidden layer and its bottom half trained in unsupervised mode, maximizing a mutual information criterion, while the top half is trained in supervised mode. Tests with the identification of breaker status illustrate the effectiveness of the approach.**

***Index Terms* – Power System Topology, Autoencoders, Neural Networks, Information Theoretic Learning.**

I. Introduction

THE knowledge of breaker status is essential in any control center for a number of reasons, from sheer security to allowing further system analyses, starting with the most simple power flow or basic state estimation. However, the estimation of the status of a breaker has relied mostly on:

- signals remotely collected and made available at the SCADA system, taken as true values, or
- heuristic rules applied when such signals are missing.

Some past approaches to estimate breaker status relied on mathematic models, other relied on neural networks. The results, for one reason or another, including practical feasibility or confidence, never had wide acceptance.

This paper explores a model of neural networks in a completely new way: the adoption of a competitive autoencoder scheme. This scheme has already proved to be effective, but a 100% accuracy was not attained. The paper deals with improving the accuracy of this scheme, to make ground for its industrial acceptance.

Autoencoders, or autoassociative neural networks, have a special architecture – the output dimension is equal to the input – and are trained to optimize a general objective: to reproduce, in the output, the input vectors belonging to a particular data cluster used in their training. Two basic models of autoencoders may be identified: in one case, there is a hidden layer with a dimension smaller that the input/output; the other case refers to architectures with all hidden layers larger than the input/output layers. Without loss of generality, this paper will refer to autoencoders of the first family.

This work is partially funded by the ERDF from the EU through the Programme COMPETE and by the Portuguese Government through FCT - Foundation for Science and Technology, project ref. LASCA PTDC/EEA-EEL/104278/2008.

V. Miranda (vmiranda@inescporto.pt), J. Krstulovic (jopara@inescporto.pt), J. Hora and V. Palma are with INESC TEC (INESC Technology and Science, coordinated by INESC Porto), Portugal. V. Miranda and J. Krstulovic are also with FEUP, Faculty of Engineering of the University of Porto, Portugal. J. Krstulovic is also with the University of Split, Croatia. J. Príncipe is with the Computational NeuroEngineering Laboratory, University of Florida, Gainesville, FL 32611 USA.

As neural networks, it is usual to think that methods applied to the training of general neural networks, such as backpropagation, should be applied also to autoencoders, minimizing a function of the input-output error. This idea is correct, yet it does not take advantage from the special properties of autoencoders nor is it based on a particular interpretation of the data transformation phenomena occurring in the autoassociative neural network.

This paper illustrates a remarkable phenomenon: it is possible to split the autoencoder in two halves, train the first part in unsupervised mode (so, no guidance exists in how the weights should evolve to produce a particular result or minimize any kind of error), then tune the second half keeping the first half fixed – and obtain a better result (better accuracy) than training the complete network in the traditional fashion.

This is actually an important technical advancement: training half of a neural network is easier and less computationally costly and because the half-network depth is half of the original one, better accuracy in weight adjustment may be achieved (the problem of adjusting first layer weights in backpropagation methods is well known).

This advancement will be illustrated in an example built to represent the estimation of the status (open or closed) of a breaker via a pair of autoencoders organized in a competitive scheme, i.e. with each autoencoder tuned to a particular case and then, in each particular test, checking which autoencoder produces a smaller error to decide the breaker status.

II. Autoencoders

Auto-associative neural networks or autoencoders are feedforward networks that are built to mirror the input space S in their output. Therefore, such a network has an input vector of the same size as the output and is trained to display an output equal to its input. This is achieved through the projection of data onto a different space S' (in its middle layer) and then re-projecting it back to the original space S. To achieve this, a trained autoencoder stores in its weights information about the training data manifold.

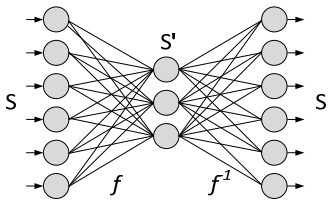

Fig. 1. Architecture of a 6-3-6 autoencoder with a single hidden layer

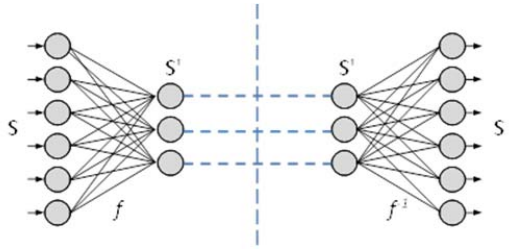

Fig. 2. Autoencoder split in two halves. The first half projects data from S to S'. The second half operates the inverse function.

Fig. 1 shows a simple neural network with a smaller single middle layer. This simple architecture is frequently adopted because networks with more hidden layers have proved to be difficult to train [1], although allowing increasing accuracy. If an autoencoder is split along its middle layer, as in Fig. 2, the first half achieves data compression.

Using the information theory concepts, an ideal autoencoder could achieve perfect data compression with minimal information loss as shown by the channel capacity theorem and rate distortion theories. This means that an optimal information flow would be present throughout the neural network, namely at the compressed middle layer.

This compression may also be seen as a process of feature reduction. In this sense, it was proven that linear autoassociative neural networks (i.e. with linear activation functions in their neurons) achieve compression in the sense of PCA (Principal Component Analysis) [2]. Therefore, the weight matrix for their first half may be calculated from a PCA and the weight matrix of their second half is the transpose of the first matrix. This avoids the need for training.

However, this does not hold for non-linear autoencoders [3]. PCA, as it is well known, projects data in a space spanned by orthogonal axes (the directions of the most important eigenvectors of the data matrix). This is no longer true with non-linear autoassociative neural networks.

III. Breaker status estimation

Single breaker status estimation is a sub-problem of the larger problem of topology estimation, which in turn is a sub-problem of the state estimation problem. One of the most important approaches relying on a mathematical formulation certainly is the so called generalized state estimation [4], which departed from mathematical definitions of open breaker branch (zero current) and closed branch (zero voltage drop) and included them in the state estimation formulation. Another proposal [5] was based on binary variables associated to breakers, represented by quadratic expressions with (0 or 1) solutions and added to a classic state estimation mathematical model. Other interesting models were proposed based on Lagrange multipliers associated to constraints related to topology [6][7][8]. A more recent approach [9] proposed a linearized mathematical model with relaxation of the binary variables associated to with breaker status to the continuous interval [0,1] to produce a network topology estimation.

In a different line, some authors have proposed the adoption of feedforward neural networks [10][11][12][13][14]. In all these cases, neural networks of the type were used. The classical approach may be summarized as follows: a neural network with a single output neuron is trained in supervised mode to generate a binary output, indicating if the input pattern corresponds to an open or closed breaker. This approach requires an external parameter to be arbitrarily fixed: a threshold splitting the numerical output in two sets, each associated to a given status diagnosis.

The original work in [15] proposed autoencoders with a single hidden layer as a means to recover missing breaker status signals, in the sense of a technique used in missing sensor signal restoration. However, in [16] and for the first time the model for missing signal restoration was replaced with advantage by a model of competitive autoencoders. Instead of trying to reconstruct an input signal, this model is based on the following idea:

1. An autoencoder is trained to learn a pattern constituted by vectors of electric values present in the network when the breaker is closed – and another trained to learn the pattern for the case with the breaker open.
2. An input vector with values depending on an unknown status will either belong to one or the other cluster.
3. One of the autoencoders should reproduce "correctly" the input vector, as this will belong to the pattern that the autoencoder has learned; the other should display a larger input-output error for the opposite reason.
4. Thus, selecting the autoencoder with min error is tantamount to identifying the pattern and therefore the breaker status associated with the input vector.

Among other advantages, this process has the clear superiority of not requiring external definitions of parameters or thresholds to separate solutions.

IV. Unsupervised training

If an optimal information flow must occur throughout an autoassociative neural network, one should then be able to train the first half of the autoencoder under some criterion that would optimize information throughput, and achieve a good weight matrix for this sub-network. As one does not know how the information should be organized in the compressed space, this amounts to a case of unsupervised training. The specification of the cost function is now discussed.

In order to define the unsupervised training criterion to be used, some basic concepts in Information Theoretic Learning (ITL) are first presented. These turn around the concept of entropy as a measure of information quantity in a distribution.

A. Renyi's Entropy

Entropy H, as a measure of information content of a probability distribution $P = \{p_1, p_2, \ldots, p_n\}$ of a random variable X, was defined by Renyi [17] as

$$H_{R\alpha} = \frac{1}{1-\alpha}\log\sum_{k=1}^{N} p_k^{\alpha} \quad \text{with} \quad \alpha > 0, \alpha \neq 1 \tag{1}$$

This represents is a family of functions $H_{R\alpha}$ depending on a real parameter α. Renyi's quadratic Entropy is ($\alpha = 2$)

$$H_{R2} = -\log\sum_{k=1}^{N} p_k^2 \tag{2}$$

This definition can be generalized for a continuous random variable Y with pdf $f_Y(z)$:

$$H_{R2} = -\log \int_{-\infty}^{+\infty} {f_Y}^2(z)dz \qquad (3)$$

We can see that Renyi's Entropy, with its sum of probabilities, is much more amenable to algorithmic implementation than Shannon's Entropy with its sum of weighted logarithms of probability. Representing the pdf f(z) induced by a discrete sample by a sum of Gaussian kernels (the Parzen windows technique [18]) achieved, in the frame of ITL, algorithmic solving feasibility.

B. The ITL Maximum Entropy criterion

Combining Renyi's definition of Entropy with a Parzen window estimate of the pdf, we reach an Entropy estimator for a continuous valued data points $\{\mathbf{y}\}$ as

$$H_{R2}(\mathbf{y}) = -\log \int_{-\infty}^{+\infty} \hat{f}_Y{}^2(\mathbf{z})d\mathbf{z} = -\log V(\mathbf{y}) \text{ , with} \qquad (4)$$

$$V(\mathbf{y}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} G(\mathbf{z}-\mathbf{y}_i, \sigma^2\mathbf{I})G(\mathbf{z}-\mathbf{y}_i, \sigma^2\mathbf{I})d\mathbf{z} \qquad (5)$$

In this expression we recognize the convolution of Gaussian functions, which has the following interesting result:

$$V(\mathbf{y}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(\mathbf{y}_i - \mathbf{y}_j, 2\sigma^2\mathbf{I}) \qquad (6)$$

$V(\mathbf{y})$ is the *information potential (IP)* of the data set. When the objective is to maximize H, one can instead minimize the information potential V. So, Min V becomes the cost function for the unsupervised training of a mapper with maximum output Entropy [19]. It is the Maximum Entropy criterion. The rationale behind its application to autoencoder training is the following: assuming that the input noise is small, if one wishes to maximize the information throughput along the neural network, then a maximum of information should flow through the middle layer. This assures that the best projection is achieved from S to S' (minimum information loss).

C. Mutual information

A different way to look at the unsupervised training of half autoassociative neural networks is to reason that one should impose a training cost function based on maximizing the transfer of information between the input and the middle layer output. To achieve this, the concept of mutual information becomes useful [20]. Considering two independent random variables X and Y, and their joint probability distribution (X,Y), the additive property of entropy holds:

$$H(X,Y) = H(X) + H(Y) \qquad (7)$$

The interpretation for this is the following: independence means that each random variable contains no information on the other variable. When there is some form of dependency, then the following holds:

$$H(X,Y) = H(X) + H(Y) - I(X,Y) \qquad (8)$$

where I(X,Y) is called the Shannon Mutual Information between X and Y, or shared by X and Y.

If two distributions are from independent variables, their joint distribution is given by the product of the marginal distributions. Mutual Information can thus be defined as the divergence between the joint distribution and the product of marginal distributions [20] – if the product and the joint are equal, the variables are independent and the Mutual Information is zero.

Maximizing Mutual Information between input and middle layer output is therefore a criterion that may be used in the unsupervised training of half autoencoders. This will tend to tune the weights in such a way that the information at the output coincides with the one present in the input data.

Not surprisingly, under some circumstances maximizing the input-output mutual information is equivalent to maximizing the entropy at the output. Therefore, the choice of the formulation to be implemented will depend on the simplicity of programming and the computational cost associated to its practical calculation.

D. The ITL Maximum QMI criterion

The Cauchy-Schwartz (CS) distance between two pdf f and g is given by [20]

$$I_{CS}(f,g) = \log \frac{(\int f^2(x)dx)(\int g^2(x)dx)}{(\int f(x)g(x)dx)^2} \qquad (9)$$

This definition is based on the well-known Cauchy-Schwartz inequality. When this distance is taken between the joint and the product of the marginal distributions of two random variables, one has

$$I_{CS}(X,Y) = \log \frac{\left(\int\int {f_{X,Y}}^2(x,y)\,dx\,dy\right)\left(\int\int {f_X}^2(x){f_Y}^2(y)\,dx\,dy\right)}{\left(\int\int f_{X,Y}(x,y)\,f_X(x)f_Y(y)dx\,dy\right)^2} \qquad (10)$$

This divergence measure can be seen as the Quadratic Mutual Information (QMI). Clearly, if the distributions f and g are independent, no mutual information is present and the QMI is zero.

The QMI expression is composed of three terms:

$$\begin{aligned} V_J &= \int\int {f_{X,Y}}^2(x,y)\,dx\,dy \\ V_M &= \int\int \left(f_X(x)f_Y(y)\right)^2 dx\,dy \\ V_C &= \int\int f_{X,Y}(x,y)\,f_X(x)f_Y(y)dx\,dy \end{aligned} \qquad (11)$$

where V_J is the *IP* of the joint PDF, V_M is the *IP* of the factorized marginal PDF and V_C is the generalized cross *IP*. Then, the QMI becomes given by:

$$I_{CS}(X,Y) = \log V_J - 2\log V_C + \log V_M \qquad (12)$$

The estimators for these terms are

$$\hat{V}_J = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{V}_1(i,j)\;\hat{V}_2(i,j) \qquad (13)$$

where $\hat{V}_k(i,j) = G_{\sqrt{2}\sigma}(x_k(i) - x_k(j))$

$$\begin{aligned} \hat{V}_M &= \hat{V}_1\hat{V}_2 \\ \text{where} \quad \hat{V}_k &= \tfrac{1}{N^2}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{V}_k(i,j) \;, k=1,2 \end{aligned} \qquad (14)$$

$$\hat{V}_C = \frac{1}{N}\sum_{i=1}^{N} \hat{V}_1(i)\hat{V}_2(i) \qquad (15)$$

where $\hat{V}_k(i) = \frac{1}{N}\sum_{j=1}^{N} \hat{V}_k(i,j)$, $k=1,2$

Accordingly, the CS estimator of the QMI is defined as

$$\hat{I}_{CS}(X,Y) = \log\hat{V}_J - 2\log\hat{V}_C + \log\hat{V}_M \qquad (16)$$

Maximizing this estimator constitutes the MQMI criterion.

V. Autoencoder Training

A. First half unsupervised raining

The unsupervised training of the first half of the autoencoder assures a maximum information flow onto the middle layer. The success of this training phase depends strongly on the iterations departing point. A possible strategy is the random initialization of the weights, but we've devised a strategy to initialize weights that proved to be much more robust and consistently leading to good results: it is to depart from the PCA point. Therefore, one determines the initial weights as if assuming to be dealing with a linear neural network (calculating the eigenvectors of the input data). The iterations to adjust the final weights depart from that point.

B. Second half supervised training

Once the weights of the first half are calculated, the supervised tuning of the weights of the second half is straightforward. A vector is input to the first half (already trained and with weights fixed), its output is fed into the second half under training. The error between the new output and the vector input helps to control the training accuracy.

VI. Competitive Autoencoders

Fig. 3 illustrates the idea of the competitive autoencoder scheme: given two clusters of data, one may train separate autoassociative neural networks to learn each cluster – each network will learn a manifold supporting the data with minimum error. A new point belonging to one of the clusters will display a small error to one of the manifolds (and hopefully a larger error to the other). This means that one of the autoencoders will "recognize" it as belonging to the cluster it has learned. The other autoencoder, however, will not be able to reproduce the point with accuracy, because it has learned another way to project (to middle layer space S') and project back (to output layer S) the input vectors.

This principle allows one to organize "diagnosis machines" that perform identification in a competitive mode: a number of autoencoders equal to the number of data clusters are trained and then set up in parallel. When facing new data, the winner (displaying the smallest error) is taken as indicating to which cluster does the new vector belong.

Fig. 4 shows the parallel arrangement of two autoencoders to detect the one producing the smallest error. This technique may be extended to more clusters. In [21], seven autoencoders were organized in parallel to build an extremely successful diagnosing system for incipient faults in power transformers.

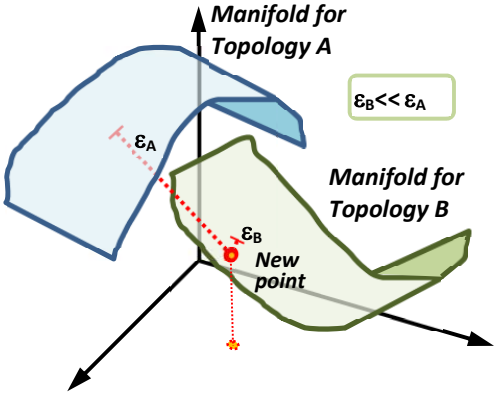

Fig. 3. Illustration of two manifolds learned by two autoencoders A and B. When a new point is input, one of the autoencoders will display a larger error because it cannot recognize the new point as close to the structure learned.

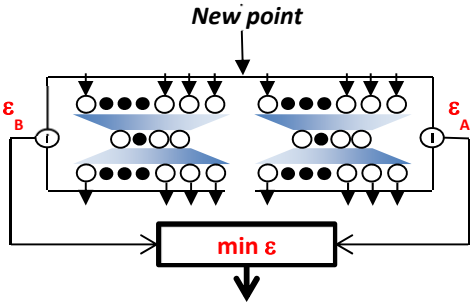

Fig. 4. Parallel competitive autoencoder scheme and the detection of the smallest error produced when the same signal is input to both.

VII. Application to Breaker Status Identification

This idea may be applied to breaker status identification, if one assumes that electric data in some significant neighborhood of the breaker form two clusters: one corresponding to the status of open breaker, the other to closed breaker. Then, two trained autoencoders might be able to perform status recognition, in the absence of any signal indicating such status. This idea has been tested with considerable success and results presented in [16].

Two factors were identified that condition the quality of the status identification: the constitution of the data clusters and the quality of the autoencoder training. In the following, one will not discuss the first item and will assume that the sets of variables to form the electric data clusters have been adequately identified and selected.

In the following, results will be presented from tests applied to a section of the IEEE RTS 24 bus system [22]. To this network, breakers were added in different places (see Fig. 5) and power flow data were collected from 10,000 experiments under the following conditions:

a. Designing of a cumulative load curve with data from [22]; then sampling load levels and constructing scenarios from valley to peak of the load curve.
b. Simulation of load variation by adding a Gaussian perturbation with σ = 5%.
c. Generation of a large set of power flow results with an OPF, with breaker status randomly defined.

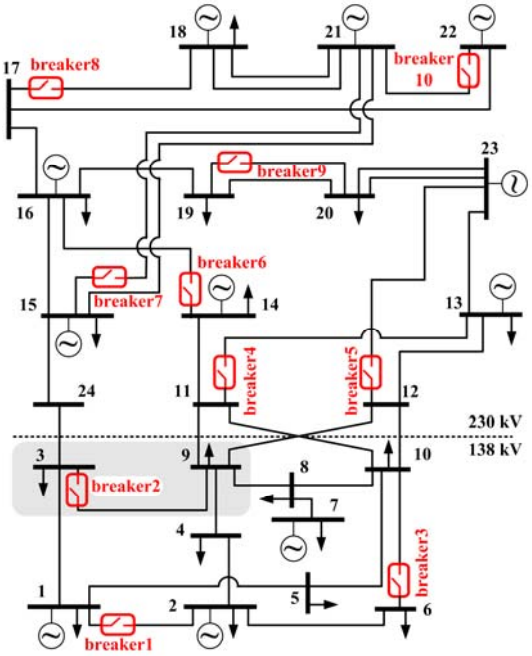

Fig. 5. IEEE RTS 24 with indication of the branches where 10 switches were introduced and the area of data collection for breaker 2 status identification.

d. Simulation of noisy measurements by adding a Gaussian perturbation to power flow solutions, with 3σ = 1% of the largest value added to power injections – a large value if one takes in account only the precision class of the measuring devices.

The object of testing in this paper is breaker 2, placed in the line between buses 3 and 9. Two clusters of local electric data (Fig. 5) were organized (cases of breaker 2 open and closed).

A. *Testing the quality of the training of a single autoencoder*

The first experiment will compare the quality of the training achieved for an individual autoencoder with only a single hidden layer (scheme used in all experiments reported):

a. Trained in a *single step* with a classical backpropagation algorithm under a MSE criterion.
b. Trained in *two steps*: 1 – Unsupervised training of the first half under a maximum entropy equivalent criterion; 2 – Supervised training of the second half under a MSE criterion.

We compare the tests made in training two autoencoders, one for the case of breaker 2 *open* and the other for breaker 2 *closed* (see Fig. 5). The results compare the classical single step global brackpropagation training (using a PROP algorithm) against our two-stage approach and are summarized in Table 1. Ten rounds with distinct seeds were done, and the best result selected. The model with unsupervised training provided better input-output match, measured in terms of the output mean square error (MSE) produced by each autoencoder.

Also, as expected, the criteria ME and MQMI gave results with errors of the same order of magnitude. Because a slight advantage was observed in the adoption of the latter, in the following we will refer only to autoencoders trained with MQMI.

TABLE 1 – ACCURACY IN THE TRAINING

Autoencoder	Model	MSE
Breaker 2 open	Single step	0.0046
	2-step ME	0.0019
	2-step MQMI	0.0014
Breaker 2 closed	Single step	0.0053
	2-step ME	0.0015
	2-step MQMI	0.0007

TABLE 2 – ACCURACY IN BREAKER 2 STATUS DIAGNOSIS

	TP	FP	FN	TN
Single step	4928	104	5	4963
2-step MQMI	4915	0	14	5067

B. *Testing the quality of the breaker status identification*

In this experiment, the two autoencoders (for *open* and *closed* data clusters) trained in a single step mode were put together in a competitive scheme, in a way similar to the method in [16]. Alternatively, diagnoses were obtained with two competing autoencoders, adopting the unsupervised training of the first half strategy proposed in this paper.

Table 2 summarizes de results. The base case was defined as "breaker 2 closed" and the number of true positives (TP) and negatives (TN) and false positives (FP) and negatives (FN) were counted. For instance, a TP case is counted when the diagnosing system predicts that the breaker is closed and it is in fact closed. Only the best autoassociative networks identified in the previous experiment were used.

Our approach led to only 14 wrong diagnoses in 10,000 test cases, against the already reasonable value of 109 achieved with single step trained autoencoders – a new remarkable accuracy of 99.86%! As there is no reason to state that FP and FN cases cause distinct costs to the operation of the power network, the diagnosis system based on the new 2-step training with unsupervised MQMI criterion must be considered indeed as a better system.

An analysis of the 14 cases of wrong diagnosis explains why 100% accuracy may not be reachable. In each case, the active and reactive power flows in line 3-9 were showing values very close to zero while the breaker was closed. This seems to be a condition triggering a FN diagnosis (indicating an open breaker when it is closed). The distribution of these 14 cases in 10000 random samples is shown in Fig. 6.

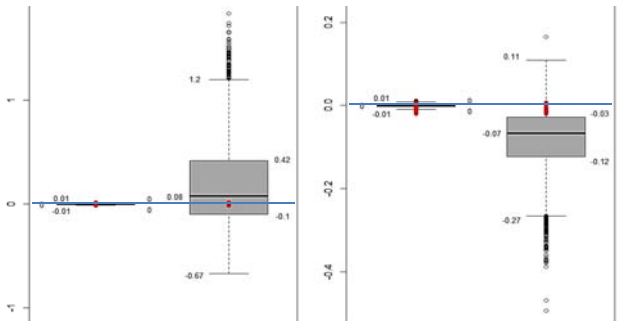

Fig. 6. 10,000 samples: active and reactive power flows in line 3-9 (left and right graphs. In each graph: distribution with open (left) and closed breaker (right). The dark (red) dots around zero correspond to the 14 False Negatives, showing simultaneous almost zero flow values in the closed breaker cases.

VIII. CONCLUSIONS

This paper introduces unsupervised learning in the training of autoassociative neural networks, and demonstrates that the adoption of such strategy leads to better results in the identification of breaker status when only local electric data are available.

This identification is very much needed in control centers because breaker status signals are many times missing (do not arrive at the SCADA) when needed – and the definition of breaker status is a necessary step to build up an image of the network topology that allows analysis algorithms to run – such as power flow or state estimation.

A previous paper had demonstrated that the concept of competitive autoassociative neural networks could be used with good accuracy in the identification of breaker status, taking in account only electric data from the network. However, given that 100% accuracy was not achieved in the diagnosis, it was important to improve it as much as possible. Boosting confidence in the approach allows it to be seriously considered to become integrated in the industrial environment of an EMS or DMS.

The idea of resorting to unsupervised learning derives from of Information Theoretic Learning and the interpretation of the work of an autoencoder as dealing with information flow. The ability to maximize the information flow (minimize information loss) is present in the criterion adopted: the maximization of the mutual information between input and middle layer, equivalent to the maximization of the entropy at the middle (hidden) layer. The remarkable accuracy achieved demonstrates the correctness of the new approach.

REFERENCES

[1] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", *Science*, Vol. 313, no. 5786, pp. 504 – 507, July 2006

[2] Baldi, P. and Hornik, K., "Neural networks and principal component analysis: Learning from examples without local minima", Neural Networks, 2(1):53–58, 1989

[3] N. Japkowicz, S.J. Hanson, M.A. Gluck, "Nonlinear Autoassociation is not Equivalent to PCA", Neural Computation, Vol. 12 , Issue 3, pp. 531 - 545, March 2000

[4] O. Alsaç, N. Vempati, B. Stott, and A. Monticelli, "Generalized state estimation", *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 1069–1075, Aug. 1998.

[5] J. Pereira, V. Miranda and J. Tomé Saraiva, "Fuzzy Control of State Estimation Robustness", *Proceedings of the 14th PSCC*, Seville, Spain, June 2002

[6] K. Clements and A. Simões Costa, "Topology error identification using normalized Lagrange multipliers", IEEE Transactions on Power Systems, vol. 13, no. 2, pp. 347-353, 1998.

[7] E. M. Lourenço, A. Simões Costa, and K. A. Clements, "Bayesian-Based Hypothesis Testing for Topology Error Identification in Generalized State Estimation," IEEE Transactions on Power Systems, vol. 19, no. 2, pp. 1206-1215, 2004.

[8] E. M. Lourenço, A. Simões Costa, K. A. Clements, and R. A. Cernev, "A topology error identification method directly based on collinearity tests", IEEE Transactions on Power Systems, vol. 21, no. 4, pp. 1920-1929, 2006.

[9] E. Caro, A. J. Conejo, and A. Abur, "Breaker status identification," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 694–702, 2010.

[10] A. A. da Silva, V. Quintana, and G. K. H. Pang, "A pattern analysis approach for topology determination, bad data correction and missing measurement estimation in power systems", *Proceedings of the Twenty-Second Annual North American*, pp. 363 - 372, 1990.

[11] A. P. Alves da Silva, V. H. Quintana, and G. K. H. Pang, "Neural networks for topology determination of power systems," *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, Seattle, pp. 297-301, 1991.

[12] A. Alves da Silva, V. Quintana, and G. Pang, "Solving data acquisition and processing problems in power systems using a pattern analysis approach", *IEE Proceedings-C*, vol. 138, no. 4, pp. 365–376, 1991.

[13] D. Vinod Kumar, S. Srivastava, S. Shah, and S. Mathur, "Topology processing and static state estimation using artificial neural networks," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 143, no. 1, pp. 99–105, 1996

[14] J. Souza, A. Leite da Silva and A. P. Alves da Silva, "Online topology determination and bad data suppression in power system operation using artificial neural networks", *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 796-803, 1998.

[15] V. Miranda, J. Krstulovic, H. Keko, C. Moreira, and J. Pereira, "Reconstructing Missing Data in State Estimation With Autoencoders", *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 604-611, 2012

[16] J. Krstulovic, V. Miranda, A. J. A. Simões Costa and J. Pereira, "Towards an auto-associative topology state estimator", *IEEE Trans. on Power Systems*, DOI: 10.1109/TPWRS.2012.2236656, 2013

[17] A. Renyi, "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Renyi*, vol 2, pp. 526-552, Akademia Kiado, Budapest, 1976.

[18] E. Parzen, "On the estimation of a probability density function and the mode", *Annals Math Statistics*, vol 33, 1962

[19] D. Erdogmus and J. C. Principe, "Generalized Information Potential Criterion for Adaptive System Training", *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1035-1044, Sep 2002.

[20] J. C. Principe, Information Theoretic Learning - Renyi's Entropy and Kernel Perspectives, New York: Springer, 2010

[21] V. Miranda, S. L. Lima, "Diagnosing faults in power transformers with autoassociative neural networks and mean shift", IEEE Tran.s on Power Delivery, vol.27, no.3, pp.1350-1357, July 2012.

[22] IEEE RTS Task Force of APM Subcommittee, "IEEE Reliability Test System", *IEEE Transactions on PAS*, Vol-98, No. 6, pp 2047-2054, Nov/Dec. 1979.

Vladimiro Miranda (M'90, SM'04, F'06) received his Ph.D. degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1982 in Electrical Engineering. In 1981 he joined FEUP and currently holds the position of Full Professor. He is also a researcher at INESC since 1985 and is currently Director at INESC Porto, the leading institution of INESC TEC – INESC Technology and Science, an advanced research network in Portugal. He is also President of INESC P&D Brasil. He has authored many papers and been responsible for many projects in areas related with the application of Computational Intelligence to Power Systems.

Jakov Krstulovic (StM'10) received his graduation degree from the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia in 2008. He is currently a PhD student at the Faculty of Engineering of the University of Porto (FEUP), Portugal. From October 2010 he is a Junior Researcher in Power Systems Unit of INESC TEC, Portugal. He also works as a teaching assistant at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Croatia. His primary research interests are power system state estimation, computational intelligence and large scale integration of the wind power.

Joana Hora graduated from FEUP, in 2009 in Industrial Engineering and is currently a researcher at INESC TEC, Portugal, in the Power Systems Unit.

Vera Palma graduated from FCUP, in 2007 in Applied Mathematics and is currently a researcher at INESC TEC, Portugal, in the Power Systems Unit.

Jose C. Principe (M'83–SM'90–F'00) is currently a Distinguished Professor of Electrical and Biomedical Engineering at the University of Florida, Gainesville. He is BellSouth Professor and Founder and Director of the University of Florida Computational Neuro-Engineering Laboratory (CNEL). He is involved in biomedical signal processing, in particular, the electroencephalogram (EEG) and the modeling and applications of adaptive systems. Dr. Príncipe is the past Editor-in-Chief of the IEEE Transactions on Biomedical Engineering, past President of the International Neural Network Society, former Secretary of the Technical Committee on Neural Networks of the IEEE Signal Processing Society, and a former member of the Scientific Board of the Food and Drug Administration. Dr. Principe was awarded the 2011 IEEE CIS Neural Networks Pioneer Award.