Extracting topology information from electric measurements: a model comparison

Cátia S. P. Silva, Jakov Krstulovic, *Student Member, IEEE*, Joana H. Martins, Vera Palma, Vladimiro Miranda, *Fellow, IEEE* and José C. Príncipe, *Fellow, IEEE*

Abstract **– This paper confirms that the network topology information lies hidden in the manifold supporting the solutions of the power flow equations and that suitable methods may make it explicit without direct information on the breaker status. A set of methods is applied to the identification of the unknown status of a switch, by dealing only with local electric information, and their performance efficiencies are compared. One of the methods uses optimal subspace projections using metric learning with an entropy functional that preserves classification accuracy. The results have direct influence on the way one may build local topology estimators either in distributed or centralized state estimation.**

Index Terms **– State Estimation, Power System Topology, Entropy, Pattern recognition, Classification., Autoencoders**

I. Introduction

THE growing acceptance of the smart grid concept leads inevitably to the need to develop new control designs for the electric power system, based on distributed architectures. This development may imply heavy investment in sensor and measuring devices and any means that may replace in part such investment should be considered.

The definition of the topology of an electric power network is an essential step previous to any kind of power system analysis. Its importance is highlighted in State Estimation (SE) procedures resident in Control Centers running Energy or Distribution Management Systems. The emergence of distributed control schemes further increases the interest in processes that may guess the status of switches or breakers without having access to direct information.

The information on breaker status has a binary form and is received at the SCADA system before being used in the SE process. However, it may be corrupted, due to sensor malfunction, which may lead to wrong topology identification, or it may even be absent. When this happens, the information on the power flow through the breaker is used to define if it is open or closed. This heuristic rule, based on deciding that the breaker is closed if some power flow throughput is detected, is however useless if this information is not available.

The conjecture that the topology information lies hidden in the electric data was addressed in previous publications [1][2]. The results achieved by adopting a competitive autoencoder scheme have been revealing: the information on breaker status is indeed spread in the electric variable values and it may be retrieved by suitable means. The efficiency of the method is therefore of paramount importance.

This paper addresses the problem of identifying the status of a single breaker inserted in a network scheme, assuming that training data about the conditions of interest have been previously obtained, by means of treating the information hidden in local electric variable values. The value of the information reported in this paper is threefold:

1. The theoretical value of, for the first time, systematically addressing the problem of breaker status identification within a framework of classification of topology-induced system states.
2. The theoretical value of presenting a novel approach to identify breaker status based on a sound mathematical formulation.
3. The practical interest of providing to the industry a comprehensive assessment of the merits of a set of techniques and producing practical solutions to identify the status of breakers in the absence of direct sensor signals.

To report the research results achieved, we have specifically selected the case of a single breaker inserted in a grid. The apparent simplicity of the case is however challenged by the complexity of the classification problem. The conclusions extracted from such example will serve as the foundation for more complex cases in the future.

The paper addresses a series of possible techniques to identify the breaker status from an analysis of local electric data, including the heuristic rule referred to above. A special focus is given to the novel technique to be described in Section IV, which attempts to recognize patterns in a reduced space by applying information entropy concepts.

The test cases were organized by selecting small sub-networks from the IEEE RTS 24-bus network [3]. All the tests are made in the absence of gross errors in the data. Noise is also absent, except the one added by the circuit modeling (branches with a pi-model) and the power flow routine used to solve the power flow equations (a classical Newton-Raphson method).

The most striking result shown is that, even in the absence of direct information on branch flows (thus rendering unusable the heuristic rule) one is able to retrieve the breaker status

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.

C. Silva (catiapsilva@gmail.com), J. Krstulovic (jopara@inescporto.pt), J. Hora (joana.hora@gmail.com), V. Palma (vera.p.ferreira@inescporto.pt) and V. Miranda (vmiranda@inescporto.pt) 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. Principe (principe@cnel.ufl.edu) and C. Silva are with the Computational NeuroEngineering Laboratory, University of Florida, Gainesville, FL 32611 USA.

information with high accuracy.

II. Brief review of past work in breaker status identification

The determination of a single switch or breaker status has been traditionally addressed within the framework of State Estimation. However, this section will not be reviewing general topology identification models with SE, but only identifying some models where breaker status identification is an almost independent feature. With this focus, several models adopting neural networks have been proposed [4][5][6][7][8]. Neural networks, with their capacity of being able to be trained to recognize patterns in data, are one of the natural candidates for the task of recognizing a breaker status within electric data.

In another line of reasoning, one must refer to [9], a true topology estimator model which attempts to find out the switching branch statuses from analog electrical measurements. It also relaxes the binary variables representing switches to the [0, 1] interval, but includes them as linearized constraints to the SE problem. The problem is then solved for these variables only as a TSE problem in the Weighted Least Squares sense.

However, from all literature reviewed, it must be said that the publications that display a sharper focus on a local estimator model for breaker status estimation are [1][2], where a competitive arrangement of local auto-associative neural networks or autoencoders is presented as a tool with an extremely high rate of success.

III. Models being compared

This section briefly describes the models that were put in competition to identify the status of a breaker in the absence of a breaker status signal. One assumes that active and reactive power measurements for line flows and bus injections/loads are available, to a higher or lesser extent – for instance, from the data collected at the SCADA in a control center. No voltage information is used in the work reported.

A. Linear Discriminant Analysis - LDA

The linear discriminant analysis (LDA) is a statistic that finds a linear combination of features which (linearly) separates the classes of objects or events. LDA is well-known in the machine learning field and it is a parametric technique, that is, it explicitly uses a Gaussian data model to find the differences between the classes of the data.

It assumes that the cases of a class k are generated according to some probability distribution $\pi_k = p(Y = k)$ and its predictor variables are generated by a class-specific multivariate normal distribution (each class is a spherical cluster)

$$X|Y = k \sim N(\mu_k, \Sigma_k)$$

i.e.

$$p_k(x) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}\langle x-\mu_k, \Sigma_k{}^{-1}(x-\mu_k)\rangle}$$

The LDA predicts the outcome as follows:

$$\hat{Y}|X = x \coloneqq argmax_k\ \pi_k p_k(x) = argmax_k\ \delta_k(x) \quad (1)$$

with the discriminant functions:

$$\delta_k(x) = -\frac{1}{2}log|\Sigma_k| - \frac{1}{2}\langle x-\mu_k, \Sigma_k^{-1}(x-\mu_k)\rangle + log\pi_k \quad (2)$$

Here, $\langle x-\mu_k, \Sigma_k^{-1}(x-\mu_k)\rangle$ is called the Mahalanobis distance of x and μ_k. Thus, LDA can be described as prototype method, where each class k is represented by a prototype μ_k and cases are assigned to the class with the nearest prototype.

The variance matrix estimated by LDA can be used to linearly transform the data such that the Mahalanobis distance $\langle x, \Sigma^{-1}y\rangle$ becomes the standard Euclidean distance in the transformed coordinates $\langle x', y'\rangle = x^T y$. This is accomplished by decomposing $\hat{\Sigma}$ as $\hat{\Sigma} = UDU^T$ with orthonormal matrix U (i.e., $U^T = U^{-1}$) and a diagonal matrix D and setting $x' = D^{-1/2}U^T x$.

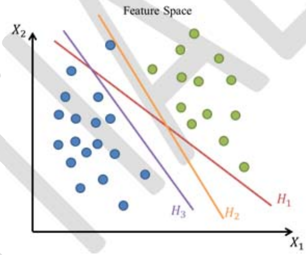

Fig. 1: For a 2-class problem, this plot shows three possible boundaries by linear discriminant analysis. The data is merely representative.

B. Quadratic Discriminant Analysis – QDA

In the most general case, decision boundaries are quadratic due to the quadratic occurrence of x in the Mahalanobis distance. This is called quadratic discriminant analysis (QDA). The difference between LDA and QDA is that now the covariance matrix can be different for each class (the classes can have a non-spherical cluster), as such, we will estimate the matrix Σ_k separately for each class k, $k = 1,2, \dots, C$.

The classification rule is similar to (1) and we also seek for the class k which maximizes the quadratic discriminant function (2).

Because it allows for more flexibility for the covariance matrix, the QDA tends to fit the data better than the LDA, but then it has more parameters to estimate.

C. Regression tree (RT)

This subsection presents a particular kind of nonlinear predictive model: regression trees (RT).

As a general term, linear regression is a global model, where there is a single predictive formula holding over the entire data space. When the data has lots of features which interact in complicated, nonlinear ways, assembling a single global model can be very difficult, and hopelessly confusing when you do succeed. An alternative approach to nonlinear regression is to sub-divide, or partition, the space into smaller regions, where the interactions are more manageable. We then partition the sub-divisions again - this is called recursive partitioning - until obtaining chunks of the space which are so tame that we can fit simple models to them. The global model thus has two parts: one is just the recursive partition; the other is a simple model for each cell of the partition.

RT use a tree to represent the recursive partition. Each of the *terminal nodes*, or *leaves*, of the tree represents a cell of the partition, and has attached to it a simple model which applies in that cell only. A point x belongs to a leaf if x falls in the corresponding cell of the partition. To figure out which cell we are in, we start at the *root node* of the tree, and ask a sequence of questions about the features. The interior nodes are labeled with questions, and the edges or branches between them labeled by the answers. Which question we ask next depends on the answers given to previous questions. An example of a regression tree is illustrated in Fig. 3.

All the data processing for the LDA, QDA and RT approaches was performed off-line using a commercial software package named Statistics from the MATLAB R2012a, The MathWorks Inc., Natick, MA, 2013 [10].

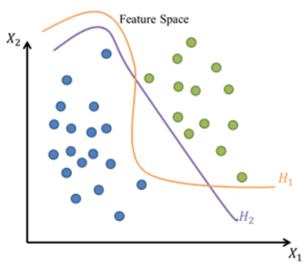

Fig. 2: For a 2-classes problem, this plot shows two possible boundaries by quadratic discriminant analysis. The data is merely representative.

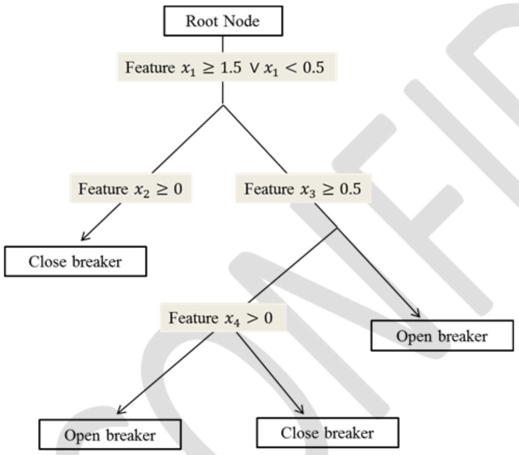

Fig. 3: Regression tree example for classification of a local breaker within a network topology. The variables are merely representative.

D. Autoencoders (AE)

An alternative to the previous methods was presented in [2], based on a competitive arrangement of autoassociative neural networks (or autoencoders).

IV. Metric Learning Using Conditional Entropy

The previous methods are all based on probability density functions (pdf) estimation. The parametric methods (LDA, QDA and RT) assumed a data distribution (Gaussian with equal cluster covariance, Gaussian with different cluster covariance and uniform) in which the model is based and developed. The non-parametric method (autoencoder) does not impose a model to the underlying distribution of the data, but resorts to pdf Parzen density estimation during the training phase. Until now, these kernels methods have been the most utilized descriptors of a pdf of the data in a non-parametric setting.

Positive definite kernels have been employed in machine learning as a representational tool allowing algorithms that are based on inner products to be expressed in a rather generic way [11] – known as kernel methods. Let $\mathcal{X}$ be a nonempty set. A function $\varkappa\colon \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is called a positive definite kernel if for any finite set $\{x_i\}_{i=1}^{N} \subseteq \mathcal{X}$ and any set of coefficients $\{\alpha_i\}_{i=1}^{N} \subset \mathbb{R}$, it follows that $\sum_{i,j} \alpha_i \alpha_j \varkappa(x_i, x_j) \geq 0$, if at least one i, $\alpha_i \neq 0$. In this case, there exist an implicit mapping $\phi\colon \mathcal{X} \mapsto \mathcal{H}$ that maps any element $x \in \mathcal{X}$ to an element $\phi(x)$ in a Hilbert space $\mathcal{H}$, such that $\varkappa(x, y) = \langle \phi(x), \phi(y) \rangle$. The above map provides an implicit representation of the objects of interest that belong to the set $\mathcal{X}$. This is possible as long as a kernel function is available. Therefore, (see Appendix) the Gram matrix obtained from evaluating a positive definite kernel on samples can be used to define a quantity based on the data with properties similar to those of an entropy measure without assuming that the probability density is being estimated.

We can thus apply the matrix framework to the problem of supervised metric learning, namely the breaker status identification. This problem can be formulated as follows. Given a set of points $\{(x_i, l_i)\}_{i=1}^{N}$, we seek a positive semidefinite matrix AA^T, that parametrizes a Mahalanobis distance between samples $x, x' \in \mathbb{R}^d$ as $d(x, x') = (x - x')^T AA^T (x - x')$. The goal is to find a parametrization matrix A such that the conditional entropy of the labels l_i given the projected samples $y_i = A^T x_i$ with $y_i \in \mathbb{R}^p$ and $p \ll d$, is minimized. This can be posed as the following optimization problem:

$$\begin{aligned} &\underset{A \in \mathbb{R}^{dxp}}{\text{minimize}} \; S_\alpha(L|Y) \\ &\text{subject to:} \\ &A^T x_i = y_i, for\ i = 1, \dots, N; \\ &\qquad tr(AA^T) = p, \end{aligned} \tag{3}$$

where the trace constraint prevents the solution from growing unbounded. According to [11], we can translate this problem to the matrix-based framework in the following way. Let **K** be the matrix representing the projected samples

$$K_{ij} = \frac{1}{N} exp\left(-\frac{(x_i - x_j)^T AA^T (x_i - x_j)}{2\sigma^2}\right) \tag{4}$$

with σ as free parameter and **L** be the matrix of class co-occurrences where $L_{ij} = \frac{1}{N}\ if\ l_i = l_j$ and zero otherwise. The conditional entropy can be computed as $S_\alpha(L|Y) = S_\alpha(NK \circ L) - S_\alpha(K)$, and its gradient at A is given by:

$$X^T\big(P - diag(P1)\big)XA \tag{5}$$

where

$$P = \big(NL \circ \nabla S_\alpha(NK \circ L) - \nabla S_\alpha(K)\big) \circ K \tag{6}$$

and

$$\nabla S_\alpha(A) = \frac{\alpha}{(1-\alpha) tr(A^\alpha)} U \Lambda^{\alpha - 1} U^*, with\ A = U \Lambda U^* \tag{7}$$

Note that instead of computing the full set of eigenvectors and eigenvalues of A, we can approximate the gradient of S_α by using a few leading eigenvalues.

Finally, we can use an algorithm of the gradient descent type to search for A iteratively. After settling on a projection matrix A, the original space $\mathcal{X}$ can be projected onto the space ruled by A, say $\mathcal{Y}$. One may then look for patterns in the transformed space $\mathcal{Y}$, applying a classifier, without ever estimating the data pdfs in the high dimensional space. One can expect to obtain similar results with minimum degradation but with a somewhat easier calculation process due to the space dimension reduction achieved.

V. Results – Comparative Tests

A. Database

The numerical results from the case study are based on the 24-bus system [3] (Fig. 4), including:

a) Insertion of breakers in 10 locations of the network;
b) Design of a cumulative load curve with data from [3], based on which load levels are sampled and a large variety of scenarios from valley to peak of the load curve are constructed;
c) Random sampling of feasible power generation values;
d) Breaker status randomly defined for each generation/load scenario;

For each scenario, the power flow equations were solved. Then, simulation real world conditions, Gaussian noise was further added to the power flow solutions, with standard deviation $\sigma = 0.01$ p.u, (1 p.u. corresponds to 100 MVA).

Thousands of scenarios were thus collected and analyzed. Then, for each of the ten breakers identified in Fig.4, a set of adjacent power measurements was chosen as defining the *local measurement set* for each breaker (only active/reactive power injection and flow measurements were used).

B. Identification of a Single Breaker Status – Local Measurements including breaker power throughput

Including the breaker active and reactive power throughput in the local measurement set allows one to verify the efficiency of the classification methods against the traditional heuristic rule: "if the flow is zero, the breaker is open, otherwise it is closed". We have examined the values of the power flows in each line with an open breaker, and determined an interval of small amplitude of the flows associated to an open device (flows are not exactly zero, even if the breaker is open, due to reactive currents and derived losses in the line, lack of absolute precision in numerical calculations, and also due to Gaussian noise added in order to mimic the real measurements).

Table I displays the results of status identification with several techniques for each of the 10 breakers, on a sample of 10,000 scenarios considering the totality of the local measurement set of a particular breaker. Krstulovic et al. [2] showed interesting results using a framework of competitive autoencoders. The application of the technique to the same data as with the other techniques is also in Table I. The last column refers to a classifier (QDA) applied after a space dimension reduction via the technique descried in Section IV.

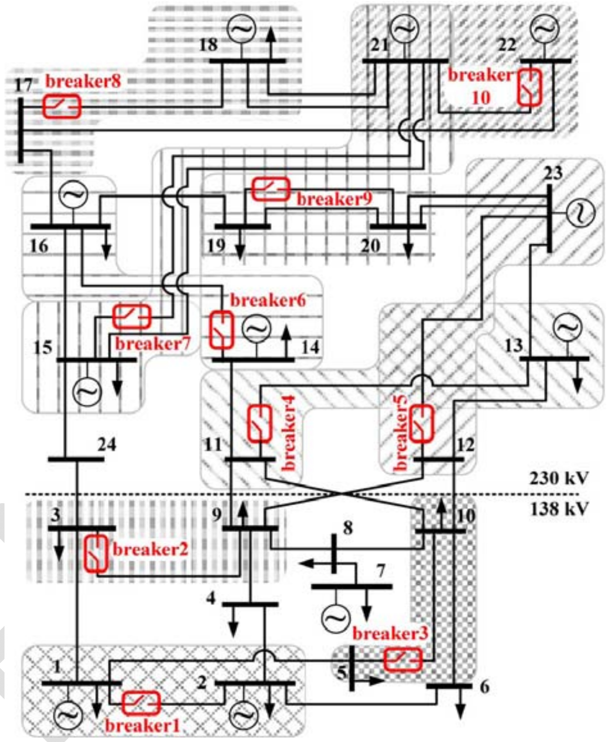

Fig. 4. IEEE RTS 24 with indication of the branches where 10 switches were introduced and the local areas of measurement collection for each breaker.

TABLE I: Comparison of Local Status Identification Methods with Local Measurements in 10,000 Scenarios

Breaker	Method (efficiency = no. correct/10,000 in %)							Projection Matrix A
	LDA Eucl.	LDA Mah.	QDA	RT	AE	Heur.	Entropy foll. by QDA	Reduced/ Original space
1	100	100	100	99.90	100	99.96	100	12/14
2	86.1	92.55	99.98	99.91	99.27	99.99	99.94	16/18
3	65.69	79.10	99.92	99.75	99.13	99.90	99.82	14/16
4	99.99	99.99	100	100	100	99.99	100	14/16
5	99.98	100	100	100	100	100	100	12/14
6	99.69	99.98	100	100	100	100	99.50	12/14
7	99.47	100	100	100	100	100	100	14/16
8	96.75	98.85	100	99.92	100	99.97	100	12/14
9	99.62	100	100	99.96	100	99.98	100	14/14
10	99.28	99.95	100	99.92	100	99.96	99.99	16/16
Avg.	**94.66**	**97.04**	**99.99**	**99.94**	**99.84**	**99.98**	**99.93**	**11% reduction**

The QDA classifier provided the best results. The heuristic rule failed in more cases because in some instances the flow through the breaker is very small due to almost equal nodal voltages and this causes confusion in the diagnosis.

These results are convincing in the sense that they demonstrate that information gathered from a vector of measurements can compete with the heuristic rule (based on information on flows through a breaker) in deciding if a breaker is open or closed. The QDA technique allowed even a marginally better result, likely deriving from dealing with more information that just breaker flows.

The last technique shows that some space dimension reduction is possible with negligible loss of precision. However, the dimension reduction permitted to maintain quality was only of 11% on average.

C. Excluding the direct Active and Reactive Power Flows

The value of the pattern recognition approach is enhanced however in the case of an absence of information on the power flows through the breaker – when no heuristic rule may be used. If the active and reactive power flows were eliminated from the data set (reproducing a case of missing signals in the SCADA), would it still be possible to establish a correct diagnosis?

Table II presents the results of the status identification for each of the same 10 breakers and on the previous sample of 10,000 distinct scenarios with breaker flows removed.

TABLE II: COMPARISON OF LOCAL STATUS IDENTIFICATION METHODS WITH LOCAL MEASUREMENTS EXCLUDING THE POWER FLOW DIRECTLY CONNECTED TO THE BREAKER

Breaker	Method (efficiency = no. correct/10,000 in %)						Projection Matrix
	LDA Eucl.	LDA Mah.	QDA	RT	AE	Entropy foll. by QDA	Reduced/ Original space
1	100	100	100	98.72	100	100	12/12
2	83.16	90.83	99.60	83.45	97.40	99.57	16/16
3	58.68	77.51	99.31	96.28	98.79	99.15	12/14
4	99.99	99.98	100	99.61	100	99.73	12/14
5	99.98	100	100	99.72	100	100	12/12
6	82.16	80.51	85.56	76.52	100	83.27	10/12
7	99.44	100	100	97.02	100	100	14/14
8	86.33	93.79	99.99	94.60	100	100	12/12
9	96.09	100	100	99.27	100	100	12/12
10	95.58	99.76	99.94	97.48	100	99.88	14/14
Avg.	**90.14**	**94.24**	**98.44**	**94.27**	**99.62**	**98.16**	**5% reduction**

The method using competitive AE proposed in [2] is now the best performing, closely followed by the QDA classifier, either applied in the original measurement space or in a reduced space after the method described in Section IV. It is interesting to notice that Breaker 6 emerges as a difficult case for the classical methods but not for the AE method.

The conclusion is unequivocal: it is indeed possible to recognize patterns in the electric power flows, associated with topologic states. And, given enough measurements, the accuracy in classification is remarkable: in 8 out of 10 breakers studied, no erroneous classification could be found in 10,000 random scenarios for each case in the case of the AE!

D. Remote Measurements

The impact of the input measurements choice is crucial to the efficiency of the method. Without discussing an optimization of the choice of the measurement set, it is however not at all proved that the topology information is merely local. On the contrary, the work in [2] already hints that at least partial information on breaker statuses is still present in remote data.

To test this hypothesis, we selected a set of measurements non-adjacent to each breaker similar to the ones reported in [2]. Fig. 5 exemplifies the remote power flow measurements chosen as inputs in the case of the breaker 6: 14 remote measurements (active and reactive power flows) were chosen. A similar procedure was adopted to select the measurements for the other 9 switches.

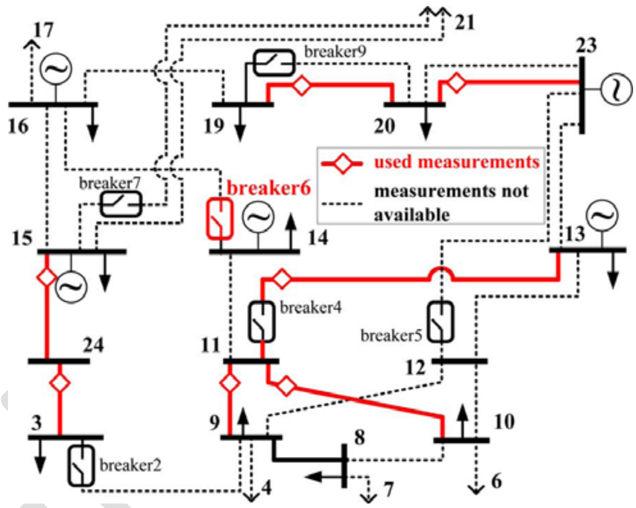

Fig. 5. Partial representation of the IEEE RTS 24 system, for the remote identification of breaker 6 status. Flow measurements are available only for full lines and not for dashed lines.

TABLE III: COMPARISON OF DIFFERENT METHODS FOR STATUS IDENTIFICATION WITH REMOTE DATA ONLY

Breaker	Method (efficiency = no. correct/10,000 in %)						Projection Matrix
	LDA Eucl.	LDA Mah.	QDA	RT	AE	Entropy foll. by QDA	Reduced/ Original space
1	76.22	78.48	80.12	74.26	63.83	76.64	10/12
2	68.97	68.42	75.08	65.1	65.40	72.67	16/18
3	58.47	67.44	75.49	60.18	60.81	71.17	18/20
4	98.44	98.06	99.21	94.23	92.51	99.22	16/18
5	65.05	63.58	65.60	45.92	57.33	62.72	14/14
6	99.23	99.06	99.24	98.64	99.18	99.20	14/14
7	60.79	59.23	58.91	44.28	55.83	58.95	12/12
8	80.48	81.97	92.57	77.33	75.27	91.99	10/10
9	53.51	54.67	54.92	39.72	53.02	53.60	10/12
10	79.33	76.09	86.33	70.98	89.57	86.05	10/10
Avg.	**74.05**	**74.70**	**78.75**	**67.06**	**71.28**	**77.22**	**7% reduction**

The results in Table III show that the QDA classifier, either applied in the original measurement space or in the reduced space (using the Gram matrix model referred to in Section IV, with 7% reduction in space dimension) performs better than the other classifiers. One could not achieve a convenient training of the AE that would match the classifying precision of QDA or LDA.

The number of misclassified cases is bigger than in the previous studies, which was expected – it is still a good number, quite remarkable in fact. This confirms the hypothesis that there is still significant information on the breaker status in a set of non-adjacent measurements.

The RT method is the worst performing model, now.

Much can be learned from the comparative analysis of Tables II and III. Take, for instance, Breakers 5 and 6 – the classification efficiency in B6 is always extremely high, while the efficiency for breaker B5 suffers from severe degrading when the local measurement set is ignored. This is a clear indication that simple engineering judgment is not enough to select the most adequate set of measurements to associate with topology states.

E. Metric learning in remote measurements

This problem without a direct branch measurement and with remote measurements is the most challenging one and therefore we made a deeper investigation into the possibility of classification in a reduced space, with the projections discovered with the Gram matrix/Entropy method described in Section **Error! Reference source not found.** With the determination of a projection matrix A, we found the best projection space of dimension. Into this we applied three different classifiers in order to identify the class labels for the 10,000 scenarios for each one of the 10 breakers: QDA, LDA Euclidian-based and LDA Mahalanobis-based.

According to [11], the higher entropy order α emphasizes the parts of the space with higher data densities. Since we do not have any prior knowledge or assumption on the breakers statuses, we choose α close to 1. As such, the following experiments used an entropy's order $\alpha = 1.01$.

Fig. 6 shows the performance curve acquired for the different classifiers in the metric learning projection, discovered by applying the model in Eq. (3), by changing the free-parameter in (4) within the set

$$\sigma = \{0.001, 0.01, 0.05, 0.1\ ,1\ , \sqrt{2}, 2{,}5\}$$

which defines the scale of the similarity in the RKHS.

We can see that we have a quality threshold for the σ parameter: values greater than 0.01 affect negatively the classifier's performance. The smaller the parameter σ is (down to a certain value), the better the classification efficiency becomes. We observed that the algorithm's accuracy converges for $\sigma \leq 10^{-3}$ until computational/numerical problems occur (for values near $\sigma = 10^{-200}$).

We have repeated this exercise for LDA-Euclidian and LDA-Mahalanobis with the same general result. Table IV presents the results acquired for the same 10,000 test scenarios using a training set of 500 random scenarios and adopting a small enough value for σ (σ = 0.001).

This study confirms that linear discriminants are not the best choice for a classifier in the problem of recognizing topology patterns in electric power data. Breaker 6 is still the easiest case – in fact, so easy that a LDA-Euclidian based model could slightly outperform the QDA model.

This argues in favor of the statement that each breaker location is a challenge on its own. And the spreading of information throughout the electric power data is uneven and strongly depends on the structure of the network. The challenge is therefore open to discover the optimal sets of measurements containing information of a breaker status.

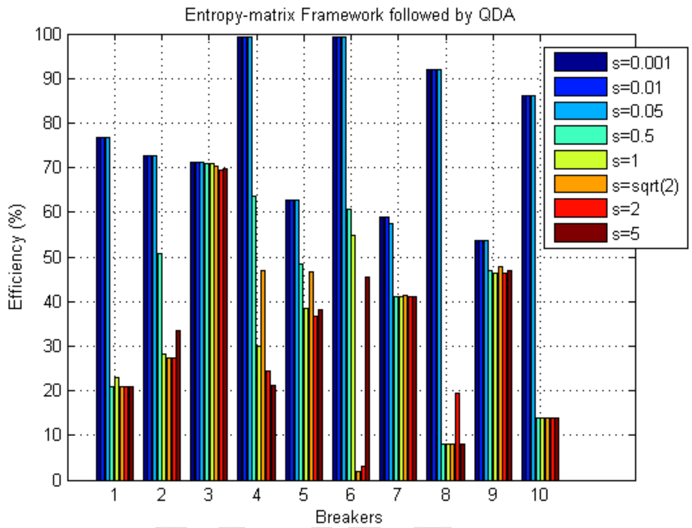

Fig. 6. Classification efficiency, using a QDA classifier, over the test scenarios projected with the learned metric governed by matrix A. Obviously, the highest efficiencies are obtained for σ < 0.01.

TABLE IV: STATE ESTIMATION RESULTS USING THE ENTROPY-MATRIX FRAMEWORK OF ORDER $\alpha = 1.01$ FOR $\sigma = 0.001$ FOLLOWED BY THREE CLASSIFIERS.

Breaker	Entropy-matrix Framework followed by		
	QDA	LDA Euclidian-based	LDA Mahalanobis-based
1	76.64%	74.74%	76.47%
2	72.67%	67.13%	69.50%
3	71.17%	56.15%	63.70%
4	99.22%	98.28%	98.30%
5	62.72%	64.74%	59.13%
6	99.20%	99.21%	99.08%
7	58.95%	60.37%	59.08%
8	91.99%	80.10%	83.94%
9	53.90%	53.25%	53.59%
10	86.05%	79.23%	77.67%
Average	**77.22%**	**73.32%**	**74.05%**

VI. CONCLUSIONS

The notion that the information on grid topology is embedded in the electrical data is intuitive. However, until now it was not demonstrated that it could be extracted reliably. But the experiments reported in this paper prove that electric data are indeed topology patterns – because several techniques identify such patterns, with a variety of degrees of accuracy.

The paper addresses several techniques competing for breaker status estimation, to add to an earlier proposal using autoencoders. The experiments reported show that the choice of an extraction technique is important if one aims at a highly accurate determination of a breaker status. After this work, one strongly believes that the area is open to research on the most efficient model to classify topology states within sets of electric data.

In industrial software applications, one presently resorts to heuristic rules (such as "if the branch current is zero then the breaker is open"), to guess a breaker status in the case of missing status signals. But as a consequence of the confirmation that topology information lies hidden in electric data, one may now confidently foresee a time when an intelligent agent will be in charge of identifying a breaker status from the pattern of electric data.

Furthermore, the assessment of such an agent is local, in the sense that it may rely on local data. This makes the technique suitable for building up a distributed topology state estimator. Or, in the absence of local data, a remote agent may still be able to produce an estimation for a breaker status, although with a lesser degree of accuracy. So, the loss of telemetry does not mean that one gets totally blind over the state in a section of a network.

An alternative use to the technique is as a complementary estimator to signals actually received at the SCADA in a control center, either confirming the status or setting up an alarm flag for possible gross error in data. A number of interesting applications of the principle investigated in this paper may be foreseen.

It must be highlighted that all the remarkable results reported in this paper were obtained with *measurement sets contaminated by noise*. This means that the topology pattern recognition is robust to noise such as the one that may be encountered in system control centers and that its integration with traditional State Estimation techniques may be foreseen.

Furthermore, these results were obtained from measurements sets including active and reactive powers only. It is likely that the addition of voltage measurements may indeed improve the quality of the pattern recognition process.

We have also shown that the difficulty of the task increases with the distance from the breaker to the measurements, which can be expected because each breaker status appears mixed differently by the circuit until the measurement. Consequently, the classification becomes progressively and unevenly harder and strongly depends on the structure of the network. This is where nonlinear projections and information theoretic quantities that fully utilize the information in the data will become more appealing.

This opens a new challenge: to adopt approaches to discover optimal sets of measurements that may maximize the efficiency in topology pattern recognition. While the subject was not addressed in the paper, we believe that such research has now a clearer purpose and benefits from more work.

VII. APPENDIX – RENYI'S ENTROPY AXIOMS AND AXIOMS FOR GRAM MATRICES

Information theoretic learning [12] showed that Renyi's second-order entropy has practical estimation advantages with respect to Shannon's entropy. An empirical plug estimator of Renyi's second-order entropy based on the Parzen density $\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} \kappa(x_i, x)$, can be obtained as follows:

$$-\log \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} h(x_i, x_j) \tag{3}$$

where $h(x, y) = \int_{\mathcal{X}} \varkappa_\sigma(x, z)\varkappa_\sigma(y, z)dz$. Since h is a positive definite kernel, there exists a mapping ϕ to a Reproducing Kernel Hilbert Space (RKHS) such that $\mathrm{h}(x, y) = \langle \phi(x), \phi(y) \rangle$; and the argument of the logarithm in (3), called the *information potential* [11], can be interpreted in this space as the mean square value of RKHS elements:

$$\left\langle \frac{1}{N}\sum_{i=1}^{N} \phi(x_i), \frac{1}{N}\sum_{i=1}^{N} \phi(x_i) \right\rangle = \left\| \frac{1}{N}\sum_{i=1}^{N} \phi(x_i) \right\|^2 \tag{4}$$

with the limiting case given by $\|E[\phi(X)]\|^2$. Thus, we can think of this estimator as a statistic computed on the representation space provided by the positive definite kernel h.

Now, let us look at the case where $\varkappa_\sigma$ is the Gaussian kernel; if we construct the Gram matrix $\mathbf{K}$ with elements $K_{ij} = \varkappa_{2\sigma}(x_i, x_j)$, it is easy to verify that the estimator of Renyi's second-order based on (3) corresponds to:

$$\widehat{H}_2(X) = -\log\left(\frac{1}{N^2} tr(KK)\right) + C(\sigma) \tag{5}$$

where $C(\sigma)$ takes care of the normalization factor of the Parzen window. As we can see, the information potential estimator can be related to the norm of the Gram matrix $\mathbf{K}$ defined as $\|K\|^2 = tr(KK)$. This reasoning can be generalized.

Real Hermitian matrices are considered generalizations of real numbers. It is possible to define a partial ordering on this set by using positive definite matrices, which admit spectral decompositions [13]. When these matrices are normalized by their trace their eigenvalues add to one; it turns out that the Gram matrix of projected data for specific kernels obey these properties. Therefore, [11] defines operators in these RKHS that obey the same axiomatic properties of entropy as defined by Renyi.

Consider the set Δ_n^+ of positive definite matrices $A \in \mathcal{M}_n$ for which $tr(A) \leq 1$. It is clear that this set is closed under a finite convex optimization and the following proposition was proven in [11].

Proposition *Let* $A \in \Delta_n^+$ *and* $B \in \Delta_n^+$ *and also* $tr(A) = tr(B) = 1$. *The functional*

$$S_\alpha(A) = \frac{1}{1-\alpha} log_2[tr(A^\alpha)],$$

satisfies the following set of conditions:

i. $S_\alpha(PA^*P) = S_\alpha(A)$ *for any orthonormal matrix* $P \in \mathcal{M}_n$.

ii. $S_\alpha(pA)$ *is a continuous function for* $0 < p \leq 1$.

iii. $S_\alpha\left(\frac{1}{N} I\right) = log_2 N$.

iv. $S_\alpha(A \otimes B) = S_\alpha(A) + S_\alpha(B)$.

v. *If* $AB = BA = \mathbf{0}$*; then for the strictly monotonic and continuous function* $g(x) = 2^{(\alpha-1)x}$ *for* $\alpha \neq 1$ *and* $\alpha > 0$, *we have that:*

$$S_\alpha(tA + (1-t)B)$$
$$= g^{-1}\left(tg(S_\alpha(A)) + (1-t)g\left(tg(S_\alpha(B))\right)\right).$$

In conclusion we define an operator in RKHS that is estimating entropy directly from projected data, without ever estimating the pdf. This is the basis of the metric learning optimization used in the paper to find sub space projections.

REFERENCES

[1] 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, May 2012

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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

[8] 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.

[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] The MathWorks Inc. [Online at www.mathworks.com].

[11] L. G. S. Giraldo and J. C. Príncipe, “Information Theoretic Learning with Infinitely Divisible Kernels”, CoRR abs/1301.3551, January 2013.

[12] J. C. Príncipe, “Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives”, Series in Information Science and Statistics, M. Jordan, R. Nowak and B. Schölkopf, Springer, 2010.

[13] R. A. Horn and C. R. Johnson, Topics in Matrix Analysis. Cambridge University Press, 1991.

Cátia S. P. Silva received her B.Sc. degree in Mathematics from the Faculty of Sciences of the University of Porto (FCUP), Porto, Portugal in 2010. In 2012, she received the M.Sc. degree in Biomedical Engineering from the University of Engineering of the University of Porto (FEUP), Porto, Portugal. She is currently pursuing the Ph.D. degree in Electrical and Computer Engineering from the Computer Neuro-Engineering Laboratory (CNEL) at the University of Florida, Gainesville, USA. Between April 2012 and June 2013, she was a Junior Researcher in Power Systems Unit of INESC TEC, Portugal.

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 H. Martins received a M.Sc. degree in Industrial Engineering in 2009 at the Faculty of Engineering of the University of Porto (FEUP), Portugal, and a M.Sc degree in Quantitative Methods in 2012 at the Faculty of Economics of the University of Porto (FEP), Portugal. She is currently a PhD student at FEUP. She was a researcher at the department of Industrial Engineering of FEUP (April 2009 - August 2011), and at Power Systems Unit of INESC-TEC (November 2011 - June 2013), Portugal.

Vera Palma graduated from the Faculty of Sciences of the University of Porto, Portugal (FCUP), in 2007 in Applied Mathematics and is currently a researcher at INESC TEC, Portugal, in the Power Systems Unit.

Vladimiro Miranda (M’90, SM’04, F’06) received his B.Sc. and Ph.D. degrees from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1977 and 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 the current Director-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. Prof. Miranda was awarded the 2013 IEEE PES Ramakumar Family Renewable Energy Excellence Award.

José C. Príncipe (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.