Application of Machine Learning on Power System Dynamic Security Assessment

Conference Paper · December 2007

DOI: 10.1109/ISAP.2007.4441604 · Source: IEEE Xplore

Emmanouil Voumvoulakis
National Technical University of Athens

Nikos D. Hatziargyriou
National Technical University of Athens

Application of Machine Learning on Power System Dynamic Security Assessment.

E. M. Voumvoulakis, A. E. Gavoyiannis and N.D. Hatziargyriou *Senior Member*

Abstract--This paper addresses the on going work of the application of Machine Learning on Dynamic Security Assessment of Power Systems. Several techniques, which have been applied for the Dynamic Security Assessment of the Greek Power System are presented. These techniques include off-line Supervised learning (Radial Basis Function Neural Networks, Support Vector Machines, Decision Trees), off-line Unsupervised learning (Self Organizing Maps) and online Supervised learning (Probabilistic Neural Networks). Results from the application of these methods on operating point series from the Greek Mainland system and the Power System of Crete island show the accuracy and versatility of the methods.

Index Terms- Decision Trees, Dynamic Security Assessment, Machine Learning Neural Networks, Probabilistic Neural Networks, Radial Basis Function Neural Networks, Self Organizing Maps, Support Vector Machines

I. Introduction

DYNAMIC Security assessment is probably the most versatile field of application of automatic learning techniques in power systems. Moreover with the present evolutions of power systems, security assessment is becoming more and more challenging.

Numerous applications of machine learning methods to power system security have been reported. A review of some recent examples is detailed in the rest of the paper. An automatic learning framework for power system security assessment, proposed in [1], is illustrated in Fig. 1. The first step is the creation of a Knowledge Base (KB). The KB comprises a large number of Operating Points (OPs), each of which is characterized by a vector of pre-disturbance steady state variables, called attributes that can be either directly measured (power, voltages, etc.) or indirectly calculated quantities (wind penetration, spinning reserve, etc.). The attributes selected to form an OP depends on the method used. The OPs must cover all possible situations of the power system and for this random sampling techniques are utilized. Load flow is performed on each OP and a disturbance is simulated using a time-domain simulation machine. The OP is then classified as safe or unsafe using a post-disturbance variable criterion (e.g. voltage at certain buses and/or frequency of the system).

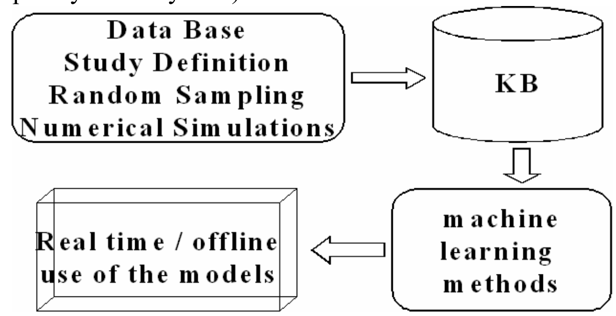

Fig. 1 Machine learning framework for security assessment

The heart of the framework is the machine learning methods that are used in a toolbox fashion and can either classify the OPs (classification) or provide a reliable estimation of the post-disturbance variables that are used as a security criterion (regression). KB is necessary for the training of these machine learning methods. The KB is split into a Learning Set (LS), used to train the machine learning algorithms, and a Test Set (TS) used for testing the developed structures.

The final step consists of using the constructed models either in real-time, for fast and effective decision-making, or in the off-line study environment, so as to gain new physical insight and to derive better system and/or operation planning strategies. This paper develops the application of some of the main machine learning techniques on the Dynamic Security Assessment of Power Systems.

II. SUPERVISED LEARNING METHODS

A. *Radial Basis Function Neural Networks (RBFNN)*

[5], proposes Radial Basis Function (RBF) Neural Networks to assess the security of the Greek mainland Power System. Radial functions are a special class of function. Their characteristic feature is that their response decreases (or increases) monotonically with distance from a central point. The center, the distance scale, and the precise shape of the radial function are parameters of the model, all fixed if it is linear. The most general formula for any RBF is

$$h(x) = \phi\left((x-c)^T R^{-1} (x-c)\right) \tag{1}$$

where φ is the function used, c is the center and R is the metric. The term $(x-c)^T R^{-1}(x-c)$ is the distance between the input vector x and the centre c in the metric defined by R. Often the metric is Euclidean. In this case $R=r^2I$ for some

E. M. Voumvoulakis is with the Department of Electrical and Computer Engineering of National Technical University of Athens Greece (e-mail: emvoumv@power.ece.ntua.gr).

N. D. Hatziargyriou is with the Department of Electrical and Computer Engineering, National Technical University of Athens, Greece (e-mail: nh@power.ece.ntua.gr).

scalar radius r and where I is unity diagonal. Its parameters are its center c and its radius r and monotonically decreases with distance from the center. Each of d components of the input vector x feeds forward to M basis functions whose outputs are linearly combined with weights w_j, $j=1...M$ into the network output f(x), as,

$$f(x)=\sum_{j=1}^{M} w_j h_j(x) \quad (2)$$

and the free variables are the weights w_j. The model f is expressed as a linear combination of a set of M fixed functions h which are often called basis functions.

The proposed method has been applied to security classification of the Greek Mainland system modified to take into account an increased wind power penetration in a critical area. The voltage level in the area is used to clarify security. For the verification of the method experiments have been performed using independent input load and wind data produced with a random procedure and the classification of each testing sample has been compared with the one of the corresponding stored testing samples. Table I presents the results of the application. Samples which belong to class A are considered to be Safe while those which belong to class B are Unsafe. The false alarms rate is defined as the ratio of the secure (OPs) that are classified as insecure to the secure OPs, while the missed alarms rate is defined as the ratio of the insecure OPs.

TABLE I.
EVALUATION OF TWO-CLASS CLASSIFICATION PERFORMANCE OF RBFNNS

	Estimation of RBF	
	Class A	Class B
Samples of class A	2176	160
Samples of class B	34	634
Success Rate	93.53%(2806/3000)	
False Alarms	6.85% (160/2336)	
Missed Alarms	5.12% (34/664)	

B. *Support Vector Machines*

The foundations of SVMs have been developed by Vapnik [6]. The method is gaining popularity due to its many attractive features, and promising performance [7,8]. The main idea of a SVC is to construct a hyperplane as the decision surface, in such a way that the margin of separation between positive and negative examples is maximized. The machine achieves this desirable property by following a principled off-line approach rooted in the statistical learning theory. More precisely, the SVM is an approximate implementation of the method of Structural Risk Minimization (SRM) [7,8].

1) *Support Vector Classifiers*

For a two-class partition a probability distribution *P(x,y)* is assumed, where for every input row (sample or operating state in our application) $x\in R^n$, $y\in\{-1,1\}$ which is the corresponding class, i.e. secure or insecure. The SVC seeks an optimal separating hyperplane between the input vectors which belong to classes. This problem can be made easier if the vectors are mapped from the input space to a higher dimensional space H which is called feature space. The mapping is, $\Phi : R^d \to H$. The hyperplane assumes the form, $\sum_{j=1}^{M} w_j\phi_j(x)$, where M is equal to the dimensionality of the feature space and $\varphi\in\Phi$ is a non-linear function. This is a linear separating hyperplane on the feature space. This can be reformulated as, $\sum_{i=1}^{N} a_i y_i K(x_i, x)$, where a_i are the Lagrange multipliers. K is called the inner product kernel defined as, $K(x,x')=\sum_{j=1}^{M}\phi_j(x)\phi_j(x')$. The optimization problem presented as a Lagrange optimization is formed as,

$$Q(a)=\sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j y_i y_j K\left(x_i^T x_j\right) \quad (3)$$

subject to the constraints,

$$\sum_{i=1}^{N} a_i y_i = 0, \quad 0 \le a_i \le C \quad for \quad i=1,...,N$$

Radial basis functions most commonly use a kernel of the form,

$$K(x,x')=\exp-\frac{(x-x')^2}{2\sigma^2} \quad (\mathbf{4})$$

Note that the number of basis functions, the center parameters, that correspond to the support vectors and the weights in the output layer are all automatically determined via the optimal hyperplane. All basis functions have the same width parameter which is specified a priori. The proposed method has been applied on the data set of section II.A. Table II shows the results of the two-class classification of 3000 new test samples by SVC.

TABLE II
EVALUATION OF TWO-CLASS CLASSIFICATION PERFORMANCE OF SVC

	Estimation of SVC	
	Class A	Class B
Samples of class A	2191	145
Samples of class B	46	618
Success Rate	93.63%(2809/3000)	
False Alarms	6.21% (145/2336)	
Missed Alarms	6.93% (46/664)	

2) *Support Vector Regressors*

The problem of approximating the set of data, $(x_1,y_1),...,(x_N,y_N)$, where $x\in R^n$, $y\in R$, by a function that has at most ε deviation from the targets (y_i) and is as flat as possible [9], is considered. Initially, a linear function of the form, f is considered, $f(x,w)=(w\cdot x)+b, \quad w\in R^n$.The task is to minimize the empirical risk, $R_{emp}=\frac{1}{L}\sum_{i=1}^{N} L_\varepsilon(y_i)$, subject to the condition $\|w\|^2 \le c_0$, where c_o is a constant. The loss function $L_\varepsilon(y)$has the form,

$$L_\varepsilon(y)=\begin{cases} 0 & for \quad |f(x)-y|<\varepsilon \\ |f(x)-y|-\varepsilon & otherwise \end{cases} \quad (5)$$

A more challenging problem is to aim at solving the nonlinear regression task. Similar to the case of nonlinear classification, this will be achieved by considering a linear regression hyperplane in the so-called feature space. Then function f becomes $w \cdot \phi(x) + b$. The optimal regression function is given by the minimum of the function,

$$n(w,\xi^*,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{6}$$

subject to the constraints,

$$y_i - w^T \cdot \phi(x_i) \le \varepsilon + \xi_i, \quad w^T \cdot \phi(x_i) - y_i \le \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0, \quad i = 1,\ldots,N \tag{7}$$

where C is a pre-specified value and a trade off between flatness of f and the amount up to which deviations larger than ε can be tolerated. ξ, ξ^* are slack variables representing upper and lower constraints on the outputs of the system. Then the problem is transformed into the dual form of the Lagrange optimization task and the solution is given by,

$$Q(\alpha,\alpha^*) = \arg\min_{a,a^*} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_i - a_i^*)(a_j - a_j^*)K(x_i,x_j) - \\ -\sum_{i=2}^{N}(a_i - a_i^*)y_i + \sum_{i=1}^{N}(a_i - a_i^*)\varepsilon \tag{8}$$

where a and a^* are the Lagrange multipliers, with constraints,

$$0 \le a_i, a_i^* \le C\sum_{i=1}^{N}(a_i - a_i^*) = 0 \tag{9}$$

where $K(x_i,x_j)$ is defined as in the SVC case, i.e. $K(x_i,x_j) = \phi^T(x_i)\cdot\phi(x_j)$. The parameters ε and C must be defined by the user. Solving equation (9) with constraints (10) determines the Lagrange multipliers, a *and* a^*, and the regression function is given by,

$$f(x,w) = w^T x = \sum_{i=1}^{N}\left(a_i - a_i'\right)K(x,x_i) \tag{10}$$

The proposed method has been applied on the data set of section II.A. Table III shows the results of the two-class classification of the same 3000 new test samples by SVR.

TABLE III
EVALUATION OF TWO-CLASS CLASSIFICATION PERFORMANCE OF SVR

	Estimation of SVR	
	Class A	Class B
Samples of class A	2175	161
Samples of class B	20	644
Success Rate	93.97%(2809/3000)	
False Alarms	6.89% (161/2336)	
Missed Alarms	3.01% (20/664)	

C. Decision Trees

The Decision Tree (DT) is a tree, structured upside down, built on the basis of a Learning Set (LS) of preclassified states [14,15,16,17]. DTs have been quite extensively studied in the context of various security assessment problems [18,14,16,19,20]. The construction of a DT starts at the root node with the whole LS of preclassified OPs. At each step, a tip-node of the growing tree is considered and the algorithm decides whether it will be a terminal node or should be further developed. To develop a node, an appropriate attribute is first identified, together with a dichotomy test on its values. The test T is defined as $T : A_{ij} = t$. The subset of its learning examples corresponding to the node is then split according to this dichotomy into two subsets corresponding to the successors of the current node. A more detailed technical description of the approach followed is described in [16,17].

The creation of the data set, starts from a base scenario, which is a snapshot of the system at its maximum loading of the year 2005. A coefficient n is considered which takes values from 0.91 to 1.10 with a step of 0.01, and represents the ratio of the total load to the maximum total load. For each of the twenty values of the coefficient n, 200 OPs are created. The total load of these OPs is normally distributed around the value $nP_{\max}$ The Load at each bus of the system is given by equation

$$P_i^D = N(nP_{0,i}^D, \sigma) \tag{11}$$

where N represents the normal distribution, $P_{0,i}^D$ is the demand of bus i at the base scenario and is σ the standard deviation (σ =4%). The generated active power of each unit follows a normal distribution with mean value the production of the unit when the system operates at maximum load, multiplied with the coefficient n, as long as it is between its technical minimum and maximum.

$$P_i^G = \min(P_{i,\max}^G, N(nP_{0,i}^G, \sigma)) \tag{12}$$

where $P_{i,\max}^G$ is the technical maximum of the generator and $P_{0,i}^G$ is the generation of the unit at base case scenario. The simulated disturbance is the loss of a combined cycle unit at Lavrio with 460 MW nominal active power. This disturbance can lead to dangerously low voltage levels, especially in the region of Athens.

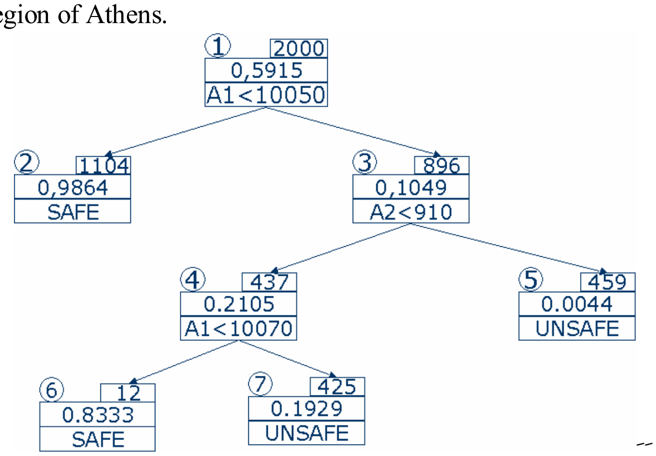

Fig. 2. Decision Tree for DSA

The DT is constructed using as candidate attributes the active load of each of the 9 areas to which the Greek power system is divided as illustrated in Fig. 3, and the total load of the Greek power system (10 attributes). Load Shedding Rules can be derived by inverse reading of the constructed DT. Fig. 2 illustrates the DT, where: A1 is the total active load of Greek Power System in MW and A2 is the total active load of Pelloponissos region in MW. Table IV illustrates the results of the DSA performed by this DT.

TABLE IV.
EVALUATION OF TWO-CLASS CLASSIFICATION PERFORMANCE OF DTS

	Estimation of the DT Model	
	Class A	Class B
Samples of class A	1092	99
Samples of class B	28	781
Success Rate	93.65%(1873/2000)	
False Alarms	8.31% (99/1191)	
Missed Alarms	3.46% (28/809)	

III. UNSUPERVISED LEARNING METHODS

Unsupervised learning methods are used to guarantee a fast DSA and a good representation of the state space. These methods find characteristic groups and structures in the input data space. One of the most successful implementation of unsupervised learning is the Kohonen Self Organizing Map ([21], [22]) A SOM. maps a high dimensional input space to a low dimensional output space. The mapping of the SOM is done by feature vectors w_j, in a way that their main distances to the training vectors are minimized. The feature vectors are structured in a neighborhood grid. If the grid is two-dimensional, the SOM offers the possibility for the visualization of its mapping. During the learning phase the input vectors are presented randomly. At each presentation of an input vector, i.e. at each step of the learning process, every neuron of the network calculates a scalar activation function which depends on the input vector and on its own weight vector wj . This function is chosen to represent distance ||..|| between the input vector and the weight vector of the neuron under consideration. Possible choices are Euclidean distance or the scalar product. The winning unit is considered to be the one with the largest activation. For Kohonen SOMs, however, one updates not only the winning units weights but also all the weights in a neighborhood around the winning units. The neighborhood's size generally decreases slowly with each iteration.

The proposed method has been applied on the data set of section II.C. A 5*X*5 SOM is utilized in order to classify the load profile of the Greek Power System. The load level of the 9 areas of the Greek Power System is used as the input vector for the construction of the SOM Fig. 3 shows the results of the mapping of the testing set onto the SOM At each node of the map the number of OPs that belong to that node, the number of Safe OPs and the number of Unsafe OPs are shown. The nodes with a majority of Safe OPs are considered as Safe nodes, while the others are Unsafe.

One of the advantages of the SOM method is the ease of its construction, since it is independent from the security criterion applied. This is an inherent characteristic of unsupervised learning methods. When the classification criterion is changed, the only modification required is the straightforward recalculation of the numbers shown in Fig. 3 for each of the nodes of the map. In the contrary, a change of the classification criterion would require for example the construction of a new DT.

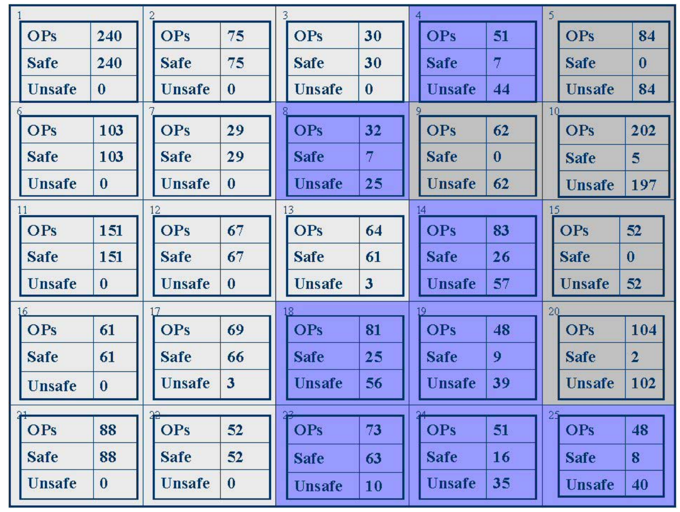

Fig. 3. Mapping of the Testing Set onto the S.O.M.

Table V illustrates the results of the DSA performed by the SOM.

TABLE V.
EVALUATION OF TWO-CLASS CLASSIFICATION PERFORMANCE OF S.O.M.

	Estimation of the DT Model	
	Class A	Class B
Samples of class A	1086	105
Samples of class B	16	793
Success Rate	93.95%(1879/2000)	
False Alarms	8.81% (105/1191)	
Missed Alarms	1.98% (16/809)	

IV. ON LINE DSA

A. Probabilistic Neural Networks

An online supervised learning method for multiclass classification of power system frequency stability, tested with on-line data from an actual power system, is presented in [10]. The method is general and can be applied to all types of stability problems, by simply changing the security classification criteria. It is based on Probabilistic Neural Networks (PNNs) [11], and it is capable of capturing automatically network changes. Fig. 4 illustrates a PNN.

The d components x_1,...x_d of the multivariate input X are forwarded to the hidden layer, whose activation functions (Gaussians) are used to compute the kernel densities $f_j(x)$, with mean μ_j and variance σ_j^2, computed from the learning procedure, for all kernels j = 1,...,K_c, of class c. The outputs of the hidden layer are multiplied by the respective weights ω computed from the learning process and the output unit sums its inputs to compute $p(x \mid c) = \sum_{j=1}^{K_c} \omega_{ij}(x)$.

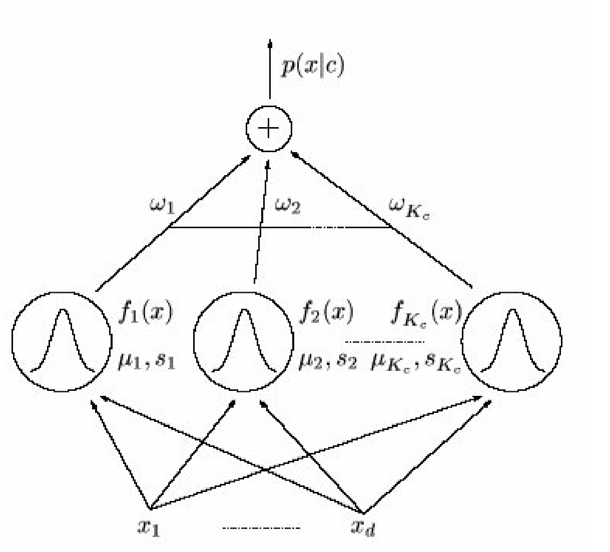

Fig. 4 Probabilistic neural network implementation for one class c with input of dimension d.

Typically there is one PNN for each possible class. Its output provides the corresponding class-conditional Probability Density Function (PDF) p(x|c), where c is the class, and x is the input vector. These probabilities can be used in a subsequent decision-making stage to arrive at a classification. For the input PDF a particular parametric model, called finite Gaussian mixture model, is used. This model allows application of a parametric method of evaluating PDFs [12,13] that can be implemented by a feed-forward neural network. It approximates the unknown PDF by a mixture of Gaussian kernels, the parameters of which are computed from the set of the input samples, which are taken sequentially, one at a time. The number of kernels, their parameters and weights are iteratively estimated from the input samples using the Maximum Likelihood (ML) process. Simple statistical tests involving the mean, the variance and the kurtosis of the kernels are employed in order to decide when a kernel should split in two or when two kernels should join in one. Similarly, a simple test involving the mixing weight of a kernel to decide upon its removal is used. This technique has been preliminary presented in [11].

The security structures developed by the on-line learning method show adaptability to changing operating conditions, when applied to the Power System of Crete. Two distinct operating situations of the system with different distributions (non stationary), are identified. First the system is considered at its current configuration and second its future operation after the installation of a pumped hydro storage plant. For the verification of the method, experiments have been performed using actual load and production time-series data provided by the EMS system. It is shown that the method is capable of adapting its security structures even to drastic changes of the system configuration.

B. Study Case System

The proposed method is applied to the Crete (island of Greece) power system in order to classify its dynamic security by on-line supervised learning. The Crete system is the largest isolated system of Greece, with an installed capacity of 572 MW, a peak load of 514 MW (summer 2002) and a low load of 114 MW. As an isolated system with no interconnections with larger ones, the Crete system experiences frequency stability problems in the event of generation outages. Operation with increased share of wind power exacerbates the problem, as wind turbines are non-controllable sources that cannot provide extra power in case of a power unbalance.

The installation of a pumped hydro plant in the Crete system is currently studied. The operation of this system offers significant economic advantages and allows a better exploitation of the available wind power.

In the proposed application two training sets are used. Each training set consists of 6552 patterns or operating points (OPs), each with its corresponding security status. The first training set corresponds to the current case of the Crete system and is based on actual operating points taken hourly from the database of the Crete Energy Management System. This is termed as system current case. The second training set corresponds to a future case of the Crete system that includes the operation of the hydro plant. This is termed as system future case.

The OPs are labeled accordingly to classes, as follows:

IF $V_{min} \leq 0.9$ THEN the system is insecure (class B) ELSE the system is secure (class A)

C. Results

Fig.5 and Fig.6 show the dynamic variation of the number of kernels (dashed line) for the training set of the future case of the system, during the training process for the secure (A) and insecure (B) class, respectively. In the described application the class-A density is approximated with a number of kernels up to 97 and the class-B density with up to 76 kernels. At each learning step n, the security of the current sample is evaluated using the PNN structures of the previous n-1 steps.

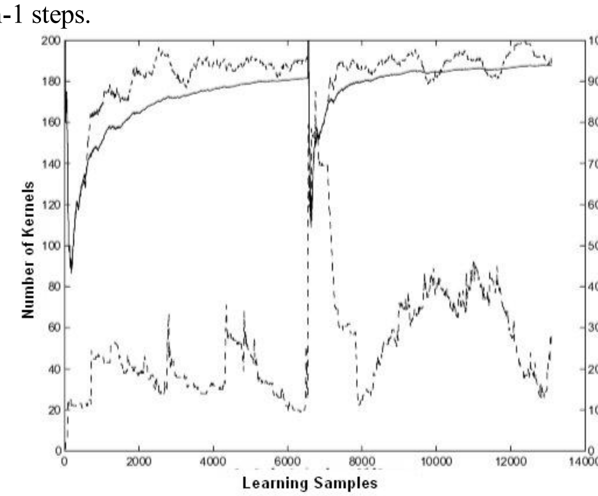

Fig. 5 Dynamic variation of the number of kernels (dashed line) compared to the success rate (continuous and dashed-dotted line) during the class A training process for the current-future case.

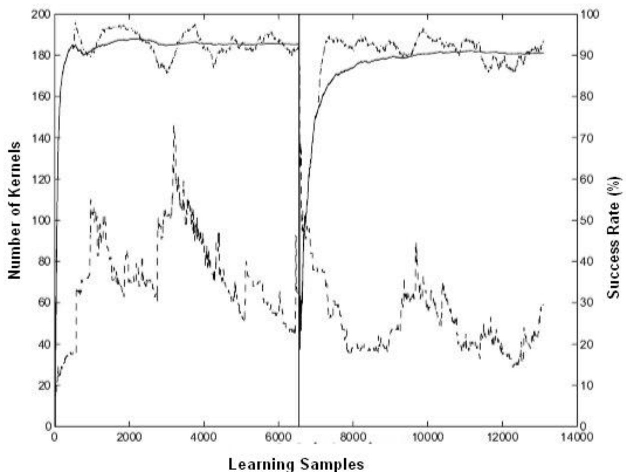

Fig. 6 Dynamic variation of the number of kernels (dashed line) compared to the success rate (continuous and dashed-dotted line) during the class B training process for the current-future case.

This is equivalent to use testing data that are decoupled from the training data. At each learning step, the success rate is calculated with two ways, first from the number of samples that were successfully classified in all previous n steps and second from the number of samples that were successfully classified in the last 500 steps. The second way provides a more objective index, as the first one has a rather masking effect on the calculations. The results are displayed by continuous (all previous steps) and dashed-dotted (last 500 steps) lines in Fig.5 and Fig.6 It can be seen that the success rate exceeds 90 percent after the first 4135 samples for the secure state and after the first 485 samples for the insecure state in case of testing with all previous samples.

V. CONCLUSIONS

The results from the application of the machine learning techniques show the accuracy and versatility of the methods. RBFNNs and SVM perform not only classification of the systems states but also regression and give an estimation of the security criterion (voltage and/or frequency) value. This is very important as it can be used as a measure for the security margin of the system. Decision Trees on the other hand provide explicit rules to the operator of the system, while the inverse reading of them can also establish load shedding schemes when the safety of the system is jeopardized. The advantage of the SOM in comparison to the offline supervised learning methods is that its construction is independent from the security criterion applied. This means that when the classification criterion is changed, the only modification required is the straightforward recalculation of the security indices for each of the nodes of the map. In the contrary, a change of the classification criterion would require the reconstruction of any of the offline supervised learning methods new DT. The advantage of on line learning of the PNNs is that it can deal with the changes in the structure of a power system without the need of completely retraining the PNN.

VI. REFERENCES

[1] Task Force 21 of Advisory Group 02 of Study Committee 38. "Power system security assessment." CIGRE Technical Brochure, 2004.

[2] Power Systems Engineering Research Center, "Integrated security analysis final project report". May 2003.

[3] Aggoune M.E. Vadari, S.V. "Online assessment of power system dynamic security using modular neural networks". *In Proc. 1991 Expert System Application to Power Systems*, pp. 338-44

[4] R. J Marks CA Jensen and M.A. El-Sharkawi, "Power system security assessment using neural networks: Feature selection using fisher discrimination". *IEEE Trans. on Power Systems*, vol. 16, no.4, pp. 757-763

[5] A. E. Gavoyiannis, E. M. Voumvoulakis, and N. D. Hatziargyriou. "Radial basis functions for dynamic security assessment of power systems with increased wind power penetration". presented at the PSCC meeting, Liege, Belgium, 2005.

[6] Vladimir Vapnik. *The nature of statistical learning theory,* New York, Springer-Verlag, 1995.

[7] A. E. Gavoyiannis, D. G.Vogiatzis, D. P. Georgiadis, and N. D. Hatziargyriou. "Combined support vector classifiers using fuzzy clustering for dynamic security assessment". *In Proc. 2001 IEEE PES Summer meeting ,Vancouver,Canada,* vol. 2, pp1281-1286 .

[8] A. E. Gavoyiannis, D. G.Vogiatzis, and N. D. Hatziargyriou. "Dynamic security classification using support vector machines". *In Proc. IEEE Int. Conf. on Intell. Syst. Applicat. Power Syst.*,Budapest, Hungary, 2001, pp. 271-275.

[9] A. J. Smola and B. Scholk, "A Tutorial on Support Vector Regression", NeuroColt 1998, NC2-TR-1998-030, http://www.neurocolt.com

[10] E. Gavoyiannis, E. M. Voumvoulakis and N. D. Hatziargyriou, "On-Line Supervised Learning for Dynamic Security Classification Using Probabilistic Neural Networks", *In Proc. 2005 IEEE PES Summer meeting San Fransisco,* June 2005, vol.3, pp. 2669-2675

[11] N. Vlassis and A. Likas. "A kurtosis-based dynamic approach to gaussian mixture modeling". *IEEE Transactions On Systems, Man, and Cybernetics,* Part A 29(4), pp. 393-399, 1999.

[12] R. L. Streit and T. E. Luginbuhl. "Maximum likelihood training of probabilistic neural networks". *IEEE Trans. on Neural Networks*, vol. 5, no. 5, pp.764-783, 1994.

[13] H. G. C. Traven. "A neural network approach to statistical pattern classification by semiparametric estimation of probability density functions". *IEEE Trans. on Neural Networks*, vol. 2, no. 3, pp. 366-377, May 1991.

[14] N. Hatziargyriou D. Georgiadis, E. Karapidakis. "Application of machine learning technologies on the dynamic security of isolated power systems with large wind power penetration". *In Proc. 2002 SETN, Thessaloniki,* pp. 257–268, .

[15] P. Georgilakis and N. Hatziargyriou. "On the application of artificial intelligence techniques to the quality improvement of industrial processes". *In Proc. 2002 SETN, Thessaloniki,* pp. 473–484.

[16] Pavella M. L.Wehenkel. "Decision trees and transient stability of electric power systems". *Automatica*, vol. 27, no. 1, pp.115–134, 1991.

[17] L. Wehenkel. *Automatic Learning Techniques in Power Systems,* Kluwer Academics, 1998.

[18] P. Kundur, *Power System Stability and Control*, Mc-Graw Hill, 1994

[19] S. Papathanassiou N. Hatziargyriou and M. Papadopoulos. "Decision trees for fast security assessment of autonomous power systems with large penetration from renewables". *IEEE Trans. Energy Conv.*, vol. 10, no. 2, pp.315-325, June 1995.

[20] J. Thorp D. Brown S. Rovnyak, S. Kretsinger. "Decision trees for real-time transient stability prediction". *IEEE Transactions on Power Systems*. vol. 9, no. 3, pp 1417-1426

[21] Alain J. Germond Dagmar Niebur. Power system static security assessment using the kohonen neural network classifier. *IEEE Transactions on Power Systems*, vol. 7, no. 4, pp. 865–871, 1993.

[22] Carsten Leder Christian Rehtanz, "Stability assessment of electric power systems using growing neural gas and self-organizing maps", *In Proc. 2000 European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 401–406.