Distance Rejection in the Context of Electric Power System Security Assessment Based on Automatic Learning

Isabelle HOUBEN and Louis WEHENKEL*

University of Liège - Sart-Tilman, B28, B-4000, LIEGE, BELGIUM
Email : lwh@montefiore.ulg.ac.be
* *Research Associate, F.N.R.S.*

Abstract. Automatic learning methods have proved to be quite attractive in the context of power system security assessment. The complementarity of various methods proposed so far, lead us to combine them in a *toolbox* in order to exploit their advantages and discard their limitations. In this paper, we show how the nearest neighbor approach could be used to face the problem of detecting outliers, i.e. cases not well enough represented in the data base used to learn the models. More precisely, such detection can be based on distance rejection which implies the choice of an appropriate distance. On a particular real life problem, we show how the simple nearest neighbor in the candidate attributes space allows to reject such cases.

Keywords : power system security assessment; automatic learning; nearest neighbor; distance rejection.

1 Automatic Learning Based Approach to Power System Security

Security is the ability of a power system to withstand in a satisfactory way all kind of external or internal disturbances. It is a concern of paramount importance for the proper design and operation of electric power systems. The increase in international interconnexions and free access organizations together with social requirements for uninterrupted electric supply implies increasing intricacy and at the same time increasing importance of security aspects.

Broadly, security covers steady-state and dynamical aspects. Also, in a broad sense, security studies are carried out within three different contexts : planning, operation planning and real-time operation. Traditional (analytical) approaches mainly consist of integrating numerically the non-linear equations of motion in the time-domain or of eigenvalue analysis of the linearized system in the frequency domain.

The limitations of analytical methods are linked to their inherent weaknesses; in particular : (i) lack of synthetic information (which are the driving parameters? how do they influence the phenomena? (sensitivity); "what if" ? (means to control)); (ii) overly specific and very detailed case by case information; in short : "black-box" type of information in that they do not provide proper hints about

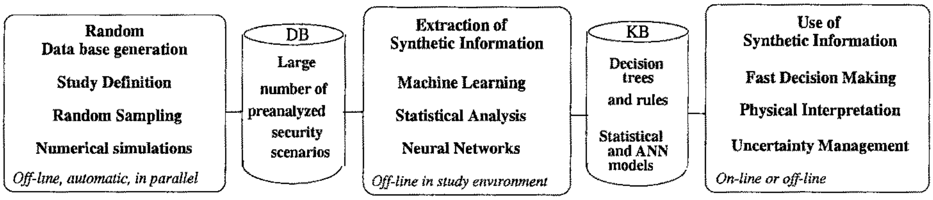

Fig. 1. Automatic learning framework for security assessment

the parameters influencing the phenomena nor suggestions to control; (iii) involved heavy real-time computations; (iv) inability to account for uncertainties.

Automatic Learning (AL) in general is concerned with the design of automatic procedures able to learn a task on the basis of a learning set of solved instances of this task. Three main families of AL methods may be distinguished, namely : (i) machine learning, a subfield of symbolic artificial intelligence; decision trees are members of this family (ii) statistical pattern recognition and regression; (iii) artificial neural network based learning. In what follows, we focus on non-parametric AL methods. Indeed, parametric methods make too strong assumptions about the shape of probability densities or classification boundaries; they therefore would not be adequate enough for the wide variety of security problems.

In the particular context of power system dynamic security assessment, the AL approach may be schematically described by Fig. 1 (see e.g. [1]) : random sampling techniques are considered to screen all relevant situations in a given context, while existing numerical simulation tools are exploited - if necessary in parallel - to derive detailed security information. The heart of the framework is provided by AL methods used to extract and synthesize relevant information and to reformulate it in a suitable way for decision making. This consists of transforming the data base (DB) of case by case numerical simulations into a power system security *knowledge base* (KB). As illustrated in the figure, a large variety of AL methods may be used in a toolbox fashion, according to the type of information they may exploit and/or produce. The final step consists of using the extracted synthetic information (decision trees, rules, statistical or neural network approximators) either in real-time, for fast 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.

2 Overview of AL Methods Used in Security Assessment

Two broad types of AL problems may be distinguished : supervised and unsupervised learning. While supervised learning usually aims at constructing a model for an assumed relationship between input and output parameters, unsupervised learning (or clustering) essentially aims at either uncovering similarities. Note that in power system security assessment, input attributes are variables describing the state and topology of a power system, and the output information characterizes its security with respect to plausible disturbances. Both types of variables may be either continuous/numerical or discrete/qualitative.

2.1 Unsupervised Learning and Clustering

Unsupervised learning is not oriented towards a particular prediction task. Rather, it tries to identify existing relationships, either among objects, or among attributes. Thus, one of the purposes of clustering is to identify homogeneous groups of similar objects, in order to represent a large set of objects by a small number of representative *prototypes*. Graphical, two-dimensional scatter plots may be used to analyze the data and identify clusters. Another application concerns the identification of similarities (and redundancies) among the different attributes used to characterize objects.

Note that large scale power system security data bases may contain several thousand samples, described by several hundred attributes. Thus, unsupervised learning is useful in order to explore large data bases to find groups of similar problems, and also to reduce problem dimensionality by identifying correlated variables.

2.2 Supervised Learning

In what follows, we discuss the three classes of methods providing three *complementary* types of information, focusing on those aspects which are most relevant to the subsequent discussions.

Symbolic knowledge via decision trees [2]. *Top down induction of decision trees* (TDIDT) is one of the most successful classes of machine learning (i.e. symbolic learning) methods. A main asset of decision trees (DTs) lies in the explicit and logical representation of the induced classification rules and the resulting explanatory capability. In particular, the tree induction method provides systematic correlation analyses among different attributes and identifies the most discriminating attributes. From the computational viewpoint it is efficient at the learning stage as well as at the prediction stage.

There are two generalizations of decision trees of interest in the context of dynamic security assessment. First, *regression* trees which infer information about a numerical output variable [3]. Second, *fuzzy* trees which use fuzzy logic instead of standard logic to represent output information in a smooth fashion [4].

In power system security assessment, the main asset of decision tree induction lies in its capability to extract simple and interpretable rules from large scale data bases. In particular, it is able to efficiently scan a very large number of candidate attributes, *select* the most relevant ones and determine their importance in terms of the *information quantity* they provide on the output information.

Smooth nonlinear approximations via artificial neural networks [5]. We restrict ourselves to multilayer perceptrons, MLPs for short. Their first or *input* layer corresponds to the attribute values, and the last or *output* layer to the desired security classification or margin information. Intermediate layers enable the network to approximate arbitrarily complex input/output mappings, provided that its topology and its weights are chosen properly.

One of the difficulties with MLPs comes from the very large number of weights and thresholds related in a nonlinear fashion, which makes it almost impossible to give any insight into the relationship learned. All in all, one can say that MLPs offer a flexible, easy to apply, but essentially black-box type of approach to function approximation.

In power system security assessment MPLs are most well adapted to infer information about security *margins*, which behave generally in a smooth fashion.

Memory based reasoning via statistical pattern recognition [6, 7]. The previous two approaches essentially compress detailed information about individual simulation results into general, more or less global security characterizations.

Additional information may however be provided in a case by case fashion, by matching an unseen (e.g. real-time) situation with similar situations found in the data base. This may be achieved by defining generalized distances so as to evaluate similarities among power system situations, together with appropriate fast data base search algorithms.

A well known such technique is the "k Nearest Neighbors" (kNN) method able to complete decision trees and multilayer perceptrons. The main characteristics of this method are high simplicity but sensitivity to the type of distances used. In particular, the kNN method is sensitive to irrelevant and/or redundant attributes, which are frequently encountered in security assessment problems.

Let us briefly describe the method developed in order to adapt the kNN method to such problems. It combines information obtained by decision tree induction and genetic algorithms so as to adapt the kNN parameters in a supervised learning stage. The latter consists of the following three steps [8, 9] :

1. selection of the relevant attributes, for the particular problem under consideration, by decision tree induction;
2. adjusting their weights in the Euclidean distance, in order to take into account their respective impact on the problem of concern (initial guess provided by the information quantities (IQ) obtained as a byproduct in decision tree induction; refinement using genetic algorithms GA);
3. choice of the appropriate value of k.

These parameters are determined using the learning set, together with the leave-one-out method to appraise generalization to unseen states. The final result is assessed using an independent test set.

Thus the learning procedure summarizes to : for the learning set at hand, build a decision tree to identify its test attributes and appraise their corresponding IQ. Note that since the computing effort for building DTs increases at most linearly with the number of candidate attributes, it is possible (and advisable) to use as many candidates as deemed necessary, in order to avoid missing relevant ones. The obtained distance gives rise to the kNN-DT+IQ method.

Common sense suggests and experience confirms that combining the information provided by DTs with the simple search of k contributes to improve significantly kNN's accuracy. Yet, it should be possible to improve it further. This is suggested by the fact that DTs provide synthetic information which moreover is obtained via local optimizations at the successive test nodes. It may therefore hide part of the detailed information and result in a globally sub-optimal solution. Hence, the identification of the relevant attributes and the appraisal of their weights may be suboptimal. Obviously, appropriate procedures relying on a global optimization such as GAs are able to further improve the kNN-DT design.

3 Use of *k*NN for Distance Rejection

3.1 Practical Motivation

A possible cause of automatic learning methods' failure to correctly assess unseen cases is the existence of outliers. These are cases which go beyond the generalization capabilities of a method, because they are not "well enough" covered in the data base used to train it.

In power system security assessment, data bases are generated by random sampling. The specifications of random sampling are defined from prior expertise about the problem. They are generally biased with respect to actual statistics, since it is necessary to ensure a good representation of insecure states, which are very infrequent in real life. In this process, the best is made in order to ensure the representativity of the data base, however there is no guarantee that this objective is fully reached. Thus, the criteria extracted (decision trees, neural nets, nearest neighbor classifiers...) are obviously valid only in the range of situations scanned in the data base from which they have been derived.

For example, one reason for an insufficient coverage is that some driving parameters have been overlooked and kept constant while generating the data base; it is then hazardous to establish valid similarities between a new case and cases of this data base. Another reason is that the new case is "far away" from all others, i.e. it has no "close enough neighbors". The former reason comes from insufficient expertise of the phenomena; the latter from insufficient scanning of the attribute space. Our aim is to focus on this latter cause and detect the corresponding outliers.

Below, we will describe the results obtained with a very simple distance rejection scheme, in the context of a large data base, stemming from a transient stability assessment problem of real large scale system. In particular, we will find out that in order to detect outliers it would be dangerous to exploit distances correlated with the data base (which happens if the distance is tuned by supervised learning). A more appropriate choice consists in using a "neutral" distance, based only on normalizing the attributes deemed relevant by the experts.

3.2 Detection Procedure

As was already said, the use of nearest neighbor techniques to assess the security of unseen cases requires the choice of an appropriate distance depending the problem at hand. However, as shown below, such a distance is not able to provide a procedure to detect outliers, i.e. cases not well enough represented in the data base used to learn the distance.

For a given data base (DB), in order to decide whether a new (unknown) case is a "normal" one (i.e. well enough represented in the DB) or whether it should be labeled as an outlier, we propose the following procedure.

(i) Consider the multidimensional space defined by the entire set of candidate attributes deemed relevant by the experts; compute the distance of the DB states to their respective nearest neighbor. Let d_{max} denote the maximum distance thus found.

(ii) Proceed similarly with the given new state, i.e. compute its distance, d_u, to its nearest neighbor in the above multidimensional space. If $d_u < d_{max}$ consider it to "belong" to the DB, i.e. to be a "normal case". Otherwise, declare it to be an outlier.

The computation of d_{max} in step (i) has to be performed once and for all, whereas the computation of d_u in step (ii) has to be repeated with each new case.

3.3 Evaluation on a Real Life Problem

We will illustrate the above procedure on a stability problem of the 22-North configuration of the Hydro-Québec power system (see [9] and [10]). To this purpose, we will consider two DBs : (i) the "normal" DB built for this configuration (denoted for short 22N); (ii) an "abnormal" DB, labeled "31-North" (31N for short), and built for another stability subproblem of the same power system. Each case is described by 74 candidate attributes and by one class, stable or unstable.

The 22N DB is decomposed in two sets :

1. the learning set (LS), composed of 2746 states (59 % unstable and 41 % stable ones) and used as reference DB, to build decision trees and learn the distance for the kNN, if necessary.
2. the test set (TS), composed of 657 states (61 % unstable and 39 % stable ones).

The 31N DB is denoted by TS2 and is composed of 1450 states (72 % unstable and 28 % stable ones).

A first series of simulations consists of classifying the cases of TS2 by the DT and the kNN-DT+IQ methods, both trained with the 22N learning set (the decision trees selects 11 attributes among the 74 candidate ones, 95% of the information being concentrated in the first 5 ones). The results are reported in Table 1. The corresponding error rates show that, obviously, the TS2 cases cannot be correctly classified by models (DTs or kNN classfiers) derived from

Table 1. 22N and 31N test states classified with kNN trained with 22N learning states

Method	Pe^{TS} (%)	Pe^{TS2} (%)
DT	3.5	22.4
"Trained" kNN	2.7	20.8

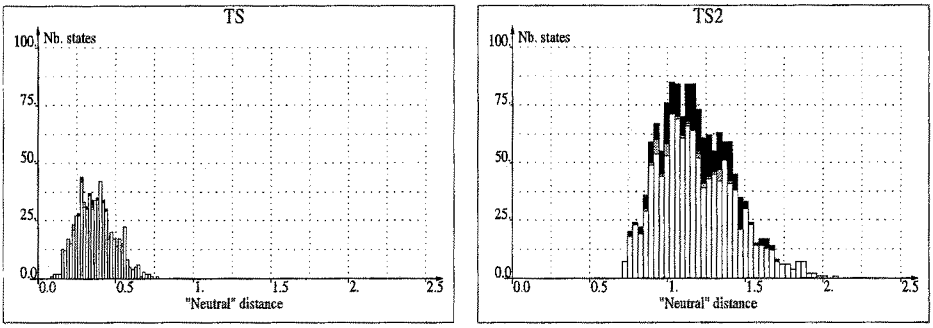

Fig. 2. "Neutral" distances to NN (NN found in 22N LS by "neutral" distance)

the 22N learning set. It should be noted that the assessment of an outlier by an automatic learning method trained with the DB of concern is not necessarily wrong; simply, one cannot guarantee its correctness.

A second series of simulations uses the detection procedure applied in the following way :

(i) consider the 22N learning set in the 74-dimensional space of all 74 candidate attributes, with uniform weights combine in the so-called "neutral" Euclidean distance;

(ii) compute in the above space the distances of all 22N test states to their respective nearest neighbor;

(iii) compute in the same space the distances of all 31N (i.e. TS2) states.

The bar diagrams of steps (ii) and (iii) are plotted in Fig. 2. In these diagrams, the overall height of the bars correspond to the number of states in the corresponding "neutral" distance interval. The white part of it corresponds to states which are correctly classified using the kNN method using the "trained" distance; the other parts correspond to different types of errors (dangerous non-detections, in black; false alarms and marginal non-detections in grey).

Figure 2 obviously shows that the two distributions are well separated. Indeed, choosing $d_{max} = 0.78$, will discard all TS2 cases as being "outliers" except for 20. Note, however, that among the latter 20 cases which lie below $d_{\max}$ only one is misclassified by the methods trained on the 22N data base. Actually, the above observations suggest that the 20 cases are not outliers, and corroborate the proposed procedure.

Note that using for the outlier detection procedure the "trained" distance, would be completely misleading. This is clearly shown by the complete overlap

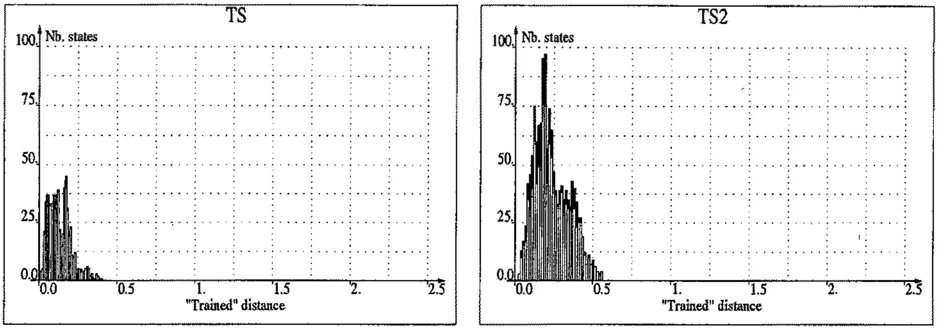

Fig. 3. "Trained" distances to NN (NN found by "trained" distance)

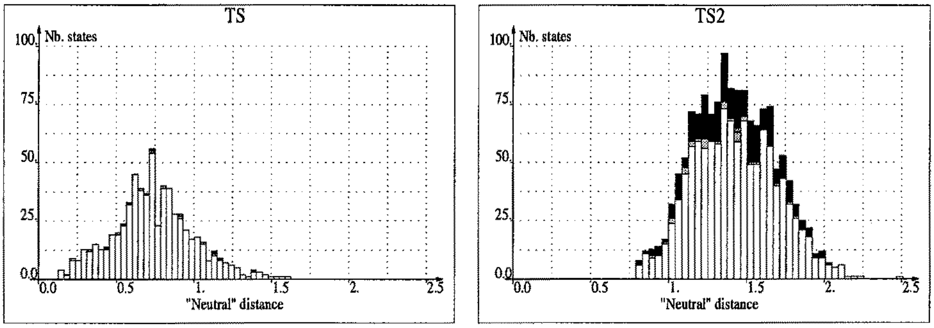

Fig. 4. "Neutral" distances to NN (NN found by "trained" distance)

of the two diagrams in Fig. 3. This is because the selected attributes and their weights in the adapted distance have been determined in the absence of the outliers that one precisely wants to detect in order to decide whether they belong to the same "problem" (i.e. DB). Their use would therefore bias this test.

Figure 4 shows the distributions obtained in an experiment combining the two distances. The nearest neighbor of a state is found using the "trained" distance, then the "neutral" distance among these two states is computed. From the computational point of view this procedure would be much faster, since the nearest neighbor search would be carried out using a much smaller number of attributes (11 instead of 74). Unfortunately, the two diagrams in Fig. 4 still overlap significantly, and a large number of outliers would not be detected.

As a complementary exercise, we have interchanged the two data bases, viz., we have trained the *k*NN models with a learning set composed of the 31N states and tested the 22N test states : the conclusion is symmetric : here, the outliers are found to be the 22N states.

We conclude that in order to detect outliers a "neutral" distance should be used, which is uncorrelated with the data base.

4 Conclusion

This paper shows that distance rejection allows to detect outliers. However this procedure implies two mains problems. The first one concerns the choice of the distance. Our experiments show that a “trained” distance automatically learned for security assessment is not adapted to outlier detection. The rejection can only be based on a “neutral” distance defined by experts who select the attributes potentially influencing the security problem. Consequently, it is harder to find a good distance. The second problem concerns the validation of the procedure because outliers have by definition a low probability of occurrence.

References

1. L. Wehenkel. *Automatic learning techniques in power systems.* Kluwer Academic, Boston, 1998.
2. J. R. Quinlan. Induction of decision trees. *Machine Learning* 1, pp. 81–106, 1986.
3. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Wadsworth International (California), 1984.
4. X. Boyen and L. Wehenkel. Automatic induction of fuzzy decision trees and its application to power system security assessment. To appear in *Fuzzy Sets and Systems*, 1998.
5. S. Haykin. *Neural networks. A comprehensive foundation.* IEEE Press, 1994.
6. R. O. Duda and P. E. Hart. *Pattern classification and scene analysis.* John Wiley and Sons, 1973.
7. P. A. Devijver and J. Kittler. *Pattern Recognition : A Statistical Approach.* Prentice-Hall International, 1982.
8. I. Houben, L. Wehenkel, and M. Pavella. Genetic algorithm based k nearest neighbors. In *Proc. CIS-97, IFAC Conf. on Contr. of Indust. Syst.*, pp. 43–48, Belfort, Fr, 1997.
9. I. Houben, L. Wehenkel, and M. Pavella. Hybrid adaptive nearest neighbor approaches to dynamic security assessment. In *Proc. of CPSPP'97*, pp. 685–690, Beijing, 1997. International Academic Publishers.
10. L. Wehenkel, I. Houben, M. Pavella, L. Riverin, and G. Versailles. Automatic learning approaches for on-line transient stability preventive control of the Hydro-Québec system - Part I. Decision tree approaches. In *Proc. of SIPOWER'95, 2nd IFAC Symp. on Control of Power Plants and Power Syst.*, pp. 231–236, Dec. 1995.