A Flexible Machine Learning Environment for Steady State Security Assessment of Power Systems

D. D. Semitekos, N. M. Avouris, G. B. Giannakopoulos
University of Patras, ECE Department, GR-265 00 Rio Patras, Greece
Email: {dsem, N.Avouris, G.B.Giannakopoulos } @ee.upatras.gr

Abstract

Machine Learning techniques have been applied to power system analysis for a number of years. The need for a flexible computing environment to support these studies is derived by the complexity of the process, the volume of data often used and the diversity of the applied tools and techniques that span many disciplines. A data warehouse can be a central component of this environment. In this paper our experience with building and using such an environment is described. Data collection and transformation tools, machine-learning tools and testing tools have been integrated in the MLPSE environment described here, used for steady state security assessment of power systems. It is argued that the proposed approach can be applied to many similar power systems analysis studies.

Keywords data warehouse, decision trees, machine-learning, steady state security assessment, contingency analysis

1. Introduction

MLPSE (Machine Learning for Power System Environment) is a flexible computing environment for the steady state security assessment of power systems using machine learning inductive techniques. MLPSE is built around a data warehouse and combines multiple tools, some of which have been developed in the process of the reported research, like the feature selection tool, the operating points definition tool etc, while some others were developed and made available by other researchers, like the Load Flow analysis package (Grady, 1995) and the Machine Learning toolkit (Weiss, 1998). A data warehouse (DW) is the central component of the architecture, shown in figure 1. A DW stores complex data of an enterprise, designed to facilitate their manipulation for decision-making and analysis (Kimball, 1996, Kroenke, 2000).

In this paper it is argued that building and using an environment like MLPSE, can be a useful approach for many power system analysis studies. In particular when the analysis has as an objective to develop run time models through machine learning approaches, i.e. using data intensive techniques, this approach can facilitate and automate the process. MLPSE was built and subsequently used for performing a steady state security assessment of part of the PPC system and for development of a run time contingency analysis module using multiple decision trees. In the following the MLPSE environment and a case study demonstrating its typical use are described. Some of the implications of using integrated machine learning environments in power system analysis studies are discussed in the final part of the paper.

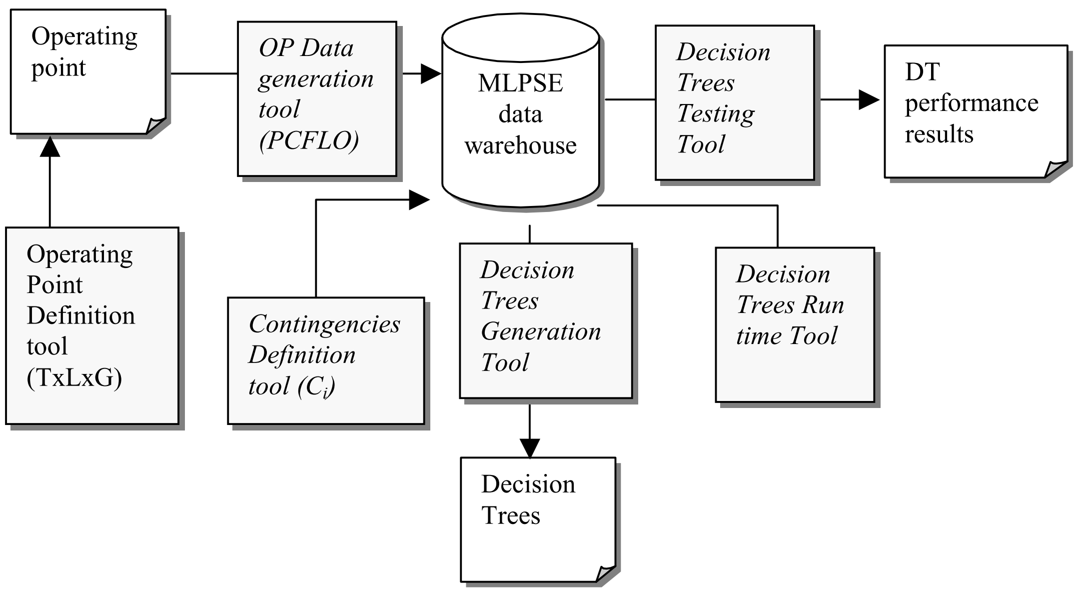

Figure 1. The MLPSE environment modules

2. The MLPSE architecture

The basic architecture of MLPSE is shown in figure 1. Two operational modes are defined. The training and testing mode

In training mode, the MLPSE tools are used for producing a system able to automatically rank predefined contingencies for future system operating points. Four (4) major modules are involved in this process: (a) The *operating points simulation* and system performance criteria module. (b) The *contingencies definition* and processing module. (c) The *features selection* module. (d) The *automatic learning* module.

In testing mode MLPSE supports definition of a testing data set. This can be used for evaluation of the developed system performance.

The run time simulation tools are independent components. They are implemented in two standalone modules. The *Run time control room adviser simulation* and the *Performance-testing* module. In the following sections some typical modules of MLPSE and an example of their use is provided.

3. The operating points definition module

Automatic learning methods need a rich representative data set of the problem space, which can be used for training and subsequently testing performance of the developed systems. The source of this data set in the case of power systems applications can be the supervisory control system of the network supported by the state-estimation application through which a number of snapshots of typical system operating points can be collected. An alternative approach is to generate a typical data set through simulation of operating points using tools like load flow analysis. The latter approach has been used during our study.

A variety of operating points that represent various load conditions, generation schemes and network topologies have been defined for the network under study. The MLPSE tool provides for the operator the opportunity to define a number of distinct operating points along these dimensions. Operating points are produced by simulation according to the following process: A base case maximum load operating case is considered. Other operating points are defined using combinations of *load*, *network connectivity* (topology) and *generation sequence* schemes.

(a) *Active and reactive load levels* are considered between a maximum and a minimum load level. All loads of the network are uniformly distributed according to a load pattern.

(b) The *network connectivity* states are derived from the base case, considering the elimination of one or more transmission lines, buses and other elements from the network.

(c) The *generators activation sequence* is pre set for every generation scenario. The number of the generators to be activated and the generation level is computed through successive load flows taking in consideration economic operation levels

The feasibility of each operating point produced is checked through a load flow study and is validated through the *System Performance Criteria* module. Invalid operating points, i.e. for topological reasons - (formation of islands in the network) are also discarded. Operating points with violating performance criteria are discarded, considered not realistic operating points. The total number of the valid operating points simulated, is according to the above algorithm, less or equal to the Cartesian product of T_i x G_j x L_k, where T_i is the i network topology states, G_j is the j generation sequences schemes and L_k is the k load levels. Valid operating points are stored in the Data Warehouse in a database. Operating points' parameters are also kept in the MLPSE Data Warehouse.

3.1. The Operating Points parameters generation module

This module includes a number of parameters concerning each operating points' network description and are stored in the data warehouse tables named *Features* and *Contingencies*. The table *Features* is related with the Operating Points data. Each defined contingency is applied to a copy of the stored operating point data.

A load flow study is run at the automatic learning training phase and the results, simulating the severity of the particular contingency to the network in the particular operating point, are stored in the table *Contingencies*.

The system parameters described in the table *Features*, refer to a pre - contingency simulated operating cases.

4. Steady state security assessment indices selection

Many indices have been proposed in the literature as criteria for steady state security assessment (Albuyeth et al. 1982, Ejebe &Wellenberrg, 1979, Wehenkel, 1998 etc). These involve overloaded lines, or bus voltages that deviate from the normal operation limits. However, violations reported are not of the same importance. For instance, many minor overloads in a set of lines may be of minor importance with regard to a single major violation in an important line and vice versa. These are the "*masking effect*" problems and a way to face them is with the assignment of weighting factors in the indices to be used.

As discussed in (Ejebe et al. 1979) the form of the index is such that a contingency that produces, for example, a single reactive power violation may be ranked as more severe than another contingency, which produces abnormal voltages at several buses. This *masking phenomenon* occurs because the percentage value of a reactive power violation can be higher than the percentage value of a voltage violation. If we are to use similar indices for state estimation reasons, it is preferable to apply them on the entire network without any discrimination on specific components. In MLPSE Data Warehouse operating points and system indices are stored under a form that can be useful for automatic learning tools, so that the effects of a certain contingency on an operating state can be automatically deduced.

With these considerations, in MLPSE aggregate pre-contingency system features like the total active generation and the total active load, can be of importance in the process of automatic learning. The proposed system state estimation criteria are the following:

I. Line Overload Index:

$$LOI = \sum_{i=1}^{NL} \left(MVALine - MVALineLimit\right) \cdot MVALineLimit$$

where:

MVALine = the apparent power in MVA

MVALineLimit = the apparent power Limit in MVA

NL = Number of Lines

The proposed *LOI* index aims at avoiding the masking effect related problems weighting the contribution of each contingent violation of MVA transfer limits according to the importance of the line where the violation occurs.

II. A composite Voltage profile index

An Index to count any Major Violations occurring at any bus.

$$MA_{VVI} = \sum_{i} VB$$

where

MA_{VVI} = Major Voltage Violations Index

VB = Buses where Major Voltage Violations occur (i.e. for which Voltage Magnitude per unit divergence in absolute value terms is greater than 10%, regardless of the Voltage Bus Base).

i = Index covering all system buses.

A second Index to count minor violations is proposed as follows:

$$MI_{VVI} = \frac{\sum_{i=1}^{NB} abs\left(VMBus_i - VMLim\right)}{NVVB}$$

where

MI_{VVI} = Minor Voltage Violations Index

$VMBus_i$ = Voltage i-Bus Magnitude expressed in per unit terms

VMLim = The upper or lower per unit limit for over voltages and under voltages respectively. A 5% limit is proposed, (configuring the upper limit at 105% p.u. and the lower limit at 95% respectively for each bus)

NB = Number of Buses

NVVB = Number of voltage violating buses (within a 5% proposed limit)

On MI_{VVI} a state estimator hypothesis in built in MLPSE comparing the index with a global percentage safety limit. Thus for a 5% proposed limit, any operation point bearing an MI_{VVI} index greater or equal to 1.05 is regarded as violating and consequently “unsafe”.

5. Selection of Operating Point features

While the indices proposed so far aim at decision making about whether a certain operating point is safe or unsafe and from a simulation point of view acceptable or unacceptable, system features aim at just benchmarking an operating point from a given operational aspect. From the many parameters describing a power system operating point, the selection of the most suitable ones to be used in an automatic learning process is a difficult task. This is known as the "feature selection problem" in machine learning literature.

Operating point features used in MLPSE are network aggregates. Their major advantage is that they can be easily applied, being network topology independent. This is not the case with features that refer to specified lines or buses, where selection is biased.

The proposed metrics refer exclusively to pre contingency operating points. This is because MLPSE aims at building learning modules for automatic contingency ranking. Post contingency metrics would be of no value for the given purpose, as they would require load flow studies for every contingency under consideration, an approach that requires high computational and time resources. In MLPSE the following features are proposed:

Load level, *total number of lines*, *total rating* of all lines that is the sum of apparent power MVA limits, *total active load*, *total active generation*, *total reactive load*, *total reactive generation*, *total apparent power* flowing in all lines, *voltage stability index* as described by Wehenkel (1998), *active apparent power margin index* expressed as the fraction on the total apparent power flowing for all lines, over their total MVA limits.

The selection of a subset of these features in experimentation with automatic learning techniques in the MLPSE environment is discussed in the following

6. An example of an MLPSE study

Various automatic learning algorithms are available in the MLPSE environment. In the experiment described here decision trees are used. Decision trees have proven to be valuable tools for the description, classification and generalization of data (Murthy, 1998). The application of decision trees to on-line steady state security assessment of a power system has also been proposed by Hatziargyriou et al. (1994). Other researchers have applied alternative automatic learning techniques like neural networks (Chan, Dunn, Daniels and Edwards, 2000), (Srivastava, Singh and Sharma 1998), and probabilistic networks (Mijuskovic, Stojnic 2000) to the same problem.

According to our approach a separate DT has been built for each contingency. The available data, representing operational points, have been randomly divided in training data and testing data at 70% - 30% quotas respectively.

The selection of the most suitable features, among those proposed in section 5, is a non-trivial process. One of the most powerful characteristics of MLPSE is the flexibility that provides the user in selection of the most suitable features. This selection process involves search for salient features that describe well the system. It should be mentioned that there are many techniques, derived from power systems theory and operation experience as well as statistics and machine learning algorithms supporting this process. For instance the decision tree-learning algorithm has always the ability to assess, which fields of input data are of greater value. Overloading the algorithm with excessive number of irrelevant or redundant features reduces the effectiveness and the efficiency of the learning algorithm.

In the described experiment a steady state security assessment system has been developed for the network of the island of Crete in Greece, a power system comprising 61 busbars and 78 lines. Seven (7) contingencies have been identified as most interesting for the study.

(a) *First experiment*. In the first case of the reported study the selected features have been the *Voltage Stability Index* and *Real Power Margin Index*, discussed in section 5. The ranking of the contingencies is based on the outcome of the load flow study that can result in no-convergence= severity high, operational violations=severity medium, no violations = severity low. Seven decision trees have been trained. Using the test data set the performance of these DTs is measured, shown in Table 1. The overall success rate, for all Contingencies is in average **82.3%,** with worse performance relating to contingency 4 (74%) and best performance of the decision tree of contingency 3 (97%).

(b) *Second experiment.* In a second experiment a different set of features has been used. In this test the number of features used is three: the *Overall load level*, expressed as a percentage over the maximum load level and *Voltage Stability Index* and *Real Power Margin Index* that were used in case (a). The performance of the DTs developed in this experiment is also shown in Table1. It is observed that the addition of the extra feature in this second study has not contributed towards improvement of the developed system performance.

	Case A	***Case B***
Contingency	***Prediction Success Rate***	***Prediction Success Rate***
1	*84%*	*84%*
2	*78%*	*75%*
3	*97%*	*97%*
4	*74%*	*74%*
5	*81%*	*81%*
6	*82%*	*82%*
7	*80%*	*80%*
Overall	**82.3%**	**81.9%**

Table 1. The performance of the developed contingency ranking models

The presented example demonstrated the trial-and-error approach in selecting features to represent operating points and the effect of this process on the overall system performance. Once a minimal optimal set of features has been selected, the system can be tested using the run time module, i.e. continuously feeding it with values, representing different system conditions, upon which it decides on the severity of the selected contingencies.

7. Conclusions

This paper discussed a set of computational tools that support automatic learning for power system security assessment. The environment is based on a data warehouse approach. The advantage of this approach is the flexibility of the performed studies. This data warehouse is a repository of various system operating point data of alternative features describing these operating points and of intermediate and final study results. Automatic learning approaches often involve a long and tedious trial and error experimentation with alternative feature selection or fine tuning of the parameters of the automatic learning algorithms used. The use of the MLPSE toolkit and of the data warehouse approach facilitates the process and permits flexibility and tracking of the process by storing study decisions in the data warehouse. An extract of such experimental study has been included in the paper, which concerns steady state security assessment for the interconnected power system of the island of Crete in Greece. Many of the features of the reported study can be applied to other studies involving high volumes of system and analysis data, as the tools developed and used can be easily adapted due to the modular toolkit approach used.

Acknowledgements

Special thanks are due to the group of engineers and in particular to A. Coronidis of the Transmission and Distribution Studies Division of PPC Greece. Financial support by PPC Greece and GSRT under the YPER Programme is also acknowledged.

References

Albuyeh F., Bose A. and Heath B., "Reactive power considerations in automatic contingency selection", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 1 January 1982, p. 107

Chan K.W., Dunn R.W., Daniels A.R. and Edwards A.R., On-line dynamic security contingency screening using artificial neural networks, IEE Proc. Generation Transmission and Distribution, November 2000, Vol. 147, Issue 6, pp.367-372

Ejebe G.C., Wollenberg B.F., "Automatic Contingency Selection", IEEE Trans. on Power Apparatus and Systems, Vol.PAS-98, No.1 Jan/Feb 1979 p.97

Grady M., PCFLO Version 2.4 User Manual, ECE Dept, The University of Texas at Austin, January 15, 1995. http://www.ece.utexas.edu/~grady/

Hatziargyriou N.D., Contaxis G.C., Sideris N.C., "A decision tree method for on-line steady state security assessment", IEEE Transactions on Power Systems, Vol. 9, No 2, p. 1052-1061, May 1994.

Kimball, R., The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley & Sons, 1996.

Kroenke, D.M., Database Processing, Prentice Hall, 2000.

Mijuskovic N.A., Stojnic D., Probabilistic real-time contingency ranking method, International Journal of Electrical Power & Energy Systems 22 (7) (2000) pp. 531-535.

Murthy S.K., Automatic construction of decision trees from data: A multi-disciplinary survey, Data Mining and Knowledge Discovery, 2: (4) pp. 345-389, 12/ 1998.

Srivastava L., Singh S.N. and Sharma J., Knowledge-based neural network for voltage contingency selection and ranking IEE Proc. Generation Transmission and Distribution, November 1999, Vol. 146, Issue 6, pp.649-656

Wehenkel L.A., "Automatic Learning Techniques in Power Systems", Kluwer Academic Publ., 1998

Weiss, S., Indurkhya N., Predictive data mining, Morgan Kaufmann, 1998.