Data Mining Classifiers for Static Security Evaluation in Power System

I. S. Saeh[1], M. W. Mustafa[2]

[1], [2] Electrical Engineering, University Technologi Malaysia
Johor, Johor Bahru, Malaysia

Abstract

This paper addresses the application of data mining approach on Static Security Evaluation (SSE) of deregulated power system. The process of building binary class classifiers is divided into two components: (i) comparison the methods, and (ii) selection of the best classifier. Preliminary results of using eleven algorithms of Decision Tree's classifiers (DTC) for SSA are presented. A comprehensive comparison of the proposed classifiers for the purpose of SSA classification is discussed. A set of training cases generated on the IEEE 30 and 300-bus system were used to train and test the classifiers that discriminates the system security. The results show that DT's classifiers are capable of system security classification. Finally, empirical results indicate that C4.5 tree can be used to design a SSAC that is lightweight, efficient and effective for real time classification.

Keywords: *Decision Tree classifiers, C4.5, Static Security Evaluation, Data Mining.*

1. Introduction

Recent shift in electric energy sector from vertically integrated to deregulation, with the intention to improve operation and efficiency, has brought along a number of issues regarding the security of large systems. The occurrence of contingencies may cause dramatic interruptions of the power supply and so considerable economic damages. Such difficulties motivate the research efforts that aim to identify whether a power system is insecure and to promptly intervene.

Security evaluation, which is defined as the ability of the power system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system load, is one of the important issues especially in the deregulated environment [1]. When a contingency causes the violation of operating limits, the system is unsafe. One of the conventional methods in security evaluation is a deterministic criterion, which considers contingency cases, such as sudden removals of a power generator or the loss of a transmission line. Such an approach is time consuming for operating decisions due to a large number of contingency cases to be studied. This paper tries to address this situation by treating power system security evaluation as a pattern classification problem.

A number of approaches using artificial neural networks (ANN), such as Back-Propagation [2] and Self Organizing Map [3], have been proposed for security evaluation in power systems over the past decade. The key problem of ANN is the determination of an optimal ANN architecture, which is decided by trial and error in the selection of number of neurons in the hidden layer. ANFIS has been implemented for static security evaluation[4, 5].The Bayes classifier may provide an alternate way to resolve this problem. Although the Bayes classifier has been applied to various areas, such as signal processing; it has seldom been used in power system applications.

Recently, support vector machines(SVM), based on statistical learning theory have been used in the different areas of machine learning [6]. Important consideration in applying SVM to power system security evaluation is the proper selection of training feature set, characterizing the behavior of the power system. Many feature selection algorithms are available in the literature such as fisher discrimination analysis and entropy maximization [7]. The main problem with the existing feature algorithms is that it works well with linearly separable classes, but not well established on non-linearly separable classes [8]. Nowadays, PR techniques are gaining more importance in many power system problems. In PR approach, the main bulk of simulation is done off-line to generate sufficient data for training set[9]. The most important task in the application of PR system is the selection of primary variables, forming the components of pattern vector. This vector is evaluated at many representative operating points to generate a good training set. Each operating condition or state is termed as a pattern.

2. Power System Security

As the increase in electric power demand outpaces the installation of new transmission and generation facilities, power systems are forced to operate with narrower margins of security. Security is defined as the capability of guaranteeing the continuous operation of a power system under normal operation even following some significant perturbations [10].

The standard approaches to the security evaluation of electrical power systems are usually classified as either static or dynamic. More specifically, the static security analysis (SSA) is the post-contingent steady state evaluation of the power system by neglecting the transient behavior and any other variations that may depend on the load-generation conditions. On the contrary, if one accounts for the transition from the pre-contingent state to the post-contingent one, in the literature it is usually referred to as dynamic security analysis (DSA) which is discussed in the literature [11-13].

One of the major objectives in power system is security analysis. As security is a major, if not ultimate, goal of power system operation and control, a fast and reliable security evaluation is necessary.

2.1 Static Security Evaluation Indices Selection

Many indices have been proposed in the literature as criteria for static security evaluation [14-18], 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 [14] 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 [15] and [16], these involve overloaded lines, or bus voltages that deviate from the normal operation limits.

Under normal operating conditions, for following constrains called as security constrains must be satisfied:

$$\sum_{i=1}^{N_g} P_{Gi} = P_D + P_{loss} \qquad P_{Gi}^{\min} \le P_{Gi} \le P_{Gi}^{\max}$$

$$i = 1,2,....N_g \qquad (1)$$

$$\left|V_k^{\min}\right| \le \left|V_k\right| \le \left|V_k^{\max}\right| \quad k = 1,2,...N_b \qquad (2)$$

$$S_{km} \le S_{km}^{\max} \qquad branches, k-m$$

Where P_{Gi} represents real power generation at bus i, P_D is the total system demand; P_{loss} is the real power loss in the transmission network; $|V_k|$ is the voltage magnitude at bus k; S_{km} ; represents the MVA flow in branch $k-m$; N_g and N_b ; and being the number of generators and buses respectively.

In static security evaluation process, the status of the power system is evaluated for various probable contingencies by solving non-linear load flow equations. The contingencies may include outage of a transmission line or a transformer or a generating unit.

The load flow solver is simulated for various disturbances and the security constraints are evaluated. The system operating state is labeled as 'Static Secure' (SS-Binary 1) if all the constraints (i), and (ii) are satisfied for a specified contingency. If anyone constraint violation is identified following a contingency, the system state is labeled as 'Static Insecure' (SI-Binary 0).

3. DATA MINING

In general, most data mining techniques evaluate knowledge through the database. In recent years, it is much more difficult to interpret complicated data as the size of database becomes larger. As a result, it is necessary to develop a systematic computer-aided method to deal with the complexity of data [18]. Data mining can be divided into two parts, classification and clustering techniques and its effectiveness and powerfully of reducing the complexity of the data, made it to be used in many areas such as medical, engineering [19, 20]. As part of data mining technique, decision tree has the capability to analyze large databases, normally related to power system security evaluation. In this work, Decision Tree is used to assess the power system security. DT sometimes combined to other techniques [21].

3.1 Decision Tree Classifiers (DTC)

The Decision Tree (DT) is a tree, structured upside down, built on the basis of a knowledge base (KB) consisting of

a large number of operating points (OPs), covering all possible states of the under study power system in order to ensure its representatives [22-25]. The knowledge base is defined as [24], these attributes are the pre-disturbance steady-state variables and characterize each operating point.

The KB is divided in a learning set (LS) used for deriving the classifier structures and a test set (TS) used to evaluate the performance of these structures on new, unobserved OPs. The construction of a DT starts at the root node with the whole LS of pre-classified 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 selected test is applied to the LS of the node splitting it into two exclusive subsets, corresponding to the two successor nodes.

Every subset (node) is characterized by its security index (SI), defined as the percentage of secure OPs belonging to this subset. The optimal splitting rule is applied recursively to build the corresponding sub-trees. In order to detect if one node is terminal, i.e., sufficiently class pure, the stop splitting rule is used, which checks whether the entropy of the node is lower than a preset minimum value. If it is, the node corresponds to a sufficiently pure subset (states belong to the same class) and is declared a leaf; otherwise, a test is sought to further split the node. If the node cannot be further split in a statistically significant way, it is termed a dead-end, carrying the two class probabilities estimated on the basis of the corresponding Ops subset. A more detailed technical description of the approach followed is described in [25].

DTs are evaluated using the Testing Set (T.S.). The most important evaluator of the D.T. reliability and performance is the rate of successful classifications, defined as the ratio of successfully classified OPs to the number of OPs tested. The decision tree results and the number of the nodes depend on the accuracy given from the user.

Initially high accuracy parameters are given in order to obtain a large and accurate tree. Afterwards the tree size is gradually reduced in order to get a tree with more practical rules, because usually the initial tree is quite large with many non-important nodes, which have very small percent of OPs. This structure is not suitable for fast security evaluation, taking into account that for corrective action it is needed to cross the tree backwards.

With this procedure finally it is obtained a decision tree, which in most cases has a little worst accuracy but has quite less nodes and gives more practical and clear rules for the security of the system.

Generally the features that may be applied to describe a power system state are:

1) The voltage magnitude of each bus load.
2) The active and reactive power flow of all the lines.

3.2 Decision Tree Classifier (DTC) for Static Security Evaluation

The (DT) methodology [22] is a non-parametric learning technique able to produce classifiers about a given problem in order to deduce information for new, unobserved cases. The DT has the hierarchical form of a tree structured upside down and is built on the basis of a Learning Set (LS).The LS comprises a number of pre-classified operating states or points (OPs) defined by a list of candidate attributes. These attributes characterize the pre-disturbance OPs. A systematic treatment of the DT methodology is provided in [26].

The construction of a DT starts at the root node with the whole LS of pre-classified OPs. These OPs are analyzed in order to select the test T which splits them "optimally" into a number of most "purified subsets. For the sake of simplicity, a Z class partition is considered in the following analysis. The test T is defined as [22].

4. RESULTS AND ANALYSIS

For steady state security, the voltage magnitude (V_k) of each bus and the thermal power (S) of all the lines are the limitations. These limitations are:

$1.06 > V_k > 0.94$ and $S < S_{\max}$.

It is to be noted that the limitations for both voltage magnitude is 1.06-0.94 and line thermal power is 100 maximum. We assume that load at bus3 needs 20 MW, load at bus 4 needs 25 MW and load at bus 5 needs 15 MW from generators 1, 2 and 3 respectively. The transactions will be as follow:

TABLE 1: load and generators transactions

Transaction No.	*Gen. No.*	*Load No.*	*MW*
1	1	3	20
2	2	3	20
3	3	3	20
4	1	4	25
5	2	4	25
6	3	4	25
7	1	5	15
8	2	5	15
9	3	5	15

3.1 Transactions Implementation

We implement the transactions individually to see the impact of the transactions on the system security. Next table shows the power flow result after every transaction. It is clearly shown that all voltages are within their limits and power lines flow are not exceed. And as result, the system status is secure.

TABLE 2: power flow transactions and power system status

Trans. no.		T1	T2	T3	T4	T5	T6	T7	T8	T9
Voltage Magnitude	V1	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
	V2	1.045	1.045	1.035	1.045	1.045	1.045	1.045	1.045	1.045
	V3	1.03	1.03	1.01	1.03	1.03	1.03	1.03	1.03	1.03
	V4	1.019	1.019	1	1.017	1.017	1.017	1.018	1.018	1.018
	V5	0.99	0.99	0.977	0.989	0.989	0.989	0.985	0.985	0.985
Thermal Power	L12	79.535	27.758	64.842	77.514	27.858	61.32	72.527	59.706	62.88
	L13	34.774	56.055	60.106	31.792	56.265	59.657	68.797	69.115	67.516
	L23	61.099	60.661	27.55	61.731	62.253	23.12	26.769	24.774	22.099
	L24	19.873	18.548	16.45	17.076	18.599	11.871	13.16	13.894	10.854
	L25	26.168	25.21	23.494	27.208	28.351	69.958	21.801	22.387	19.636
	L34	43.826	44.763	47.07	63.189	61.681	22.906	53.493	52.619	57.238
	L45	14.029	14.191	13.809	13.451	13.295	14.423	18.917	18.676	19.96
System status		secure	secure	secure	secure	secure	secure	secure	secure	secure

3.2 Decision Tree's Comparison

For the same train and test data used, eleven various algorithms are used for a comparison in term of accuracy, computational time and root mean square error (RMSE) and tabulated in next table. Learning algorithms of the trees are presented in [27].

Table (1): Performance of various Decision Tree algorithms in Train and Test Set

(a) 30 Bus system		***AT Tree***	***BF Tree***	***Stump Tree***	***J 48 Tree***	***J 48 graft***	***LMT Tree***	***NB Tree***	***C 4.5 Tree***	***R Tree***	***Rep Tree***	***Simple Cart***
Train	***Accuracy***	85.3	86	77.3	92	91	89.7	77.8	95.7	93	88.2	70.5
	Time(S)	0.02	0.01	0.01	0.03	0.023	0.009	0.01	0	0.009	0.004	0.01
	RMSE	0.065	0.0622	0.065	0.064	0.055	0.075	0	0	0.045	0.034	0.038
Test	***Accuracy***	86.5	90	80.6	91.5	89	90	79.5	97	92	87	76
	Time(S)	0.01	0.007	0.01	0.0093	0.055	0.12	0.006	0.001	0.012	0.001	0.01
	RMSE	0.025	0.034	0.0371	0.0510	0.0355	0.047	0.001	0.001	0.0144	0.009	0.008
(b) 300 Bus system												
Train	***Accuracy***	88	87.5	80	95	93.4	89.7	80	97.5	92	87	74
	Time(S)	0.04	0.007	0.01	0.055	0.026	0.009	0.015	0	0.01	0.005	0.007
	RMSE	0.013	0.006	0.045	0.055	0.008	0.009	0.005	0.008	0.01	0.023	0.003
Test	***Accuracy***	88.3	92	83	92	89	94	83	97	94	88	79
	Time(S)	0.017	0.009	0.009	0.01	0.07	0.17	0.005	0	0.012	0.001	0.01
	RMSE	0.026	0.038	0.038	0.055	0.037	0.049	0.081	0.001	0.0119	0.01	0.0138

The table illustrates the accuracy, computation time and RMSE for both train and test mode in two different system sizes. From the previous table it can be seen strongly that in both small and large size system, C4.5 got best accuracy (95.7) with minimum computation time (0) second. These results can be also observed for the recall mode where C4.5 got 97% accuracy and 0.001 second for computation time. And finally, for RMSE in the train mode the Random Tree and C 4.5 was 0.001 and 0.0816 respectively. In the recall mode the J48 graft and C 4.5 got 0.473 and 0.5185 respectively.

5. Conclusion

This work has presented the results and discussions. The study of implementation data mining techniques on various test system involved suitability of using eleven DT's for SSE classification. From the studies, it is observed that DT promises alternative and successful method of evaluation for the large power system as compared to the conventional method. All these DT's methods can successfully be applied to assess SSA of deregulated power systems in real time. By considering the computation time and accuracy of the networks, it can be concluded that C4.5 is well suited for online SSE of deregulated power systems. In general, this classifier technique holds promise as a fast online classifier.

Acknowledgements

The authors would like to thank Ministry of Higher Education, Malaysia (MOHE) for providing financial support under E-science grant. Authors also like to thank Department of Electrical Engineering, Universiti Teknologi Malaysia (UTM) for providing necessary facilities and resources for this research work.

6. References

[1] J. Flory, "Electricity Transactions in an Open Access Market," Power Engineering Review, IEEE, vol. 16, p. 15, 1996.
[2] I. S. Saeh and A. Khairuddin, "Static security assessment using artificial neural network," in Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International, 2008, pp. 1172-1178.
[3] W. Rosehart, et al., "Optimal power flow incorporating voltage collapse constraints," in Power Engineering Society Summer Meeting, 1999. IEEE, 1999, pp. 820-825 vol.2.
[4] I. Saeh and A. Khairuddin, "Implementation of Artificial Intelligence Techniques for Steady State Security Assessment in Pool Market," 2009.
[5] I. Saeh and A. Khairuddin, "Anfis and ANN comparison for static security assessment," 2008.
[6] M. Mohammadi and G. Gharehpetian, "Power system on-line static security assessment by using multi-class support vector machines," Journal of Applied Sciences, vol. 8, pp. 2226-2233, 2008.
[7] C. A. Jensen, et al., "Power system security assessment using neural networks: feature selection using Fisher discrimination," Power Systems, IEEE Transactions on, vol. 16, pp. 757-763, 2001.
[8] J. Sa Da Costa and N. Munro, "Pattern recognition in power-system security," International Journal of Electrical Power & Energy Systems, vol. 6, pp. 31-36, 1984.
[9] C. K. Pang, et al., "Security Evaluation in Power Systems Using Pattern Recognition," Power Apparatus and Systems, IEEE Transactions on, vol. PAS-93, pp. 969-976, 1974.
[10] K. Morison, et al., "Power system security assessment," Power and Energy Magazine, IEEE, vol. 2, pp. 30-39, 2004.
[11] K. Sun, et al., "An online dynamic security assessment scheme using phasor measurements and decision trees," Power Systems, IEEE Transactions on, vol. 22, pp. 1935-1943, 2007.
[12] E. Voumvoulakis, et al., "Application of machine learning on power system dynamic security assessment," 2007, pp. 1-6.
[13] I. Kamwa, et al., "Development of rule-based classifiers for rapid stability assessment of wide-area post-disturbance records," Power Systems, IEEE Transactions on, vol. 24, pp. 258-270, 2009.
[14] F. Albuyeh, et al., "Reactive power considerations in automatic contingency selection," Power Apparatus and Systems, IEEE Transactions on, pp. 107-112, 1982.
[15] G. Ejebe and B. Wollenberg, "Automatic contingency selection," Power Apparatus and Systems, IEEE Transactions on, pp. 97-109, 1979.
[16] L. A. Wehenkel, Automatic learning techniques in power systems: Kluwer Academic Publishers, 1998.
[17] M. Marsadek, et al., "Risk based static security assessment in a practical interconnected power system," 2008, pp. 1619-1622.
[18] H. Mori, "State-of-the-art overview on data mining in power systems," 2006, pp. 33-34.
[19] F. Camara, et al., "Privacy Preserving RFE-SVM for Distributed Gene Selection," 2012.
[20] S. Jun, "A Clustering Method of Highly Dimensional Patent Data Using Bayesian Approach," 2012.
[21] A. A. Singh and M. Thingujam, "Fuzzy ID3 Decision Tree Approach for Network Reliability Estimation," 2012.
[22] N. Hatziargyriou, et al., "Decision trees for fast security assessment of autonomous power systems with a

large penetration from renewables," Energy Conversion, IEEE Transactions on, vol. 10, pp. 315-325, 1995.

[23] P. Georgilakis and N. Hatziargyriou, "On the application of artificial intelligence techniques to the quality improvement of industrial processes," Methods and Applications of Artificial Intelligence, pp. 745-745, 2002.

[24] E. Voumvoulakis, et al., "Decision trees for dynamic security assessment and load shedding scheme," 2006, p. 7 pp.

[25] L. Wehenkel and M. Pavella, "Decision trees and transient stability of electric power systems," Automatica, vol. 27, pp. 115-134, 1991.

[26] L. Wehenkel, et al., "An artificial intelligence framework for online transient stability assessment of power systems," Power Systems, IEEE Transactions on, vol. 4, pp. 789-800, 1989.

[27] I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann, 2005.

I. S.Saeh received his Bsc. Eng. degree (1997). Msc (2009) from university Technology Malaysia. His research interests include deregulated power system security and AI Techniques. Since2009, He is PhD student at University Technology Malaysia. His current research is Deregulated power system Security using AI techniques.

Mohd W.Mustafa received his Bsc. Eng. degree (1988), Msc, in (1993) and PhD (1997) from university Strathclyde, Glasgow. His research interests include power system stability, deregulated power system distribution automation, FACTS and power quality. He is currently An Associate Professor and Deputy Dean of Faculty if Electrical Engineering, University Technology Malaysia.