Critical Voltage Monitoring Using Sensitivity and Optimal Information Machine Learning

Jovan Ilić, *Member, IEEE*, Le XIE,*Member, IEEE*, and Marija D. Ilić, *Fellow, IEEE*

***Abstract*—This paper is motivated by the basic need to develop methods for on-line detection of abnormal conditions in large electric power systems. In order to implement truly effective near-automated tools for this purpose, it is necessary to overcome several problems such as: (1) excessive computational complexity; (b) unacceptable approximations; and, (3) dependence on full state measurements. In an attempt to overcome these major roadblocks, we combine tools capable of producing accurate results over broad ranges of conditions, such as off-line data mining and machine learning, with the approximate, well-understood deterministic methods, such as sensitivity-based methods. The resulting approach indirectly overcomes the dependence on full state measurements; the actual choice of the most relevant measurements becomes a result of such a combined approach. The proposed approach is illustrated on an example of detecting a given voltage threshold violation.**

***Index Terms*—Sensitivity analysis, information theory, decision trees, electrical distance, event monitoring, abnormal conditions.**

I. Introduction

IN the electric power systems, there is frequently a need to monitor an unaccessible or unmeasurable quantity. On a large scale, this would constitute state estimation of the entire system. On a smaller scale, it could be just monitoring a troublesome voltage, angle, or even a parameter. In this paper, we illustrate a possible approach to determining when voltage at a particular bus may be falling below a pre-specified value. We start by using two qualitatively different methods for assessing accuracy, benefits and shortcomings of each. The first method uses deterministic sensitivities, and the second method is a probabilistic supervised machine learning (ML) method. We point out that both methods have strong and weak points and are capable of monitoring tasks with certain efficiency and accuracy. We ultimately end up with combining the two methods into a novel method which draws on strengths of each method

This work was supported in part by the U.S. National Science Foundation under award CNS-0428404, and, in part, by the U.S. Department of Energy, National Energy Technology Laboratory, under Research and Development Solutions, LLC contract number DE-AM26-04NT41817.305.01.21.002.

Jovan Ilić is a Scientific Specialist in the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 (e--mail: jilic@ece.cmu.edu).

Le Xie is a PhD student in the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 (e-mail: lx@ece.cmu.edu).

Marija D. Ilić is a Professor in the Department of Electrical and Computer Engineering and Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, 15213 (e--mail: milic@andrew.cmu.edu).

II. Sensitivity and Supervised Machine Learning

Sensitivity calculations are widely used in the electric power systems for different purposes. Here we use static voltage sensitivity which can be calculated from a part of systems power flow Jacobian to indirectly estimate voltage of interest. While there are different types of machine learning, we will focus here on a type that is based on entropy functions and theory of information since it is related to sensitivity and electrical distance concepts as reported in [1]. First, we will introduce voltage sensitivity and different approximations and then the ML method used for comparison followed by an example.

A. Voltage Sensitivity

A sensitivity between almost any two power system values can be defined in some way. The usual power systems description describes it as a nonlinear physical system with nonlinear first order sensitivities. Some of the sensitivities are part of the Jacobian used to solve the power flow while others must be calculated in sometimes ingenious ways. Due to computational complexity, linear approximations or P/Q decoupling assumption are used. For example, voltage sensitivity $\left(\frac{\partial V_i}{\partial V_j}\right)$ can be defined as:

$$\frac{\partial V_i}{\partial V_j} = \frac{\frac{\partial V_i}{\partial Q_j}}{\frac{\partial V_j}{\partial Q_j}} \tag{1.1}$$

and

$$\Delta V_i = \sum_{\substack{i=1 \\ i \neq j}}^{N_{bl}} \frac{\partial V_i}{\partial V_j} \times \Delta V_j \tag{1.2}$$

where N_{bl} is the number of load buses in the system. For the purpose of estimating a bus voltage:

$$V_i^{k+1} = V_i^k + \Delta V_i \tag{1.3}$$

with V_i^0 obtained from AC power flow. Since $\partial Q/\partial V$ is, under small angle approximation assumption, close to the susceptance matrix B, it is customary to use matrix B to calculate the sensitivity matrix. This means that there is a linear dependency between system bus voltages regardless of the operating point. If higher accuracy is needed, $\partial Q/\partial V$ or even the entire Jacobian must be calculated. Sensitivity matrices are either direct byproduct of the power flow or the power flow is required to find the operating point, that is completely solve the system equations making the sensitivity calculations redundant for many purposes. In this work, the system is assumed to be P/Q decoupled so that only $\partial Q/\partial V$ part of the Jacobian is used. Mathematically, this means that if the Jacobian is given by:

$$J = \begin{bmatrix} \partial P/\partial \theta & \partial P/\partial V \\ \partial Q/\partial \theta & \partial Q/\partial V \end{bmatrix} \tag{1.4}$$

and the inverse of the Jacobian:

$$J^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{1.5}$$

Then for $C \approx 0$ and using Schur complements:

$$D = \left(\frac{\partial Q}{\partial V}\right)^{-1} - \frac{\partial Q}{\partial \theta} \times \left(\frac{\partial P}{\partial \theta}\right)^{-1} \times \frac{\partial P}{\partial V} \tag{1.6}$$

which is approximated by $\left(\partial Q/\partial V\right)^{-1}$. It is clear that even if the P/Q decoupling assumption is valid, $C \approx 0$, there is still an error term introduced with this approximation which will be
illustrated with an example. Another, more serious, drawback of using sensitivity for monitoring purposes is that due to its number crunching requirements it cannot be used in real time. There is no doubt that classical theories can provide the most accurate results but they are often not suitable for real time, preventive tasks.

B. *Supervised Machine Learning*

A range of AI methods were tried in power systems including ANN, GA, ML, and expert systems. Most of these methods produce results which are not self explanatory and impossible to back track the solution logic. Also, most of these methods make no attempts to mathematically explain AI results using classical theory. Due to these reasons, we chose an ML algorithm for event detection which produces easily understandable results, has well developed mathematical background, and has some interesting features applicable to power systems. The method under consideration is ID3 [2]. Originally, ID3 was designed using categorical attributes only. Since its introduction, it was expanded to use continuous attributes and multiple classes which made it usable for engineering purposes. The ID3 algorithm itself is based on Top Down Induction of Decision Trees (TDIDT). A large number of off line simulations or on line collected data is used for estimating probability distributions used for entropy estimation. Entropy of a set with n elements, n_c different classes (values), and n_i number of elements belonging to each class is calculated using Shannon's entropy function as:

$$E\left(X_a\right) = -\sum_{i=1}^{n_c} \frac{n_i}{n} \log\left(\frac{n_i}{n}\right) \tag{1.7}$$

If a set A is split into two subset using query q, the entropy of classification, $E\left(X_a, q\right)$, is defined as:

$$E\left(X_a, q\right) = -\sum_{r \in \{yes, no\}} \frac{n_r}{n} \sum_{i=1}^{n_r} \frac{n_{r,i}}{n_r} \log\left(\frac{n_{r,i}}{n_r}\right) \tag{1.8}$$

where n is the number of data samples in the sample set, n_r is the number of samples with query result r and $n_{r,i}$ is the number of samples with query result r which are members of class i. The query is a criterion used to split the set into two subsets using either categorical or continuous attribute. Once the subsets are found, the entropy of each subset is calculated. The best partitioning of the set X_a is the one with the maximum information gain obtained by finding the subsets with minimum entropy (1.8). If a set X_a is recursively split in this manner, the process can be represented by a decision tree with best splitting decisions at each node and leaves belonging to a single class. The splitting criteria can be of different type, they can be equality between a categorical attribute and a constant, inequality between continuous variable and constant threshold, or equality and inequality between a constant and a function of attributes as used in oblique decision trees [3]. In this work, we use node decisions based on continuous variables compared to constant thresholds with ability to use categorical attributes as well. This type of decision making results in stair wise splitting hyperplanes. If starting with a well defined, non-conflicting training set, stair wise splitting surfaces is the only approximation contributing to significant classification errors. The basic characteristics of ID3 are that it is 100% accurate on the training set, it can be used to extract the variables carrying the most information related to observed events (classes), it can estimate the variable and its margin leading to a different

class, and once trained it can be used in real time. The deficiencies of the ID3 algorithm are that it can over fit the data, separation hyperplanes are stair wise, and as most of the ML algorithms suffers from "the curse of dimensionality". There are quite a few variations of decision trees to address these issues as well as different types of learning mechanisms. ID3 is used in this paper due to its mathematical basis in theory of information that can be mathematically related to classical power systems theory to some extent [1], [4].

III. EXAMPLE

The small network shown in Fig. (1). is used to demonstrate use of approximate sensitivities and ID3 approach for the purpose of detecting a voltage threshold at a particular node. Voltage drop $V_5 \leq 0.96$ needs to be detected using some indirect measurements. This will be done using both voltage sensitivity coefficients and ID3. Impedance of all of the branches is set to $0.1 + j0.1$. Loads L_2, L_3, and L_5 are uniformly sampled in the range between 50MW and 100MW each and the ID3 training set is formed by running power flow for each load sample. In addition to load levels, load bus complex voltages, and P/Q generation levels are collected.

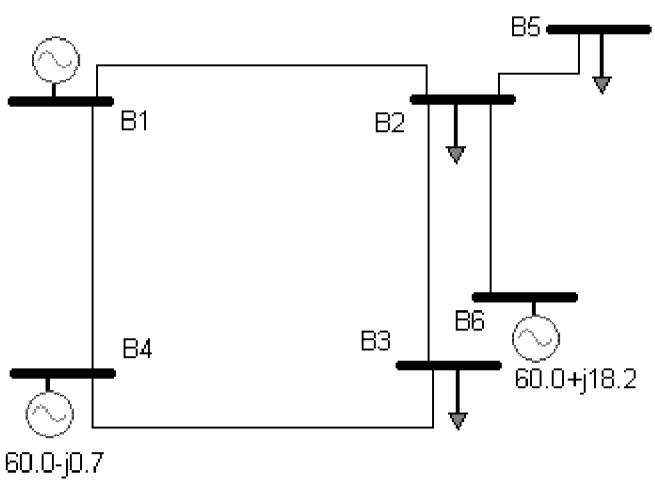

Fig. 1. Six Bus network

A. Using Sensitivity to Detect Voltage Drop

The sensitivity approach simply calculates voltage V_5 indirectly from sensitivity coefficient and corresponding voltage increments using equations (1.2) and (1.3). In the following, sensitivity matrix $\partial V_i / \partial V_j$ is calculated from $\partial Q / \partial V$ matrix. This method assumes P/Q decoupled system. Figure (2) shows voltage V_5 calculated using Newton-Raphson power flow and using sensitivities as described above.

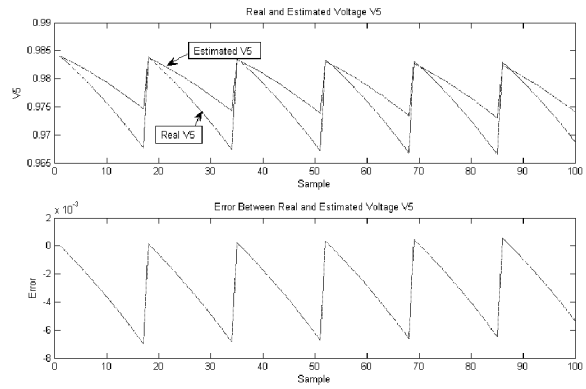

Fig. 2. Actual and Estimated V5

It is clear that there is a small difference between these two methods. Even more serious problem becomes obvious if the error between the real (using AC power flow) and estimated V_5 is plotted over a wider range of L_2, L_3, and L_5. Figure (3) shows that the error is integrated over time becoming unbounded. This should also be obvious from the equations (1.3) and (1.6). Rather large spikes in Fig. (3) are caused by jumps in L_5, L_3, and L_5 sampling space causing ΔV to be large.

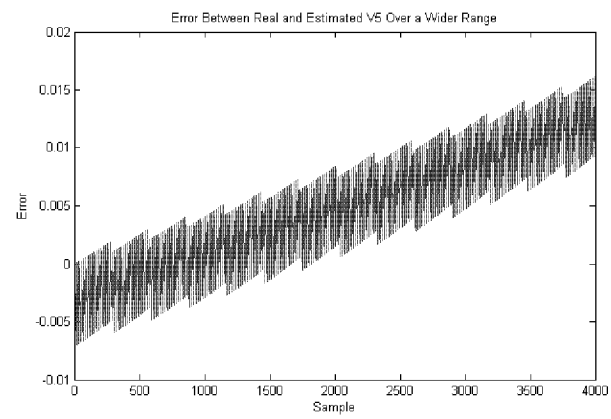

Fig. 3. V5 Over a Wide Sampling Range

B. Using ID3 to Detect Voltage Drop

In the first run, ID3 training set contained all the voltages including the voltage to be detected, V_5. As expected, this is a trivial case and the result is trivial decision tree with only one node, $V_5 \leq 0.96$, as shown in Fig. (4). The significance of this trivial case is that ID3 is capable of identifying the most significant (closest) variable and use it correctly.

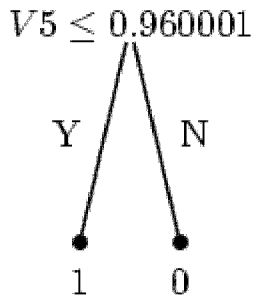

Fig. 4. Trivial Decision Tree

If V_5 is omitted from the training set, the resulting decision tree is shown in Fig. (5). The closest voltage to V_5 is V_2 and this is still in agreement with electrical distance. ID3 is consistent with electrical distance theory but produces more complex decision trees to identify system state. It is important to note that if the sensitivity were really linear, there should be no need for training data for ID3 since the needed probability distributions would be readily available from sensitivity relationships. However, if the sensitivity were truly linear or a known nonlinear function, there would be no need for machine learning or estimation. If V_5 threshold to be detected is lowered, ID3 introduces next closest voltages in its decision process since the region is becoming more nonlinear.

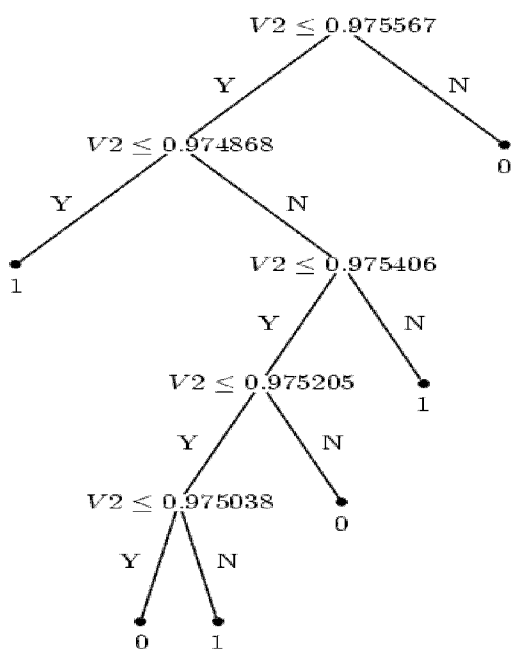

Fig. 5. Decision Tree Detecting V5 < 0.96

The machine learning approach to this type of problems has a number of advantages. First of all, supervised learning is able to learn the nonlinearities of the variable that it is trying to monitor in an optimal way. If properly trained, decision trees can properly generalize unseen cases from the seen ones. Different decision tree variations can deal with different issues such as over fitting, stair wise separation hyperplanes, etc. depending on the problem characteristics. Decision trees can also discover dependencies that are not obvious from the physical description of the system. In this example, if the ID3 training set contained all of the system variables, the resulting decision tree is more accurate and smaller as it can be seen in Fig. (6).

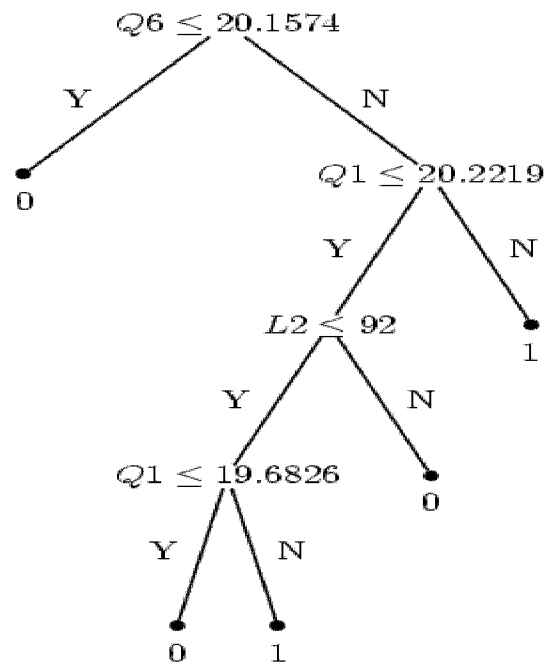

Fig. 5. Six Bus System Decision Tree with Expanded Training Set

Maybe the best incentive to use a decision tree in power systems is its fast execution in real time. A typical decision tree executing on a modern DSP processor can produce a decision within a millisecond which makes it suitable for preventive monitoring. One of the main arguments against using machine learning methods is "the curse of dimensionality". However, if it is known in advance what are the most important variables for ID3 training, the entire training set can be reduced to this set of training variables. For example, if an induced decision tree uses only two variables to make a decision, it is sufficient to use only these two variables in the training set. Because of the computational complexity of the induction algorithm, it is important to use a minimal set of decision variables. Unfortunately, this set is not known. There are two possible solutions to this problem. One is to induce a decision tree from a small size but complete training set. Then, increase the size of the training set but keep only variables that are found to be significant in the previous step. Another approach, specific to the electric power systems, is to use a couple of closest measurements, induce a decision tree and expand the radius of the used measurements until the set of the used

decision tree variables is the same between the last two inductions, in a concentric relaxation-like manner. At this point, it should be safe to conclude that the minimal set of the decision variables is identified and a larger training set with only these variables can be used.

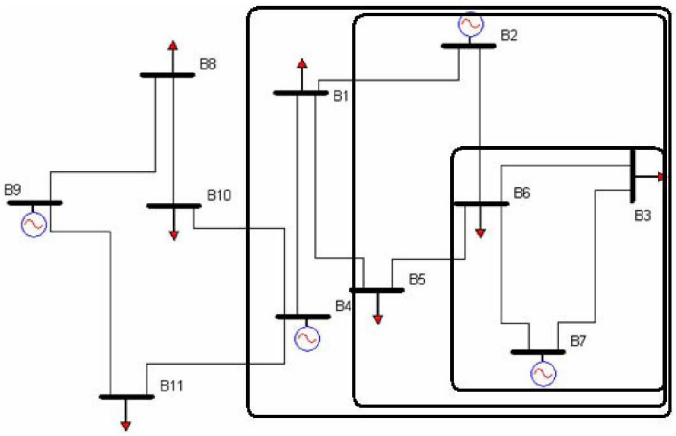

Fig. 6. Eleven Bus System

As an example, in Fig. (7), we want to detect voltage drop on bus B_3 using a minimal set of decision variables. First, V_6 and Q_{g7} are used and then Q_{g2} and V_5 are added. All of these variables appeared in the induced tree meaning that they are all important. Then the radius was expanded with Q_{g4} and V_1 which did not appear in the induced decision tree meaning that the minimal radius is reached. This works because of the local nature of voltage and it makes sense from the network topology point of view.

IV. Conclusions

Machine learning algorithm, ID3, based on theory of information and classical sensitivity method were used to design a voltage monitoring scheme. Results from both these methods were critically evaluated and compared. Sensitivity theory has advantage of being based on firm theoretical background although it suffers from using approximation techniques. Sensitivity calculations based on $\partial Q/\partial V$ also require power flow solution. Because of this, voltage monitoring based on sensitivity also cannot be used for preventive tasks due to high computational burden. Approaches with approximate sensitivity calculations are probably not a good choice for any kind of large variable or parameter change estimation or event detection due to the error integration. This type of methods could be used for estimating small variations around the operating point. ID3 is a supervised machine learning and as such all of the intensive simulations and computations are done off--line. The resulting decision tree is very fast to execute making it suitable for real-time operations. ID3 is capable of approximating nonlinear voltage dependency. ID3 is able to identify significant variables consistently with electrical distance measurements. As all of artificial intelligence methods, ID3 suffers from the "curse of dimensionality" but in the case of power systems, classical results can be used to reduce the size of the training set and to significantly reduce the computational burden. Note that there is still a significant simulation and data collecting process. Based on a series of simulations of a small electric power network, it can be concluded that both sensitivity and ID3 have advantages and shortcomings. Depending on the monitoring purposes, either a voltage sensitivity or ID3 approach can be used. The sensitivity theory can be used for estimating small variations in a monitored variable. The ID3 algorithm dose not suffer from any system approximations and can use the best event indicators from all of the information available. An induced decision tree is also very fast to execute in real time. These features make ID3 and other decision tree algorithms ideal for real time event detection. It is clear that there exists a close connection between the theory of information based machine learning and classical theory as it was pointed out by [1] but there is still lots of research to be done to make the full connection. For the time being, this paper represents an attempt to carefully combining the two approaches for developing an algorithmic approach to monitoring events in large scale electric power systems. More theoretical work is under way in our group.

Acknowledgments

The authors greatly appreciate the financial support which has made this work possible. The first and third author also acknowledge their continuing discussions and collaboration with Professor Louis Wehenkel from the University of Liege in Belgium.

References

[1] P. Lagonotte, J. Sabonnadière, J. Leost, and J. Paul, "Structural analysis of the electrical system: Application to secondary voltage control in France", *IEEE Trans. Power Syst.,* vol. 4, pp. 479-486, May 1989.

[2] J. R. Quinlan, "Induction of decision trees", *Machine Learning,* vol.1, pp. 81-106, 1987.

[3] S. K. Murthy, S. Kasif, and S. Salzberg, "A system for induction of oblique decision trees", *Journal of Artificial Intelligence Research, Aug. 1994.*

[4] P. Lagonotte, "Probabilistic approach of voltage control based on structural aspect of power systems", *Third International Conference on Probabilistic Methods Applied to Electric Power Systems,* pp. 208-213, July 1991.