Precise Measurements in Power Systems Conference, Washington, D.C., Oct. 25-29, 1993.

PARALLEL DECISION TREES FOR PREDICTING GROUPS OF UNSTABLE GENERATORS FROM SYNCHRONIZED PHASOR MEASUREMENTS

Stein Kretsinger *
Steven Rovnyak **
Donald Brown *
James Thorp **

* University of Virginia
** Cornell University

Key words: Classification, clustering, decision trees, parallel, power systems, prediction.

ABSTRACT – Large–scale electric power systems can benefit from having accurate predictions about the future state of the system. Knowing future behavior will assist in selecting control actions to stabilize the system after a fault or disturbance. Features of this prediction problem include: an extremely large number of possible classes, the need for fast training, the need for very fast real–time prediction, and the need for confidence in the prediction. We develop a pattern recognition approach, parallel decision trees, to predict the future groupings of unstable electric generators using measurements of the post–fault generator angles. The parallel decision trees decompose the large problem into a set of smaller problems that can be solved simultaneously. The outputs of these decision trees are synthesized into a grouping prediction using a nearest neighbor search through the set of known groupings. The Hamming distance to the nearest known grouping configuration turns out to be useful for assessing confidence in the prediction. Confidence in the prediction accuracy is very high when the Hamming distance is zero.

OVERVIEW

With new systems capable of making real–time, system–wide measurements on large–scale power systems, predicting future behavior has become an important area of research [1]. This prediction could be used in out–of–step relaying and in special protection systems (remedial action schemes) for example. A technique of parallel classifiers was developed, where the outputs of many decision trees are synthesized into a prediction about the system as a whole. By conditioning on the number of individual outputs that are in agreement with one of the known patterns of generator groupings, most cases can be predicted with a high degree of confidence.

INTRODUCTION

The supply of electric power is expected to be available on demand, and without interruption whenever possible. The system's ability to withstand severe disturbances without experiencing large–scale shut–downs is a fundamental concern for reliable operation. It is common in a disturbance situation for parts of the system to be disconnected by underfrequency relays and other actions. When a system goes unstable it tends to separate into groups of synchronized generators. Sometimes these groups are self sustaining islands, and other times there are wide–scale outages. This paper shows how to predict which generators are in which group, faster than real–time, from a sampling of the post–fault phasor measurements. This information could be used to select an appropriate real–time control option such as a controlled separation of the system subgroups, and/or emergency load tripping in the areas projected to go underfrequency.

Recent advances in satellite technology have allowed for the accurate measurement and transmission of generator phase angles to a central location in real–time with a small communication delay [2]. These phasor measurements reflect the current state of the system and could act as the inputs for a classifier. The classifier would predict the future state of the system based on the current phasor measurements. The most fundamental prediction is whether the system remains stable or goes unstable. This two–class problem was readily handled by standard decision tree techniques as described in [3]. The present work goes a step further by predicting the specific grouping of generators in the event of instability.

PROBLEM DESCRIPTION

The problem is to predict the future grouping pattern of generators given a short sampling of the post–fault generator phase angle measurements (phasor measurements). Prediction must be made quickly to allow for early control action. Utilities demand a high degree of accuracy in such predictions.

BACKGROUND

When a power system sustains a disturbance such as a lightning strike or a loss of generation, the state of the system is moved away from its previous operating point, and the system dynamics are usually altered due to circuit breaker operation. The state of the system, immediately after the fault (such as a short circuit to ground) is removed, together with the post–fault system configuration determine the eventual outcome. The governing differential equations can be solved numerically, although the idea of solving them in real–time is still a subject of on–going research. This motivates the investigation of a pattern recognition approach, where the post–fault state of the system is classified according to its subsequent behavior (e.g. stable vs. unstable).

The state vector for the power system model consists of all the generator angles along with their derivatives. Present technology allows one to measure the phase angles with precision of approximately 0.02 electrical degrees [4]. The idea in this paper is to investigate the performance of a classifier for predicting stability four seconds into the future based on a sampling of the post–fault phase angle measurements. In previous work we showed that a single tree can correctly predict stability or instability 95 percent of the time [3]. The parallel decision tree method, in addition, predicts the resulting subgroups of synchronized generators in the event of instability. This information could potentially be used to implement a controlled separation of the diverging areas, for example.

We used the New England 39 bus system to test the classification accuracy of these approaches. A feature common to both methodologies is that they are trained to handle disturbance locations anywhere in the network. Specifically, the training sets contain fault scenarios ranging from 1 to 10 cycles in duration on all the busses and transmission lines. Prediction accuracy is then tested for a set of randomly located faults on all the transmission lines. In the parallel decision tree methodology, to say that an evolving transient event was correctly classified means that the correct grouping of generators was predicted by the parallel decision trees.

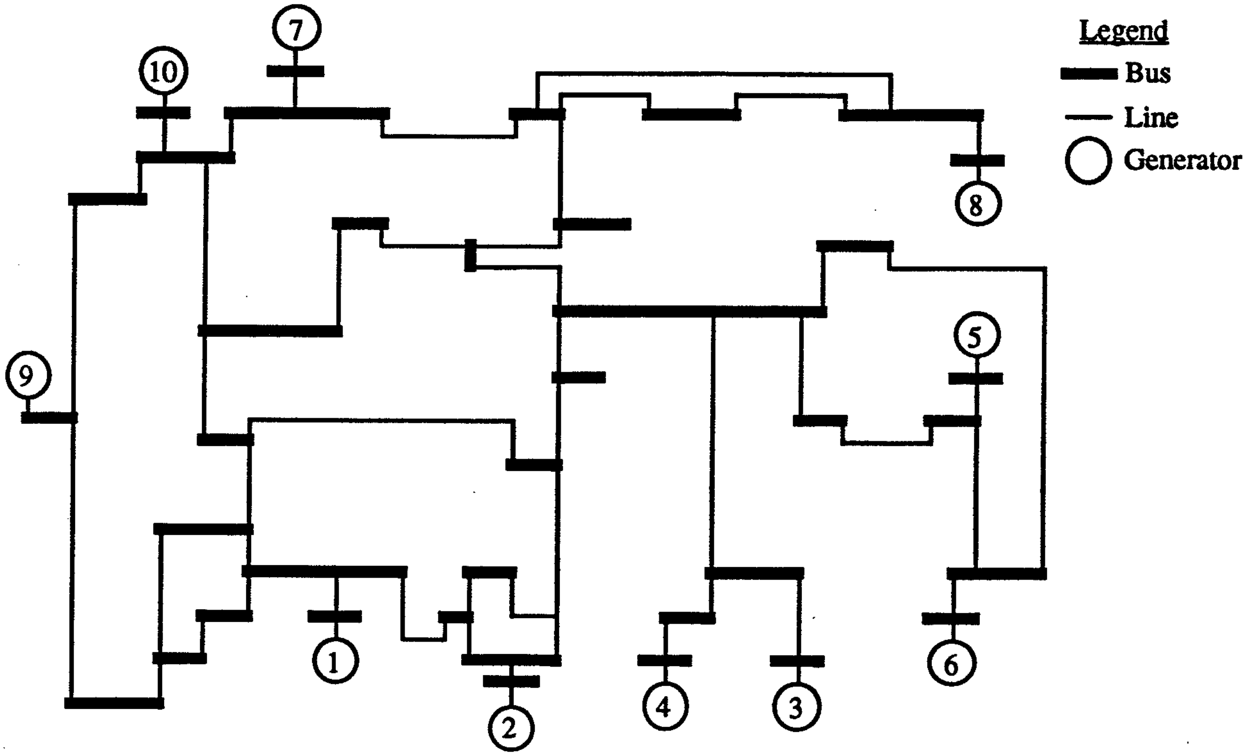

Figure 1: Schematic diagram for the New England 39 bus system.

In order to predict instability, simulated phasor measurements were taken from the post–fault swing curves immediately after fault–clearing time. Each phasor measurement was truncated to three decimal places in units of radians (0.001 radians corresponds to 0.057 degrees), and then two velocities and one acceleration were computed from the truncated generator angles. Denoting the three angle measurements from the i'th generator $\delta_i(0)$, $\delta_i(1)$, $\delta_i(2)$, we compute

$$
\begin{aligned}
v_i(0) &= 10 * [\delta_i(1) - \delta_i(0)] \\
v_i(1) &= 10 * [\delta_i(2) - \delta_i(1)] \\
a_i(0) &= 20 * [\delta_i(2) - 2 * \delta_i(1) + \delta_i(0)]
\end{aligned}
$$

Hence for each generator there are three angle measurements, two computed velocities and one computed acceleration. With 10 generators in the system, there are 60 predictor variables as inputs to the classifier.

CLASSIFIER DESIGN

The classifier output in this experiment was the grouping pattern of the electrical generators four seconds into the future. Representing all the possible grouping combinations turns out to be an underlying issue in this classification problem. For 10 generators, there are 115,975 different possibilities as explained below. Since there is no ordinal sense to the set of possible groupings, these different outcomes must be regarded as categorical variables. Fortunately, only a small subset of these classes are found in the data.

The grouping pattern can be represented as an N x N similarity matrix, which is an inherent feature of the parallel decision tree method. The element d_{ij} depends on the relationship between generators i, and j, namely if they are in the same group. The value of d_{ij} is either a 1 if the generators are in the same group or 0 if they are in different groups. The following is a similarity matrix found in the data:

$$\begin{matrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1
\end{matrix}$$

This matrix represents the grouping {1 2} {3 4 5 6} {7 8 10} {9}. The similarity matrix is always symmetric with 1's on the diagonal.

PROBLEM COMPLEXITY

With ten generators to group the total number of possible groupings is very large. For example

{1 2 3 4 7 8} {5 6} {9 10}
{1 2 3 4 7 10} {5 6} {8 9}
{1 2 7 9} {3} {4 5 8} {10}
{1 2 7 9} {3} {4 5 10} {8}
{1 2 3 4 5 6 8 9 10} {7}
{1 2 3 4 6 7 8 9 10} {5}

are just a few. Stirling numbers of the second kind give the number of ways to place n objects into k non–empty sets. The Stirling numbers are defined recursively by:

$$\left\{ {n \atop k} \right\} = k \left\{ {n-1 \atop k} \right\} + \left\{ {n-1 \atop k-1} \right\}$$

Since the number of groups formed by the generators can range from 1 to n the total number of possible groupings is the sum from 1 to n of the Stirling numbers:

$$\text{Total} = \sum_{k=1}^{n} \left\{ {n \atop k} \right\}$$

where n is the number of items to group. In our case n=10 and the total is 115,975.

PATTERN RECOGNITION APPROACHES

Since the power system follows a well understood set of differential equations most of the analysis in this field has relied on the numerical solution of these equations. Direct real–time integration is a potential solution for highly reduced–order models of the system, but execution speed increases as the size and complexity of the model increase. Integration requires a complete state description but the exact "post–fault" state of the system is not known in real time. A pattern recognition approach, on the other hand, can be trained to associate an incomplete description of the state with the prediction of future behavior.

Predicting the groups of generators is a difficult problem for any pattern recognition technique for the following reasons. The different possible groupings present an extremely large number of categorical values. The classifier must train quickly to accommodate the drifting operating point and operate very quickly to be effective in real time. A high degree of accuracy is also required for special protection systems (remedial action schemes) applications.

Decision trees (DT's) were selected from the list of established pattern recognition techniques because they fit the problem design constraints very well. Decision trees have had previous successful applications in the area of electric power systems [3, 5]. The CART software developed by Breiman [6] meets the criteria for fast training and very fast prediction. Another positive feature of decision trees is their interpretability and ease of implementation, so that relay engineers could easily interact with the tree–building process. In [3] we make the argument that a decision tree methodology can help to automate the programming and updating of adaptive relays.

PARALLEL CLASSIFICATION APPROACH

Forty–five binary decision trees were built to predict whether each pair of generators are in the same synchronized group. The outputs from the 45 trees were arranged in a similarity matrix, as described in the section on classifier design. The decision trees were built with the CART algorithm, withholding one third of the data for pruning. Substantial gains in performance were achieved when different thirds of the data were withheld for pruning different trees.

The potential for one or more of the parallel decision trees to produce an incorrect answer necessitates additional processing on the 45 output variables. Each set of 45 parallel outputs is represented as an output matrix, and it is worthwhile to note that the number of possible output matrices is much larger than the total number of possible groupings:

$$\begin{array}{ccc} 2^{45} \approx 10^{13} & >> & 115{,}975 \approx 10^{5} \\ \text{\# of possible matrices} & >> & \text{\# of possible groupings} \end{array}$$

An output matrix does not correspond to any actual grouping whenever the law of transitivity is violated. For example when A is in the same group as B, and B is in the same group as C, but A is not in the same group as C. An output matrix which violates the principle of transitivity is called an inconsistent similarity matrix.

The first step toward eliminating inconsistent outputs from the parallel decision trees was to prevent the generation of inconsistent cases in the training data. Inconsistent cases would occur if the generator groups were just beginning to separate when the simulation stopped. These cases were corrected by integrating them for one or two additional seconds until the groups of generators could be clearly distinguished. A simple test for transitivity was developed for this purpose.

Although they are trained with consistent matrices, it is still quite possible for the parallel decision trees to produce an output matrix that is physically impossible. Therefore it is necessary to reconstruct the global behavior from the output matrix. This task was accomplished by calculating the Hamming distance from the output matrix to each of the 23 matrices observed in the training set data. The output matrix is assigned to the closest known matrix in terms of the Hamming distance. Frequently, the 45 outputs exactly match one of the known matrices, for a Hamming distance of zero. Nonzero Hamming distance values are multiples of 2 because the similarity matrices are symmetric.

SIMULATION RESULTS

Because only 23 different possible groupings occur in the training set, it was possible to compare the parallel decision tree method with a single classifier approach. The latter uses one tree to decide between the 23 cases. Both the parallel trees and the single tree achieved good results on the training data of 2040 cases, and had hit scores of 88.6% and 89.5% respectively on the the test data with 544 cases. A difference between the two methods is that the Hamming distance from the output matrix to the nearest known matrix turns out to be a very good measure of confidence in the predictions of the parallel technique, as explained below.

The number of errors in prediction decreases dramatically when the parallel decision tree output is very close to one of the stored patterns. This makes intuitive sense because the Hamming distance is (twice) the number of parallel decision trees in disagreement with the known matrix. In fact, when the parallel output exactly matches one of the known matrices, the error rate is very low. Furthermore, this happens about 70% of the time. The following table breaks down the classification success rates as a function of the Hamming distance.

test set

ham.dist	#right	#wrong	#total	%right
0	378	1	379	99.74%
2	56	9	65	86.15%
4	22	15	37	59.46%
6	18	13	31	58.06%
8	5	13	18	27.78%
>8	3	11	14	21.43%
total	482	62	544	88.60%

Table 1: Success rate broken down by Hamming distance.

Out of 544 cases, the trees predict 379 cases with zero Hamming distance, and have only one error when the Hamming distance is zero. We therefore have a high degree of confidence in the prediction when the Hamming distance to the nearest stored pattern is zero. This idea is made more precise below.

To estimate the error rate of a classifier, one divides the number of errors by the total number of cases. This division yields a point estimate of the true or underlying error rate. In order to describe the distribution for this error rate we applied a Bayesian model. Specifically the observed error rate is modeled as an observation from a set of Bernoulli trials. We used a Beta distribution of the conjugate prior for Bernoulli trials, as the prior distribution for the error rate [7]:

$$h(y|\theta) = \frac{\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}}{\mathrm{Beta}(y+\alpha, n-y+\beta)}$$

where:

θ = True proportion of success
y = number of successes
n = number of trials

From this distribution we can perform the necessary statistical tests. What we are looking for is a way of isolating the predictions made with high confidence from the predictions made with less certainty. When the Hamming distance is equal to 2, for example, there are 56 successes out of 65 test cases, for a maximum likelihood performance estimate of 86.15%. However, the Beta distribution must be analyzed in order to establish confidence in the accuracy of this observed percentage. The narrowness of the Beta distribution determines the confidence in the measured success rate. A 90% confidence margin is defined to be the interval (r,1], such that there is a 90% probability that the true error rate is somewhere between r and 1. It is therefore necessary to know the confidence margin associated with a particular observed success rate in order to project its reliability. Having a substantial number of cases is necessary for obtaining a narrow confidence window.

The Beta distribution analysis can also be applied to the observed success rates of the different terminal nodes from the single classification tree approach. Since the cases in the test set are spread out among several terminal nodes, there are not as many cases in each of the terminal nodes as there are cases with the Hamming distance equal to zero. Consequently, the Beta distributions corresponding to the observed success rates in the terminal nodes are much wider than the Beta distribution for Hamming distance zero. Figure 2 shows the Beta distributions corresponding to the observed success rates of the best two nodes of the single–tree approach, and for the parallel technique when the Hamming distance to the nearest known matrix is zero.

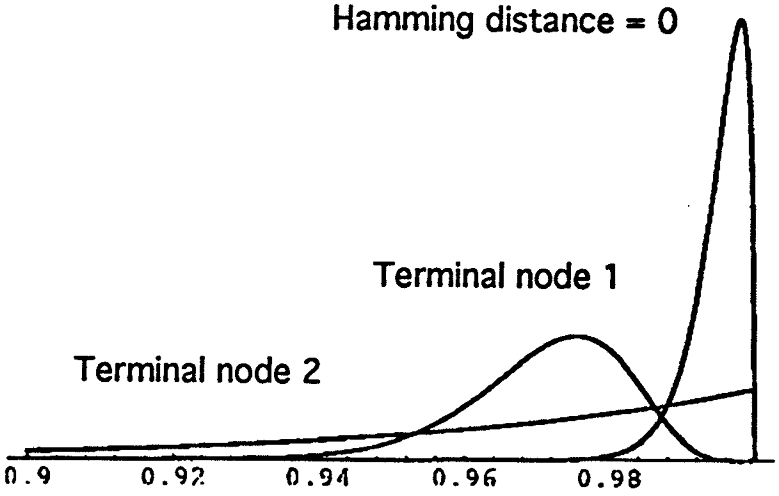

Figure 2: Probability distributions for the classifier accuracies, conditioned on the two best terminal nodes of the direct approach, and on zero Hamming distance for the parallel technique.

It is clear from Figure 2, that the classification accuracy is expected to be the best when the parallel trees classify a case with zero Hamming distance from the output matrix to the nearest known matrix. The distribution for Hamming distance zero is largely contained above .99, while the distributions for the best two nodes of the direct classification tree are much broader. Given the limited number of cases seen by each node in the direct classification tree, it is more difficult to establish a lower bound on their classification accuracy. With the parallel decision trees, 70% of the cases are decided with zero Hamming distance, and in these cases the prediction is only wrong in one instance. It would be possible to achieve tighter bounds on the classification accuracy for the single tree's terminal nodes by using a larger test set. In an actual implementation, however, one would prefer to condition on the intuitively obvious criterion of zero Hamming distance, rather than perform extensive testing on the nodes of the single tree classifier.

CONCLUSIONS

Simulations on the 39–bus system have shown that decision tree classification techniques permit future behavior to be predicted in real–time. In the domain of electric power systems, predicting system stability ahead–of–time presents new opportunities for emergency remedial control actions. In order to predict the specifics of future behavior, we have looked at predicting the future islanding pattern of unstable generators. Features of the problem include: an extremely large number of possible classes, the need for fast training, the need for very fast real–time prediction, and the need for confidence in the prediction.

To solve this problem a method of parallel decision trees was developed and implemented. The parallel decision trees predict whether every pair of generators are in the same or in different groups. All the predictions are compiled into an output matrix. The grouping corresponding to the output matrix is taken to be the nearest known matrix as determined by the Hamming distance. This parallel method of classification was compared to the single–tree method of classification where each grouping was assigned a categorical value, and prediction was implemented using the default CART options.

The averaged performances of the single classification tree and the parallel classification trees are similar on test data. The direct method has a hit score of 89.5% while parallel method has a score of 88.6% . Advantages the parallel technique include potentially being able to classify many thousands of clusterings while the direct method is limited in the number of categorical variables it can handle. The major benefit of the parallel technique is the level of confidence associated with the Hamming distance between the output matrix and the nearest known matrix. The Hamming distance provides an intuitive and statistically justifiable method for assessing confidence in the results of the parallel prediction technique.

REFERENCES

[1] A.G. Phadke, "Synchronized Phasor Measurements in Power Systems", *IEEE Computer Applications in Power*, Vol. 6, No. 2, pp. 10–15, 1993.

[2] A. G. Phadke and J. S. Thorp, "Improved Control and Protection of Power Systems through Synchronized Phasor Measurements", *Control and Dynamic Systems*, Vol. 43, pp 335–376, Academic Press, New York, 1991.

[3] S.M. Rovnyak, S.E. Kretsinger, J.S. Thorp, and D.E. Brown, "Decision Trees for Real–Time Transient Stability Prediction", IEEE PES 1993 Summer Meeting, 93 SM 530–6 PWRS.

[4] A.G. Phadke et al., "Synchronized Sampling and Phasor Measurements for Relaying and Control", IEEE PES Winter Meeting, Columbus, Ohio, February 1993 (93 WM 039–8–PWRD).

[5] L. Wehenkel, M. Pavella, "Decision Trees and Transient Stability of Electric Power Systems", *Automatica*, Vol. 27, No. 1, pp. 115–134, 1991.

[6] L. Breiman et al., *Classification and Regression Trees*, Wadsworth, Belmont, California, 1984.

[7] S.J. Press, *Bayesian Statistics, Principles, Models and Applications*, John Wiley and Sons, New York, 1989.