CHOICE OF REFERENCE IN CART APPLICATIONS USING PMU DATA

Shravan Garlapati[1], James S. Thorp[2]

Department of ECE, Virginia Polytechnic Institute and State University
Blacksburg, Virginia-24060

[1]gshra09@vt.edu

[2]jsthorp@vt.edu

Abstract — Classification and Regression trees (CART) are being used for various applications in Power Systems. PMU phasor data is used in most of these applications to build decision trees. A noticeable difference in results may be seen by changing the reference bus in CART simulations. Recently studies on PMU placement for security-dependability based adaptive relay protection scheme using data mining were proposed. Using this topic as a case study, with the help of CART simulations we have investigated the effects of changing the reference bus. In addition to this the issues related to CART simulation in representing phasor data as a single complex number attribute or as two different attributes i.e. a number associated with the real part of the complex number as one attribute and a number associated with imaginary part of the complex number as a different attribute are also discussed.

***Keywords*— PMUs, Adaptive relaying, security-dependability, decision trees, CART.**

1. INTRODUCTION

The invention of phasor measurement unit (PMU) gave birth to numerous applications in the area of power system state estimation, protection and wide area measurement systems (WAMS) control. One such application is adaptive relay protection scheme that aids to improve the reliability of the power grid. It is evident from the recent blackouts that protection relays play crucial role in the reliability of power system [1-3]. In some occasions, lack of situational awareness (WAMS) led to miscellaneous relay trips further resulting in cascading failures like blackouts. Using adaptive relays which have the ability to change their functionality, operation and logic according to power system current state can be helpful in avoiding erroneous trips and can mitigate grid from undesirable events like cascading failures [4].

Power system relay protection reliability depends mainly on two aspects i.e. dependability and security. As per IEEE [5] dependability is defined as “the degree of certainty that a relay or relay system will operate correctly” [6]. IEEE defined security “relates to the degree of certainty that a relay or relay system will not operate incorrectly” [6]. Dependability and security are two conflicting concepts. Improving security means loss of dependability and enhancing dependability means loss of security. One of the main objectives of the power system protection engineers is to provide an optimum balance between dependability and security.

Power system topology and stability allowed the traditional relay protection equipment design to be prejudiced towards dependability. Due to the manner in which the power system has evolved over the last few decades the design and operation of relays need to be reviewed inorder to provide smooth and safer operation of the system and prevent it from undesirable catastrophic events like blackouts. Therefore when the power system is in stressed state a favourable bias of relays towards security is beneficial. The 2% per year increase in electricity consumption over the last three decades have forced the power system to be operated under stressed conditions. Thus at some critical locations it has become essential to avoid relay trips when it is not necessary i.e. to be biased towards security. Protection relay bias towards dependability is desired when the power system state is in a “safe” state. Under such circumstances a fault that is not cleared has a worst impact on the system compared to a relay mis-operation due to lack of security. When the power system is in “stressed” state erroneous trips may aggravate the severity of line outages, leading to cascading failures like blackouts, thus a bias towards security is preferred.

In [7] a methodology to implement an adaptive relay protection scheme that can alter its security/dependability balance to suit to current power system state is proposed. In this paper using [7] as case study, we studied the effect of changing reference bus on resultant decision tree using CART. The main contributions of this paper are:

1) Multiple CART simulations are done using all 500KV buses as references. The reference used in [7] was a 115KV bus near the load flow swing bus. The fact that a PMU must be installed at the reference location prompted the search for a 500KV reference. After analyzing the results from different simulation runs, conclusions were drawn to use different reference for heavy winter and heavy summer 4000 California bus models for employing adaptive relay protection at a critical location, which preserved the results presented in [7].

2) A combination of decision trees and CART simulations are used for numerous applications in power systems [8]-[11]. Most of the research works used voltage magnitude, current magnitude, generator angles, real and reactive power flows, as attributes for CART simulations. It is illustrated with an example that a considerable difference in results can be seen by representing PMU phasor data used in CART simulations in the following two different formats

a) A single voltage or current complex number attribute.

b) Splitting each voltage or current complex number attributes into two different attributes. A number associated with the real part of the complex number as a different attribute and a number associated with the imaginary part of the complex number as a different attribute.

The remainder of the paper is organized as follows. Section II briefly explains CART algorithm. We obtained the data set generated using 115KV generator bus as reference from [7]. Section III summarizes the algorithm used in [7] to generate the initial data set. Our approach is discussed in section IV. Section V presents the results. Section VI explains few interesting observations from our analysis. Section VII concludes the paper.

2. CART

CART is a recursive partitioning method which builds classification and regression trees for predicting continuous dependent variables and categorical predictor variables [12]. Classification trees are built for predicting categorical predictor variables whereas regression trees are built for predicting continuous dependent variables. Breiman et al proposed classic CART algorithm [13]. Classification trees are used in power system research related to classifying PMU data. Consider a learning sample data L is used to build decision trees. A set of measurements $C = \{C_1, C_2, \ldots, C_n\}$ are used as measurement vectors. Each column of measurement vector C_i is called as attribute. Each and Every measurement vector should be classified as one of the mutually exclusive set of classes $A = \{A_1, A_2, \ldots, A_n\}$. A sample decision tree is as shown in figure 1. Head node reads input learning sample data and splits into two parts. If expected accuracy is not achieved in first split, data is split further by splitter nodes. End nodes indicate that all the measurement vectors are classified into one of the mutually exclusive set of classes. Refer to [13] for detailed CART algorithm.

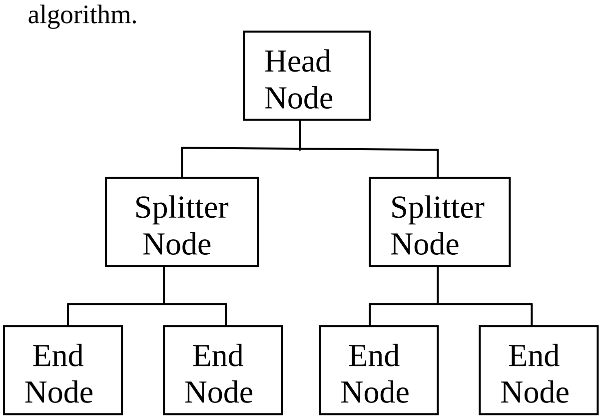

Figure 1: Sample Decision Tree

3. RELATED PRIOR STUDY

This algorithm explained in [7] is summarized in following four points.

1) According to the theory of adaptive protection, relays can change their characteristics according to the varying power system state. With the help of PMUs wide area measurements are done to obtain phasor measurements at specific buses to provide enough information to differentiate the protection relays bias towards security or dependability. If the power system is stressed, bias the protection relays towards security i.e. enable the voting scheme. On the flip side if the system state is found to be safe the voting scheme will be disarmed, only a single relay takes protective action i.e. relay operation is favourable towards dependability. Enabling the voting scheme means multiple primary relays (mostly three) protect a single transmission line and they communicate among themselves before arriving at a conclusion whether to trip a line or not. It is mentioned in [7] that adaptive relay protection is only employed at critical locations in the power grid. The above description gives rise to following questions: How to determine optimal location of PMUs from which system state can be inferred correctly? How to categorize system state as “safe” and “stressed”? Where to place the adaptive protection scheme i.e. how to identify critical locations in a power grid?

2) Decision trees are used to distinguish the system state as “stressed” or “safe”. Providing the details of decision trees are out of the scope of this paper. The details of the algorithm for growing decision trees can be found in [7]. The main goal of the algorithm is to partition the power system state space intelligently inorder to develop decision rules to adjust security/dependability balance of the relay protection scheme.

CART (Classification and Regression Trees) software is used for building decision trees in this paper. CART’s algorithm initially grows a decision tree as large as possible and selectively prunes it upwards. Cost complexity criterion is used to systematically prune the branches in a tree. The objective is to attain a minimum sized tree along with minimized cost complexity. Cost complexity criterion and branches varies depending on application. In this paper error rate to distinguish system state as “stressed” or “safe” is the cost complexity criterion and number of branches in a tree corresponds to number of PMUs.

3) Critical locations are those locations in the power grid where an unwanted line trips may result in catastrophic events like blackouts. Critical locations are intended for placement of adaptive relay protection scheme. Two different indices are used to identify critical locations in a system i.e. static index and dynamic index.

a) Static index uses load flow analysis to sort out comprehensive list cases. A double contingency which is a real fault followed by the removal of two transmission lines is considered a case. A large number of cases are created using exhaustive simulations of the real power system under study. The total number of cases created is directly proportional to the system topology, operating condition and the region of vulnerability of protection relays. For each case pre and post operating conditions are compared by solving load flow. If the system is stable after a hidden failure based fault and two line trips binary outcome is set to zero. On the other hand if the system is unstable binary outcome is set to one.

b) A shortened list is created with all cases classified as ones using static index. This list is further condensed using dynamic index. A ten second dynamic simulation run is typical for any dynamic simulation studies. Integral Square Generator Angle (ISGA) proposed in [14] is used as an index to predict angular instability. It is obtained by calculating the weighted sum of difference between generator rotor angles and center of angle. The highest value of ISGA gives the critical location of power system.

4) Having determined the critical location for adaptive relay protection scheme the next objective is to determine the placement of PMU's. The algorithm presented in [7] includes two sections developing the learning sample and training the decision tree. 4000 bus California model is used for their studies. Numerous operating points are created by load scaling at different load buses present in the power system and simultaneous load flow analysis. For every case that is created measurements are only done at 500KV buses. These measurements are enough to predict the appropriate security/Dependability balance of adaptive protection relays. For every operating point voltage phasor angles of 500KV buses, real and imaginary current flows through 500KV lines change by a considerable amount. Therefore they are used as attributes for building the decision trees. A fault and a hidden failure are assumed to occur at every operating point and load flow is performed. If load flow converges and bus voltages in pu, current flows through transmission lines and bus angles remain within their limits that particular case is considered as "safe" and is classified as zero. On the other hand if the load flow converges and any of the bus voltages, bus angles and transmission line current flows exceed their limits that particular case is considered to be "stressed" and is classified as one. CART (Classification and regression trees) algorithm is used to grow the decision trees [15]. PMU locations are indicated by splitting attributes in final decision tree.

4. OUR METHOD

In [7] PITTSBURG 115KV generator is used as reference. As a part of our work we performed simulations in [7] using multiple buses as references. Repeating simulations using multiple reference buses is time consuming. Therefore we used the following procedure to perform simulations using multiple reference buses which results in better solution in less time.

1) The Figure 2 shows a tabular representation of the learning sample data generated for a single case in [7]. Where *Va* is the voltage phasor angle at a 500KV bus. *Ir* and *Ii* are the real and imaginary currents flowing through 500KV transmission lines. There are thirty 500KV buses i.e. 30 V_a attributes and fifty one transmission lines i.e. 102 attributes related to real and imaginary current flows. In total there are 132 attributes. The total number of operating points that can be created depends on the model i.e. heavy summer and heavy winter California bus model and also on the number of load scaling operations performed on the model.

2) Instead of changing the reference for every simulation run, data obtained by using a single simulation run can be modified in such a way that the

Static Va1 Va2….Va30 Ir1 Ii1 ………Ir51 Ii51
1 0 va11 va12…va130 ir11 ii11……ir151 ii151
2 1 va21va22…va230 ir21 ii21 …….ir251 ii251
3 0 va31va32….va330 ir31ii31………ir351 ii351
. . . . ….. . . . . . ………. . .
. . . . ….. . . . . . ………. . .
N 1 vaN1vaN2…..vaN30 irN1 iiN1…..irN51iiN51

Figure 2: Learning Sample Data

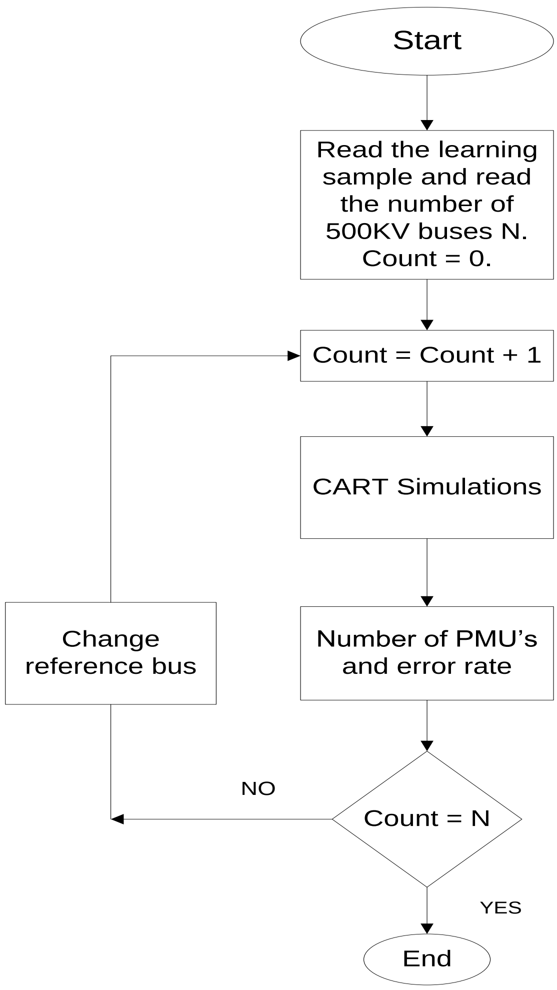

Figure 3: Flow Chart of our method

modified data appears as if it is obtained by performing simulations using a new reference. For example let us consider that the sample data pattern shown in Figure 2 is obtained by using bus 1 as a reference. Inorder to modify this data and to make it appear as if it is obtained by using bus 2 as a reference the following two steps are executed:

a) Subtract *Va2* from all angle attributes.

b) Convert rectangular coordinates of currents (real and imaginary) through a transmission line into polar coordinates. Reference bus angle *Va2* is subtracted from the polar form angle. Again the polar form current is converted back to real and imaginary values of current.

Using the above two steps we have repeated simulations using the twenty nine 500KV buses as reference. The flow chart related to our simulations is as shown in Figure 3.

5. RESULTS

We have demonstrated our algorithm on two different 4000 California bus models. The results are as shown in table 1. The first column shows the reference bus used for a particular simulation run. HS indicates the number of PMUs necessary for heavy summer model to distinguish the power system state as “stressed” or “safe” with the error rate shown in column 3. In column 4, HW indicates the number of PMUs required for heavy winter model to differentiate “stressed” or “safe” states of power grid with an error rate shown in column 5.

The initial learning sample used in our algorithm is obtained from [7]. The total numbers of training cases of heavy winter model are 4150 and the total number of cases related to heavy summer model is 11367. The difference in number of cases is due to the differences in models. From the simulation results table it can be seen that for heavy winter case there are many simulation runs (18 out of 29) where number of PMUs necessary to distinguish power system state are one. Error rate can be used to sort these out. The simulation run when TESLA is used as reference bus has very low error rate of 0.6%. So for heavy winter model simulation run using TESLA is considered to be the best PMU placement location. More importantly that PMU location happens to be the reference TESLA itself. The decision tree for heavy winter is shown in Figure 4. IR1106 in Figure 4 is the real current flowing through the line connecting two 500KV buses TESLA and LOSBANOS. Decision tree classified 1627 out of 1645 cases correctly and remaining 18 cases are misclassified. These 18 cases belonging to class 1 (relays biased to dependability) are wrongly classified as class 0 (relays biased to securability). So, misclassification or error rate is 1.1%. In the case of heavy summer model there are multiple simulation runs (9 out of 27) where number of PMU’s needed are two. All of these cases have high error rate in the range of 2 to 3.8. With the objective of achieving error rate around 1% we started looking at simulation runs with the number of PMU’s as 3. The case with VACADIXON as reference has an error rate of 1% and

Reference	HS	% Error	HW	%Error
IMPRLVLY	9	3.7	7	3.2
MIGUEL	8	3.6	5	3.9
N.GILA	3	3	1	1.9
ELDORADO	5	3.8	1	2.9
LUGO	4	3.5	1	3.5
MIRALOMA	6	3.6	1	3.7
MOHAVE	3	3.2	1	2.7
SERRANO	3	3.1	2	3.1
ADELANTO	4	3.4	1	3.5
MARKETP	2	3.4	2	3.2
RINALDI	3	3.5	3	3.8
TOLUCA	4	3.6	4	3.4
VICTORVL	3	3.8	1	3.2
RINALDI2	4	3.4	3	3.6
CRYSTAL	2	3.6	3	2.3
ROUNDMT	2	3.7	3	3.1
TABLEMT	2	3.8	1	2.3
OLINDA	2	3.7	4	3.7
MAXWELL	2	3.7	1	1.9
TRACY	2	3	1	2.1
METCALF	8	2.5	1	2.1
MOSSLAND	2	3	1	3.6
LOSBANS	4	3	1	2.4
GATES	4	3	1	2.4
DIABLO	3	3.3	1	1.6
MIDWAY	3	3.8	1	2.4
TESLA	2	2	1	0.6
VACADIXON	3	1	1	2.5

TABLE 1: SIMULATION RESULTS

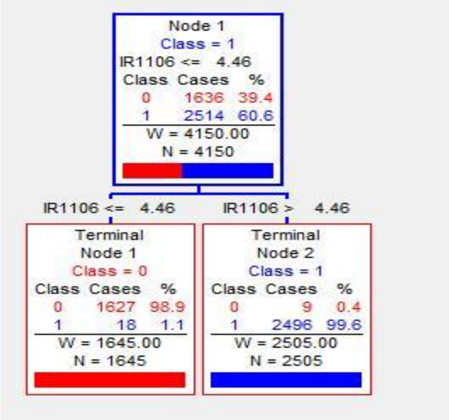

Figure 4: Decision tree for Heavy winter using TESLA as reference

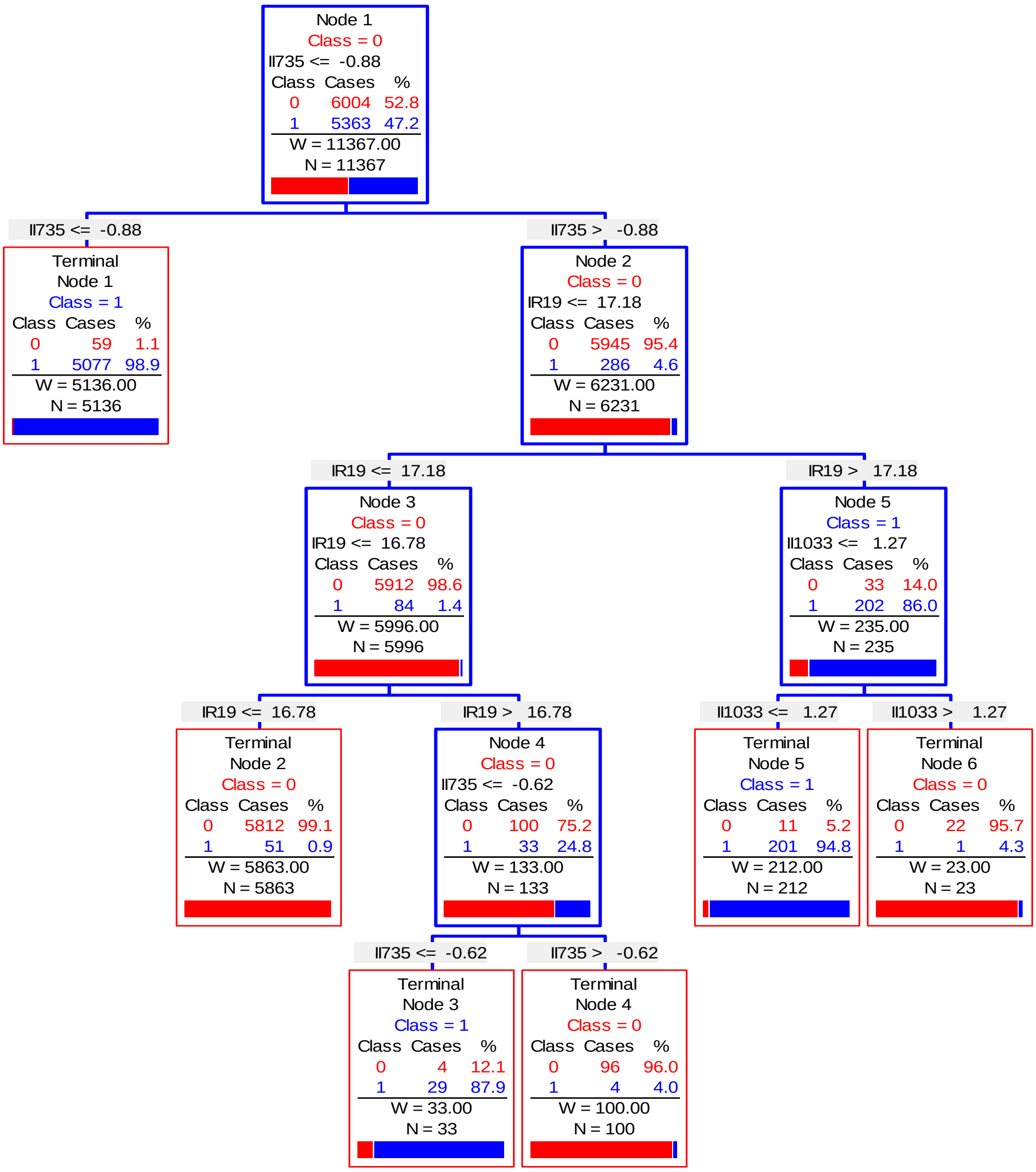

Figure 5: Decision tree for heavy summer model using VACADIXON as reference

3 PMUs to distinguish power system state. Interestingly one of those PMUs appears to be at TESLA which is one of the PMU locations in heavy winter case. Therefore just two extra PMUs are required for heavy summer case with VACADIXON as reference and error rate is also low i.e. 1%.

The decision tree for heavy summer model is as shown in Figure 5. II735 is the imaginary part of the current flowing through the 500KV line connected between buses *Devers* and bus *ValleySC*. IR19 is the real part of the current flowing through the 500KV line connected between buses *Paloverde* and *Devers*. II1033 is the imaginary value of the current flowing through 500KV line connected between buses *Diablo* and *Midway*. Therefore total numbers of PMU's required for combined HW and HS cases are four. They are at TESLA, VACADIXON, DEVERS and DIABLO. The decision tree obtained after CART simulations for both HW and HS are tested for robustness using out of sample data. Some more operating points are created by simulating outages in generators with capacity greater than 200 MW, loads consuming more than 200MW, tripping 230KV and 500KV transmission lines.

The total number of out of sample cases created for heavy winter is 660 and heavy summer is 1155 cases.

Again the difference is due to the differences in models. The performance of the decision trees for heavy summer and heavy winter models using training cases and out of sample cases is as shown in table II. Taking both training cases and out of sample cases into consideration the error rate is 220/17732 i.e. 1.26%. The number of PMUs is same as those required in [7] and error rate is also approximately same around 1%. Therefore by changing the reference bus the results from [7] can be preserved.

Our analysis is only tested on 4000 bus California model. If the approach presented in section III is applied to other power grids there may be reduction in number of PMUs. Thus before installing PMUs for adaptive relaying protection it is better to analyse the power grid using our approach for better accuracy and reduction in cost. Including communication and retrofitting the installation cost of a single PMU is very high. Apart from this each installed PMU also has recurring cost for its maintenance throughout its lifetime. If the number of PMUs is reduced, associated costs can be avoided for the reduced number of PMUs.

The interesting part of our results is that different reference is used for different models. For heavy summer (HS) 4000 bus California model VACADIXON is used as reference and TESLA is used as reference for 4000 bus heavy winter (HW) California model. Both these buses are 500KV buses. In [7] PITTSBURG which is 115KV bus is used as swing bus for both HW and HS.

6. OBSERVATIONS

This section explains some important observations made from our simulations. Consider a sample scenario shown in Figure 6 and Figure 7. The sample size used in figures is very small (40) just for illustration purpose. Red dots indicate power system operating condition in stressed state and blue dots indicates safe state. Figure 6 shows the data presented to CART for a possible splitting node. The dotted line parallel to x-axis is the best possible single column splitter for this data. Splitting on a single column is the simplest and quickest approach and yields satisfactory results in most applications. It is obvious that two blue dots which are supposed to be above the dotted line are below it, resulting in a 5% error. The remaining splitting nodes will capture the two blue dots.

Model	Training Cases	Errors	Out of Sample	Errors
Heavy Winter	4150	27	660	21
Heavy Summer	11367	130	1155	42
Total	15517	157	1815	63

TABLE 2: PERFORMANCE OF DECISION TREES

However, there is a problem in splitting on the real and/or imaginary part of complex measurements in a decision tree with real time data. There must be a reference angle for complex measurements such as synchrophasor. If a load flow is used to generate training data the swing bus can be used, for example. But the actual application will require a physical reference and angles will have the measurement of the reference subtracted from the other measurement angles. The problem develops when the reference is changed because the PMU on the reference fails or the utility decides to install the needed PMU elsewhere. Splitting on real and/or imaginary parts of currents require a new tree for a new reference.

The tilted line in Figure 7 illustrates the solution to the reference problem. It is a better solution than the horizontal line in Figure 6 and is a linear combination split in CART. It is of the form

$$\alpha \text{ real} + \beta \text{ imag} = \gamma,$$

Only linear splits involving the real and imaginary part of a complex measurement are sought to accomplish the separation in Figure 7. It is hard to limit CART to only those linear combination splits and allowing all linear splits is quite time consuming. While the performance of the linear combination splitter for Figure 7 is better than the performance of the single variable splitter of Figure 6, it is the use of linear combination splitters to solve the reference angle problem is even more important. For speed and simplicity assume the training data has some reasonable reference and that single columns are used for the tree. A tree and splits on lines such as

$$\text{Real}(y_j) = g_1 \text{ or } \text{Imag}(y_j) = g_2,$$

The effect of changing the reference is to rotate the data as shown in Figure 8. The circles show the effect of a change of reference while the tilted line is the horizontal splitter of Figure 6 rotated by the same angle. The line *imag(y) =1.95* has been rotated to the line *0.5 real(x) + 0.866 imag(x) =1.95*

The conclusion is that the performance of an original decision tree built using CART simulations with complex number PMU phasor data attributes can be maintained with changing reference by rotating the splits on a single column to linear combination splits based on rotating the horizontal and/or vertical splits of the original tree. A lot of researchers in power systems are using decision trees on PMU data this discussion may be helpful to them.

7. CONCLUSION

We have studied the problem of changing reference for building decision trees using CART simulations with PMU phasor data. When a phasor complex number data is represented as two different attributes, using a security-dependability based methodology for adaptive relay protection scheme demonstrated on a 4000 bus California model as a case study it is shown that the output decision tree, number of PMUs and error rate change with changing reference. On the other hand if synchrophasor data is represented as a single complex number attribute, it is illustrated with example that the performance of original decision tree can be preserved by rotating the horizontal/vertical splits to linear combination splits.

ACKNOWLEDGEMENT

This work was supported by the California Institute for Energy and the Environment Contract TRP-08-06**,** funded by the California Energy Commission's Public Interest Energy Research Program. This paper does not necessarily represent the views of the Energy Commission, its employees or the State of California. It has not been approved or disapproved by the Energy Commission nor has the Energy Commission passed upon the accuracy or adequacy of the information.

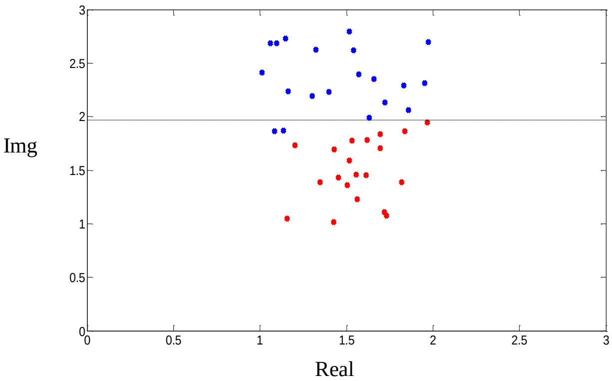

Figure 6: Dotted line shows a splitter as a result of CART simulations

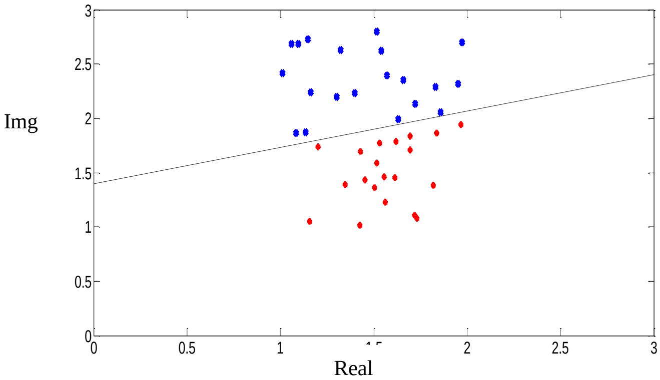

Figure 7: Dotted line with a slope – may be best possible splitter

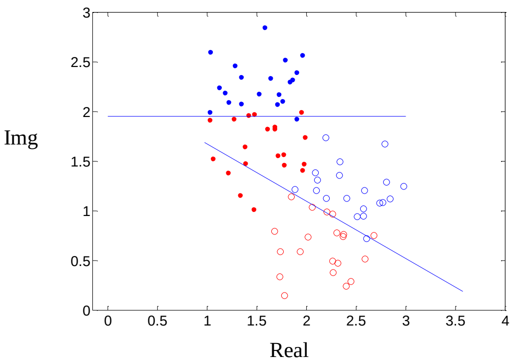

Figure 8: Both the data and splitter from Figure 6 rotated by 30^0 representing a shift in the reference

REFERENCES

[1] NYISO, "Blackout August 14, 2003 Final Report", 2005.
[2] UCTE, "Final Report: System Disturbance on 4 November 2006".
[3] UCTE, "Final Report of the Investigation Committee on the 28 September 2003 Blackout in Italy", 2004.
[4] Horowitz, S.H., A.G. Phadke, and J.S. Thorp, "Adaptive transmission system relaying. Power Delivery, IEEE Transactions on, 1988. **3**(4): p. 1436-1445.
[5] IEEE, "Standard for Relays and Relay Systems Associated with Electric Power Apparatus" 1989.
[6] NERC, "Addressing the directives issued by FERC, in Order 706 relative to the approved Cyber Secuirty Standards CIP-002-1 through CIP-009-1", 2008-06.
[7] Emanuel Bernabue, James S.Thorp "A Methodology for a Security-Dependability Adaptive Protection Scheme on Data Mining" http://scholar.lib.vt.edu/theses/available/etd-12172009-201325/unrestricted/ Bernabeu_EE_D_2009.pdf
[8] Steve Kretsinger, Steven Rovnyak, Donald Brown and James Thorp "Parallel Decision Trees for Predicting Groups of Unstable Generators From Synchronized Phasor Measurements", Precise Measurements in Power Systems Conference, Washington, D.C., Oct 25-29, 1993.
[9] Steven Rovnyak, Steve Kretsinger, Donald Brown and James Thorp "Decision Trees for Real Time Transient Stability Prediction" IEEE Transactionson Power System, Vol. 9, No.3, August 1994.
[10] Anil Swarnkar and K.R.Niazi "CART for online Security Evaluation and Preventive Control of Power Systems", proc of 5th WSEAS/IASME Intl. Conference of Electric Power Systems, High Voltages, Electric Machines, Tenerife, Spain December 16-18 2005.
[11] Ruisheng Diao, Kai Sun, Vijay Vittal, Robert J. O'keefe, Michael R. Richardson, Navin Bhatt, Dwayne Stradford, Sanjay K. Sarawgi "Decision Tree Based Online Voltage Assessment Using PMU measurements" IEEE Transactions of Power Systems VOL. 24, No.2, May 2009
[12] Statistics Methods and Applications by Pawel Lewicki and Thomas Hill.
[13] Classification and Regression Trees by Leo Breiman, Jerome H Friedman, Richard A. Olshen and Charles J. Stone.
[14] Northeast Power Failure: November 9 and 10, 1965. 1965, Federal Power Commission.
[15] CART. Salford Systems, salford-systems.com.