A Hybrid Intelligent System for Estimating a Load Margin to Saddle Node Bifurcation Point of Voltage Stability

Hiroyuki Mori *Naoto Ishibashi*
Department of Electronics and Bioinformatics
Meiji University
Kawasaki, 214-8571, Japan
hmori@isc.meiji.ac.jp

***Abstract*—This paper proposes a hybrid intelligent system for estimating a load margin to the saddle node bifurcation point of voltage stability. The proposed method is based on the integration of Regression Tree (RT) and Artificial Neural Network (ANN). Voltage stability analysis is one of the main concerns in power system operating and planning. Voltage stability analysis aims at evaluating the saddle node bifurcation point on PV or QV carves. So, it is necessary to estimate a load margin to the saddle node bifurcation point of voltage stability efficiently. In this paper, a new method is proposed to estimate the load margin with the hybrid method of RT and ANN. RT is used to classify data into terminal nodes and extract rules from each terminal node. ANN is constructed to estimate the load margin to the bifurcation points at each terminal node. Also, a new method for generating power system conditions is presented to consider the correlation of the nodal specified values. The proposed method is successfully applied to the IEEE 30-bus system in terms of computational accuracy and computational time.**

Keywords-Voltage Stability; Hybrid Intelligent System; Load Margin Estimation; Continuation Power Flow; Regression Tree; Artificial Neural Network; Data Mining

I. INTRODUCTION

This paper presents a hybrid intelligent system for estimating load margin to the saddle node bifurcation point of voltage stability in a electric power systems. The proposed method makes use of the integration of RT (Regression Tree) and ANN (Artificial Neural Network). Voltage instability phenomena occurs due to the following reasons:

- sudden increase of load
- outage of transmission lines
- behavior of transformer taps, *etc*

As a framework of security assessment, voltage stability assessment has been studied to avoid voltage collapse of the worst scenario [1]. The conventional methods on voltage stability may be classified as follows:

1) Method of Dobson [2,3]

2) Method of Alvarado [4]

3) Continuation Power Flow Calculation [5,6]

4) Look-ahead method [7]

Methods 1) and 2) evaluate the saddle node bifurcation points directly, but they do not give the margin to it. In that sense, their methods are not acceptable from a stand point of system operation and planning. Method 3) evaluates PV or QV curves with the extended power flow calculation. It is one of robust methods that give the saddle node bifurcation point. Also, the method easily gives the load margin to it. Method 4) makes use of the results obtained by Method 3) to estimate the saddle node bifurcation point with the approximate curve. In recent years, voltage stability assessment becomes more important due to the uncertainties of the power system liberalization and distributed generation. As power systems are inclined to be more complicated, voltage stability assessment requires an efficient method that estimates the saddle node bifurcation point.

In this paper, a new hybrid intelligent method [8,9] is proposed to estimate the saddle node bifurcation point. It speeds up computational efficiency in estimating the saddle node bifurcation point with high accuracy. It is based on the integration of RT [10-12] and ANN [13,14]. The former is used to preprocess learning data of ANN in a way that it is classified terminal nodes. It is known that classifying learning data into cluster gives better results in the prediction problem []. As RT, this paper uses CART (Classification and Regression Trees) [10] that has functions to extract if-then rules and evaluate the variable importance. The latter plays a key to estimate the saddle node bifurcation point with high accuracy. As ANN, this paper employs MLP (Multi-layer perceptron) that consists of three layers to estimate the load margins. Although the proposed method with the hybrid intelligent system requires a lot of computational time for the learning process, it quickly evaluates estimates for unknown system conditions. Also, a new method is presented to consider the correlation between the nodal specified values in generating power system conditions. The moment matching method is used to include the correlation in Monte Carlo Simulation (MCS) [15,16]. The effectiveness of the proposed method is demonstrated in the IEEE 30-bus system in terms of computational time and computational accuracy.

II. CONTINUATION POWER FLOW

This section describes the continuation power flow calculation that is used for learning data. It is the technique to draw PV or QV curves effectively by tracing a solution curve. In this paper, CPFLOW proposed by Chiang, *et al.* [6] is explained. It is one of the continuation power flow calculation methods with the predictor-corrector technique. Fig. 1 shows the concept of CPFLOW, where a set of power flow solutions are evaluated by changing the nodal specified value. The predictor-corrector technique continually evaluates a senior of solutions in a way that the predictor estimates the initial solution and the corrector calculates the solution with the initial solution for the N-R method. In the next paragraph, the predictor and the corrector are outlined.

A. Predictor

The predictor determines an initial point to trace a solution curve. CPFLOW uses two kinds of the tangent and the secant predictors. The tangent predictor employs a tangent vector on the solution curve at a current solution. It is time-consuming to calculate the tangent vector. On the other hand, the secant predictor is a simple algorithm that extrapolates two known points to obtain a new initial solution. Its calculation time is much faster than the secant predictor. Therefore, CPFLOW has a feature to switch from the predictor to the secant one of after calculating the first solution with the initial solution.

B. Corrector

The corrector is employed to perform the power flow calculation by the N-R method using an initial value provided by predictor. CPFLOW formulates the extended power flow equation by load parameter λ into the conventional one. In this paper, the extended power flow equations may be written as:

$$P(V,\theta)-(1+\lambda)f_{Pi}=0 \quad \text{..................} (1)$$

$$Q(V,\theta)-(1+\lambda)f_{Qi}=0 \quad \text{..................} (2)$$

where

$P(V,\theta)$: nodal active power

$Q(V,\theta)$: nodal reactive power

λ : load parameter

f_{Pi} : specified value of active power

f_{Qi} : specified value of reactive power

The use of load parameter λ works to overcome the singularity of Jacobian matrix of the power flow equation. However, it is necessary to add a constraint to evaluate the solution. CPFLOW defines a variable to be called pseudo arclength as the constraint. It may be written as:

$$\sum_{j=1}^{2(n-1)}\left\{\left(x_j^{k+1}-x_j^k\right)^2\right\}+\left(\lambda^{k+1}-\lambda^k\right)^2=\Delta s^2 \quad \text{..................}(3)$$

where

n : number of number

k : step

x : voltage vector

Δs : pseudo arclength variation

The Euclidean norm of the pseudo arclength variation is equal to that of the updating term of all the variables.

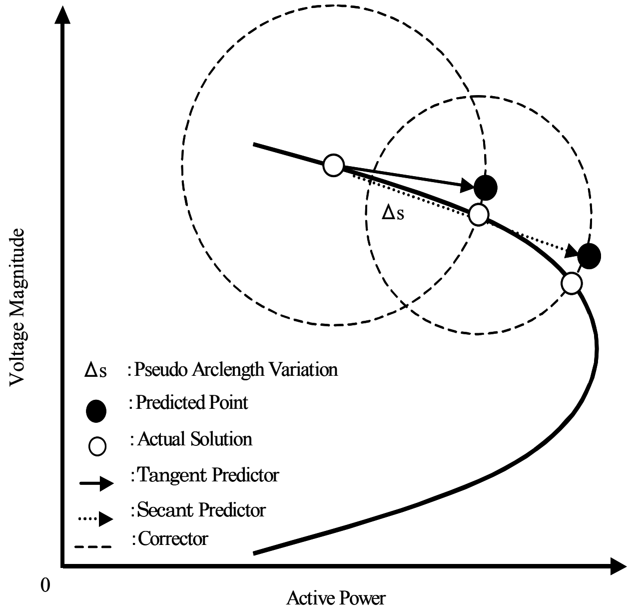


Fig.1. Concept of CPFLOW.

III. REGRESSION TREE

In this section, the regression tree (RT) is explained. It is one of the data mining techniques that classified data into clusters and extract the if-then rules. It is one of the decision trees. Fig. 2 shows the concept of decision tree, where it consists of splitting and terminal nodes. The diamond shape nodes show a splitting node to classify data and the square nodes indicate terminal node. As RT, this paper makes use of CART [10] proposed by Breiman, *et al.* After it brings up a tree to the biggest tree, the process of pruning is carried out to the smallest tree. During the pruning, it calculates the cross validation error to select the optimal tree. Also it clarifies input and output relationship with calculating variable importance statistically.

A. Growth of Tree

The construction of CART divides data into two children nodes in a way that the between-groups sum of squares is maximized at the parent node. The tree is grown up to the maximum one by splitting the obtained children nodes with the parent one. The between-groups sum of squares in a parent node may be written as:

$$\Delta S(t) = S(t) - S(t_L) - S(t_R) \quad \cdots\cdots (4)$$

where

t : number of node

$\Delta S(t)$: between-groups sum of squares at node t

$S(t)$: sum of square at parent node

$t_{L(R)}$: number of left(light) children node

$S(t_{L(R)})$: sum of square at node $t_{L(R)}$

Also, the sum of square at the parent node and at the left (right) node may be written as:

$$S(t) = \sum_{i=1}^{N} (y_i - \bar{y})^2 \quad \cdots\cdots (5)$$

$$S(t_{L(R)}) = \sum_{i=1}^{N_{L(R)}} (y_{L(R)i} - \bar{y}_{L(R)})^2 \quad \cdots\cdots (6)$$

where

N : number of data at parent node

y_i : object variable at parent node

$\bar{y}$: average of y_i

$N_{L(R)}$: number of data at left (right) children node

$y_{L(R)i}$: object variable at left (right) children node

$\bar{y}_{L(R)}$: average of $y_{L(R)i}$

B. Pruning of Tree

The tree grown up to the maximum is often overfitted for learning data. It is necessary to prune the tree and obtain a reasonable model. Specifically, the process of pruning is carried out by calculating complexity degree parameter α of (7) at the splitting node and replacing the node with the minimum with the terminal node. It repeats this process until the tree becomes the minimum with only one node. The complicated degree parameter is given as follows:

$$\alpha(t) = \left(\frac{S(t) - \sum_{T_i \in \tilde{T}} S(T_i)}{|\tilde{T}| - 1} \right) \quad \cdots\cdots (7)$$

where

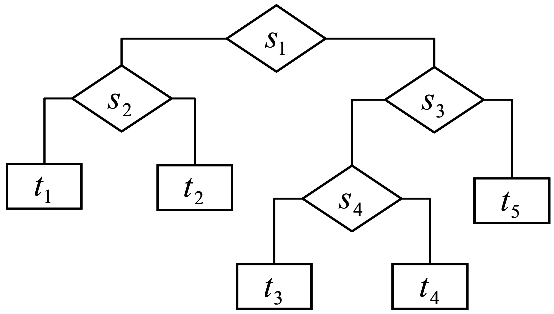
$\alpha(t)$: complicated degree parameter

T_i : number of terminal nod

$\tilde{T}$: number of terminal nodes under node t

C. Selection of Optimal Tree

In the pruning process, CART uses the cross validation method as the error criterion. It is the error evaluation method that is useful for cases of insufficient learning data or equalization of biased learning data. First, it divides learning data into ν groups and use one group as test data while using



Note) s : Splitting Node, t : Terminal Node, i : Node Number

Fig. 2. Concept of decision tree.

the (ν - 1) group as the learning data. The cross validation error may be written as:

$$R^{cv}(d) = \frac{1}{\nu} \sum_{i=1}^{\nu} R^{ts}(d_i) \quad \cdots\cdots (8)$$

$$R^{ts}(d_i) = \frac{\sum_{i=1}^{N} (y - \bar{y})^2}{N} \quad \cdots\cdots (9)$$

where

d : cross verification tree

$R^{cv}(d)$: cross validation error

ν : cross validation number of times

$R^{ts}(d_i)$: error at test data

Also, the standard error may be written as:

$$\sigma(R^{cv}(d)) = \sqrt{\frac{1}{\nu} \sum_{i=1}^{\nu} (R^{ts}(d_i) - R^{cv}(d))^2} \quad \cdots\cdots (10)$$

where, $\sigma(R^{cv}(d))$: standard error

CART calculates the cross validation error each time the pruning process is carried out. The SE rule is used to select the optimal tree. The SE rule may be written as:

$$R^{cv}(T_{cand}) \leq R^{cv}(T_{\min}) + \sigma(R^{cv}(T_{\min})) \quad \cdots\cdots (11)$$

where

T_{cand} : candidate of optimal tree

$T_{\min}$: error smallest tree

D. Variable Importance

The variable importance is the index that shows which input variable is more important statistically. It is evaluated by the between-groups sum of squares for used and unused variables. The variable importance may be written as:

$$VI_j = \frac{\sum_{t \in T_{opt}} \Delta S(x_j, t)}{\max\left(\sum_{t \in T_{opt}} \Delta S(x_j, t)\right)} \times 100 \quad (12)$$

where

VI_j : variable importance

x_j : input variable

T_{best} : optimal tree

$\Delta S(x_j, t)$: between-groups sum of squares at splitting node t

IV. PROPOSED METHOD

This paper proposes a hybrid intelligent method of CART of data mining and MLP of ANN for estimating the load margin to the saddle node bifurcation point of voltage stability in power systems. The evaluation of the load margin estimation is one of important problems in power systems. Various studies have been performed to estimate the load margin. However, there is still room for improvement in terms of accuracy and computational time. In this paper, a hybrid intelligent system is proposed to improve the method for estimating the load margin. Fig. 3 shows the outline of the proposed method. First, the proposed method makes the scenarios of the system conditions by the multidimensional normal random number to consider the correlation of each load bus and evaluates the load margin using CPFLOW. Next, the proposed method constructs RT for learning data obtained by CPFLOW and classifies learning data into the terminal nodes. ANN is constructed to estimate the load margins at each terminal node.

It is known that Monte Carlo Simulation (MCS) in consideration of the correlation between input variables gives better solutions [15,16]. In this paper, the moment matching method is used to consider the correlation between the nodal specified values of the power flow calculation. As the moment matching method, the combination method is employed to adjust the first and the second order moment of the random number [15,16]. The combination method may be written as:

$$y = C\tilde{C}^{-1}(x - m) + \mu \quad (13)$$

where

y : input scenario vector

C : lower triangular matrix such that

$s = CC^T$

where, s : covariance matrix of learning data

$\tilde{C}$: lower triangular matrix such that

$\tilde{S} = \tilde{C}\tilde{C}^T$

where, $\tilde{S}$: covariance matrix of samples

m : mean vector of sample

μ : test data vector

Finally, the algorithm of the proposed method may be written as:

Step 1: Set the system conditions.

Step 2: Generate the multidimensional normal random number with correlation of the nodal specified value and evaluate the load margin with CPFLOW.

Step 3: Construct the decision tree with CART for learning data that includes power system conditions and the load margin.

Step 4: Learn MLP at each terminal node.

Step 5: Give unknown power system conditions and estimate the load margin.

V. SIMULATION

A. Simulation Conditions

The proposed method is applied to the IEEE 30-bus system. The following simulation conditions are used:

1) To create learning data, 1250 power system conditions are generated by the multidimensional normal random numbers. The range of the load variations varies from –60% to +60% of the original normal specified value of the load flow calculation.
2) To calculate the load margin, CPFLOW is employed for 1250 power system conditions. The termination conditions of CPFLOW is that the solution on the lower point of the solution curve passes through the solution on the lower point corresponding to the initial solution on the upper point. The power mismatch the N-R method is 10^{-3}.

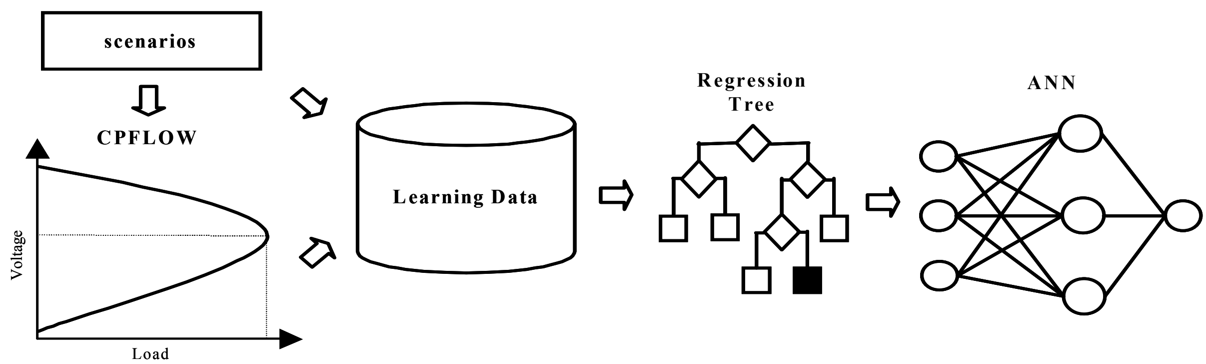


Fig. 3. Outline of the proposed method.

Table 1
Parameters of each method

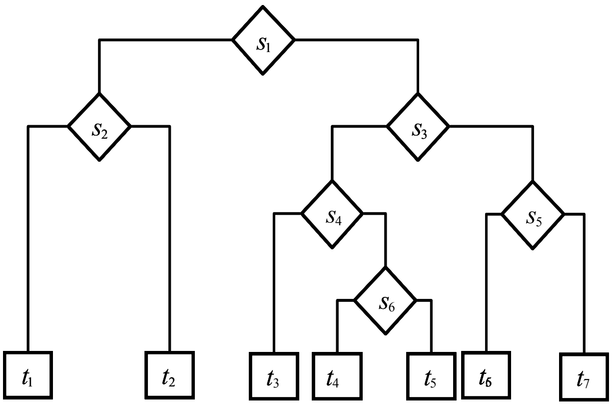
Methods	Terminal Node Number	Step Length	Tolerance Value	$\Delta\lambda$max	ν	No. of Hidden Units	Learning Coefficient α	Learning Coefficient β	Moment Coefficient	Learning Iterations
CPFLOW	-	0.4	1.00E-03	0.7	-	-	-	-	-	-
CART	-	-	-	-	10	-	-	-	-	-
MLP	-	-	-	-	-	10	0.3	0.3	0.1	30000
MLP with CART	1	-	-	-	-	10	0.65	0.65	0.1	30000
	2	-	-	-	-	10	0.5	0.5	0.1	30000
	3	-	-	-	-	10	0.1	0.1	0.1	30000
	4	-	-	-	-	10	0.35	0.35	0.1	30000
	5	-	-	-	-	10	0.5	0.5	0.1	30000
	6	-	-	-	-	10	0.3	0.3	0.1	30000
	7	-	-	-	-	10	0.5	0.5	0.1	30000

Table 2
Features of splitting nodes

Splitting Node Number	No. of Data	Average [p.u]	Splitting Variable	Splitting Value [MW]
1	1000	0.309	P_{30}	9.039
2	105	0.236	P_{30}	7.626
3	895	0.318	P_{30}	11.455
4	698	0.310	P_{30}	10.114
5	197	0.346	P_{30}	12.899
6	534	0.315	P_{30}	10.785

Table 3
Features of terminal nodes

Terminal Node Number	Learning Data		Test Data	
	No. of Data	Average [p.u.]	No. of Data	Average [p.u.]
1	44	0.200	7	0.215
2	61	0.262	12	0.261
3	164	0.293	48	0.291
4	355	0.312	82	0.311
5	179	0.322	45	0.321
6	150	0.338	38	0.338
7	47	0.371	18	0.366



Note) s : Splitting Node, t : Terminal Node, i : Node Number

Fig. 4. Regression tree of load margin

3) CART is created by using data obtained from MCS in consideration of the correlation. The number of learning data is 1000 while the number of test data is 250. Also, parameter ν of the cross validation method is set to be 10 (see Table 1). Also, the parameters of MLP at each terminal node are given in the same table. They are determined by the preliminarily simulation.

4) To demonstrate the effectiveness of the proposed method, it is compared with MLP of the conventional method in terms of model accuracy and computational time. The parameter of MLP is also given in Table 1.

5) All the calculation is performed on the Dell Dimension 9150 (Intel(R) Pentium(R) D CPU 3.20GHz 3.19GHz, 3.62GB RAM).

B. Simulation Results

Fig. 4 shows the results of CART for the load margin. From the figure, learning data are classified in 7 terminal nodes through 6 splitting nodes. Table 2 gives the feature of the splitting nodes, where the number of splitting nodes, the number of data assigned to the splitting node, the average of data, the splitting variable and its value are shown. It can be seen that splitting variable P_{30} is more important since it is used at all the splitting node. Table 3 shows the feature of terminal nodes, where the number of data and the average at each terminal node are given for learning and test data, respectively. It can be seen that terminal node 4 has the most common power system conditions. On the other hand, terminal nodes 1, 2 and 7 correspond to comparatively rare power system conditions due to a few of data.

Next, Fig. 5 shows the maximum and the average errors of the conventional and the proposed methods. The proposed method improved 1.363% in the maximum error and 0.094% in the average error than the conventional method. This is because the classification of data works to suppress the variance of data so that data similarity is obtained.

Table 4 shows a comparison of computational time for CPFLOW, MLP, and the proposed method. CPFLOW spent

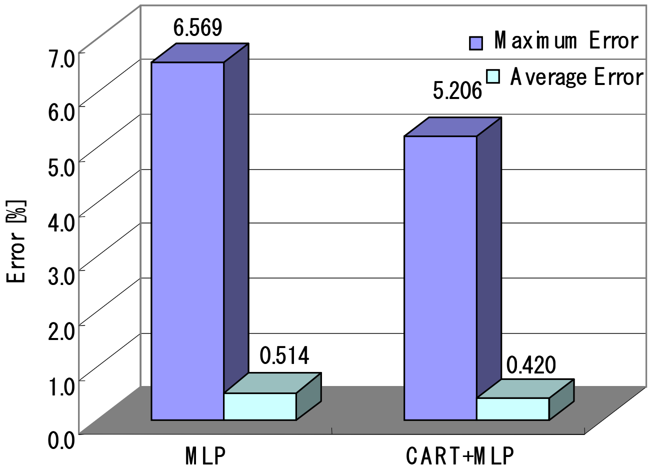


Fig. 5. Maximum and average errors of each method

Table 4
Comparison of computational time

Methods		CPU Time [s]	
		Total	Average
CPFLOW		45.72	3.658E-02
MLP	Learning	309.6	-
	Test	0.016	6.400E-05
CART + MLP	Learning	327.8	-
	Test	0.026	1.040E-04

45.72[s] for 1250 power system conditions, which means that the average CPU time is 3.658×10^{-2}[s]. MLP and the proposed method take 309.6[s] and 328.8[s] for learning data, respectively while they spent 0.016[s] and 0.026[s] for test data, respectively MLP and the proposed method gives a solution quickly once the learning process is finished. Thus, MLP and the proposed method are useful for contingency screening that requires computational efficiency.

VI. CONCLUSION

This paper has proposed a hybrid intelligent system with CART of data mining and MLP of ANN for estimating the load margin to the saddle node bifurcation point of voltage stability. CART was used to works as a preprocessing technique that data is classified into clusters. MLP was employed to estimate the load margins using learning data obtained by CPFLOW of the continuation power flow calculation. To generate realistic power system conditions for learning data, the calculation of the nodal specified value is considered by the moment matching method. The proposed method was successfully applied to the IEEE 30-node system. The proposed method succeeded in reducing 1.363% of the maximum and 0.094% of the average errors for MLP of the conventional intelligent method. Regarding computational time, the proposed method was a little worse than MLP, but it was about 365-times faster than CPFLOW. Therefore the proposed method with high accuracy and computational efficiency is very useful for contingency analysis in voltage stability.

REFERENCES

[1] H. Mori and Y. Komatsu,"A Hybrid Method of Optimal Data Mining and Artificiall Neural Network for Voltage Stability Assessment," Proc. of 2005 IEEE PowerTech, CD-ROM, Petersburg, Russia, Jun. 2005.

[2] I. Dobson, H. D. Chiang, J. S. Thorp and L. Fekih-Ahmed, "A model of voltage collapse in electric power systems," Proc of 27th IEEE Conf. Decision Cont, pp.2104-2109, 1988

[3] I. Dobson and H. D. Chiang, "Towards a theory of voltage collapse in electric power systems," Syst. and Cont. Lett, Vol.13, pp.253-262, 1989

[4] C. A. Canizares and F. L. Alvarado, “Point of Collapse and Continuation Methods for Large AC/DC Systems”, IEEE Trans. on Power Syst., Vol.8, No.1, pp.1-8, Feb. 1993

[5] K. Iba, H. Suzuki, M. Egawa and T. Watanabe, “Calculation Loading Condition with Nose Curve Using Homotopy Continuation Method”, IEEE Trans. on Power Syst., Vol.6, No.2, pp.548-593, May. 1991.

[6] H. D. Chiang, A. J. Flueck, K. S. Shah, and N. Balu, “CPFLOW: A Practical Tool for Tracing Power System Steady-State Stationary Behavior Due to Load and Generation Variations”, IEEE Trans. Power Syst., Vol.10, No.2, pp.623-634, May. 1995.

[7] K. Seki and H. Mori, “A Fast Estimation Method of Voltage Stability Margin Considering N-1 Based Contingency”, Proc. of Joint Technical Meeting on Power Engineering and Power System Engineering, IEE of Japan, PE-08-91/PSE-08-100, Kumamoto, Japan, Aug. 2008.

[8] H. Mori and A. Yuihara, “Deterministic Annealing Clustering for ANN-Based Short-Term Load Forecasting,” Trans of IEEE on Power Systems, Vol.16, No.3, pp.545-551, Aug. 2001.

[9] H. Mori and A. Awata, "Data Mining of Electricity Price Forecasting with Regression Tree and Normalized Radial Basis Function Network," Proc. of 2007 IEEE International Conference on Systems, Man and Cybernetics (SMC2007) (CD-ROM), pp.3743-3748, Montreal, Canada, Oct. 2007.

[10] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, “Classification and Regression Trees,” Wadsworth International Group, Belmont, California, USA, 1984

[11] H. Mori, Y. Sakatani, T. Fujino, and K. Numa, “An Efficient Hybrid Method of Regression Tree and Fuzzy Inference for Short-term Load Forecasting in Electric Power Systems,” Proc. of RASC 2002, pp. 1-6, Nottingham, UK, Dec. 2002.

[12] H. Mori and N. Kosemura, "Optimal Regression Tree Based Rule Discovery for Short-term Load Forecasting," Proc. of 2001 IEEE PES Winter Meeting, Vol. 2, pp.421-426, Columbus, USA, Jan. 2001.

[13] H. Mori, and H. Kobayashi: “A Fuzzy Neural Net Short-Term Load Forecasting,” Proc of ISAP’94, Vol.11, pp. 775-782, Monpellier, France, Feb. 1996.

[14] S. Chakrabarti and B. Jeyasurya, “On-line Voltage Stability Monitoring Using Artificial Neural Network”, Proc of IEEE Large Engineering System Conference on Power Engineering 2004, LESCOPE2004, pp. 71-75, Jul. 2004.

[15] H. Mori and D. Iwashita, "ANN-Based Risk Assessment for Short-Term Load Forecasting," Proc. of ISAP'05, pp.446-451, Washington D.C., U.S.A, Nov. 2005.

[16] H. Mori and Y. Yamada, "An Efficient Multi-objective Meta-heuristic Method for Distribution Network Expansion Planning," Proc. of IEEE PES Power Tech2007 (CD-ROM), Lausanne, July 2007.