International Journal of Fuzzy Logic and Intelligent Systems, vol. 5, no. 2, June 2005 pp. 95-101

A New Learning Algorithm of Neuro-Fuzzy Modeling Using Self-Constructed Clustering

Jeong-Woong Ryu, Chang-Kyu Song, Sung-Suk Kim, Sung-Soo Kim

School of Electrical & Computer Engineering, Chungbuk National University

Abstract

In this paper, we proposed a learning algorithm for the neuro-fuzzy modeling using a learning rule to adapt clustering. The proposed algorithm includes the data partition, assigning the rule into the process of partition, and optimizing the parameters using predetermined threshold value in self-constructing algorithm. In order to improve the clustering, the learning method of neuro-fuzzy model is extended and the learning scheme has been modified such that the learning of overall model is extended based on the error-derivative learning. The effect of the proposed method is presented using simulation compare with previous ones

Key Words : Clustering, Neuro-Fuzzy Modeling, TSK Fuzzy Model, Self-Constructed Clustering, System Identification

1. Introduction

Basically, the modeling of neuro-fuzzy are consisted of the structure identification and the parameters identification[1][2]. In the structure identification[2], the type of model and number of fuzzy rules are determined, whereas in the parameter identification, the learning algorithm is optimized by adjusting the parameters of models. The number and the type of membership functions are important factors that determine the parameters of the structure, such as particular characteristics[3][4].

General method adopts the process of partitioning the input space in grid type, and it uses all surface of input dimension to connect relevant neuro-fuzzy rules. In this case, the grid partition method has an advantage that covers all input space, however it invokes the increasing dimension of input space and number of membership functions. As a result, the neuro-fuzzy rules are increased exponentially by connecting premise part and rule base[2]. The problem of grid partition is solved by using the clustering method which assigns the neuro-fuzzy rules to the space with data not to the empty space and using a new approach that counts only the number of associated number of functions without counting on the dimension of input[2,3,4]. The clustering methods are classified into two groups. The first method determines the number of clusters without information on the shape of density function. The second method estimates the parameters involved in the shape of density function with the knowledge on the number of clusters. In the first case, the data density is generally used estimate the number of clusters using density function, such that the process of optimization may be missed. Also, when the number of clusters is determined, the algorithm performs well when the number of clusters and the number of patterns. However, when the number of clusters fails to represent the whole data, the algorithm's performance cannot be guaranteed.

In the modeling of neuro-fuzzy system, the clustering methods are used for the initial parameters estimation of the model[3,5]. In this case, the clustering algorithm and the neuro-fuzzy modeling are used in sequence. The input space partition optimization by clustering becomes ineffective when the parameters involved in the neuro-fuzzy modeling are changed. Basic learning method of the general neuro-fuzzy modeling is based on updating the membership function of premise part using the error-derivate of model output.

In the paper, we propose a new learning algorithm that simultaneously inferences a number of cluster and optimizes these parameters using the membership function of Takagi-Sugeno-Kang fuzzy model in the process of learning. Also, we extend the supervised learning to the unsupervised learning with the clustering method that is using the feedback of error in every step of the overall model. In the learning of overall model, a proposed algorithm extended the learning scheme to the clustering method based learning from the error-derivative based ones, such that self constructed the neuro-fuzzy rules for optimize these parameters. Finally, we apply the proposed method to the benchmark problems and obtain better results than previous ones.

2. Proposed Clustering Algorithm

Basic concept of clustering is grouping the data that includes the clusters with high similarity[2][6]. For this purpose, clustering algorithm uses a function of similarity measure. Proposed method is use the generalized function of references[2,3]

Manuscript received May. 10, 2005; revised May. 31, 2005.
This work was supported by the research grant of the Chungbuk National University in 2005.

$$r_{ij} = \exp\left(-\frac{1}{2}(x_i - c_j)^T \Sigma_j^{-1}(x_i - c_j)\right) \tag{1}$$

where c_j and Σ_j are the mean and the covariance of jth cluster. The number of data is represented by $i = 1, 2, \cdots, N$ and the number of clustering is presented by $j = 1, 2, \cdots, c$. The generalized function has the membership as 1 when data are similar, and decreases exponentially when data are away from the mean. The data groups which are not affected by the proposed method still effective measuring the similarities between groups through cluster parameter estimation. So, we use the constrained magnitude of similarity using the predetermined threshold value ζ.

$$r_{ij} = \begin{cases} r_{ij}, & \text{if } r_{ij} > \zeta \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In this case, by manipulating ζ, the results of clustering change. When ζ is large, it eliminates the information of the current clusters and when ζ is small, ζ has no effect in the algorithm. So, in order to solve this problem, we modified the function of similarity measure from the equation (1) to equation (3) such as

$$r_{ij} = \exp\left(-\frac{1}{2}(x_i - c_j)(\zeta \cdot \Sigma_j)^{-1}(x_i - c_j)\right) \tag{3}$$

Adding the conditions of similarity constraint and the function of similarity measure in inverse type to the system, the clustering results are stabilized when sudden changes of threshold occur. In the Fig. 1, the range of similarity measure is showed as changing the predetermined threshold ζ.

$$c_j = \frac{\sum_{i=1}^{N} r_{ij} x_i}{\sum_{i=1}^{N} r_{ij}} \tag{4}$$

After inferencing the mean of cluster, the covariance Σ_j is inferred using Maximum Likelihood Estimation (MLE)[7][8] with the prior probability as follows.

$$\Pr(x_i) = \frac{1}{N} \sum_{i=1}^{N} u_{ij} \tag{5}$$

where u_{ij} is a element of partition matrix U in the j-th cluster after the normalization step followed by the product step with prior probability in equation (6) and (7).

$$p_{ij} = \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \times \exp\left(\frac{1}{2}(x_i - c_j)^T \Sigma_j^{-1}(x_i - c_j)\right) \tag{6}$$

$$u_{ij} = \frac{p_{ij} \Pr(x_i)}{\sum_{i=1}^{N} p_{ij} \Pr(x_i)} \tag{7}$$

The initial number of clusters for the proposed method is equal to the number of data patterns. In the processing the algorithm, cluster parameters converge to the high densities of data space. And we introduce the modified Subtractive clustering[2,12] in the learning step to solve that performed the converging speed, difficulty of converging in the case of large data set. In the algorithm, the modified subtractive algorithm makes a meaningful cluster using cumulative densities of the inferred cluster only the meaningful cluster is represented in equation(8).

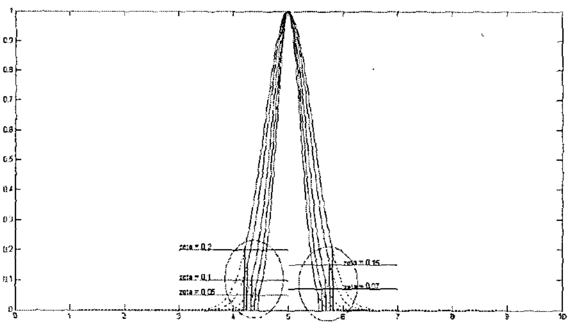

Fig. 1. Range of similarity when the mean of cluster is Eq. (4)

$$c_{new} = c_j, \quad \text{if } D_j = \max\left(\sum_{i=1}^{N} u_{ij}\right) \tag{8}$$

Using the mean that infers clusters, we eliminate the similarities in algorithm as shown in equation (9) and then the neighbour clusters are eliminated.

$$s_{i,new} = \exp\left(-\frac{1}{2}(c_i - c_{new})^T \Sigma_{new}(c_i - c_{new})\right) \tag{9}$$

On protect the eliminate with unnecessary clusters, we introduce convergent limit value ε as equation (2), algorithm can re-calculate as equation (10).

$$s_{i,new} = \begin{cases} 0, & \text{if } s_{i,new} < \varepsilon \\ s_{i,new}, & \text{otherwise} \end{cases} \tag{10}$$

Fig. 2 shows the results of effects that eliminate at limit ε to 0.9

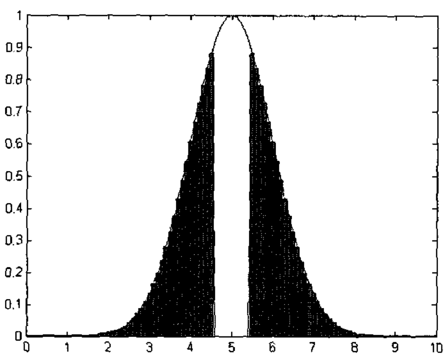

Fig. 2. Effect of ε

After this step, the centers of new meaningful clusters are inferred using equation (11).

$$c_{new}' = c_j', \quad \text{if } D_j' = \max\left(D_{new} - \sum_{i=1}^{N}\left(\sum_{j=1}^{c} s_{i,j}\right)\right) \tag{11}$$

Procedure from equation (8) to (11) repeats until termination conditions are satisfied. The clustering algorithm learns one epoch then checks the termination condition of overall clustering algorithm after above steps.

3. Proposed Neuro-Fuzzy Model and Learning (Modeling)

We use Tagaki-Sugeno-Kang(TSK) fuzzy model[2] which has the properties that linguistic inputs and polynomial output show its model from layer 1 to layer 5 in figure 3.

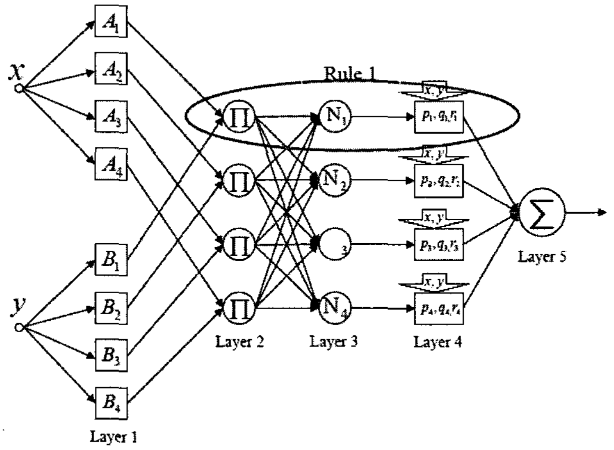

Fig. 3. TSK Fuzzy Inference System

We use the model whose outputs organize linear equation type and have the de-fuzzifier step. The model has compact structure and also has advantage to obtain parameters using least mean square method. If the model is organized with the two dimensional inputs and one dimensional output, the relation between input and output is described by the i-th rule represents as (12).

$$R^i : \text{IF } x \text{ is } A_1 \text{ and } y \text{ is } B_2 \quad \text{THEN } f_i = p_i x + q_i y + r_i \tag{12}$$

Fig. 4 shows the concept of the input-output relation of TSK fuzzy models.

The overall output of model is inferred by weighted average using each rule as (13)

$$f = \frac{\omega_1 f_1 + \omega_2 f_2 + \ldots + \omega_c f_c}{\omega_1 + \omega_2 + \ldots + \omega_c} = \overline{\omega_1} f_1 + \overline{\omega_2} f_2 + \ldots + \overline{\omega_c} f_c \tag{13}$$

where ω_i is the weights of layer 3 which is the product of membership of fuzzified input and each rules. And membership function of model is represented in (14).

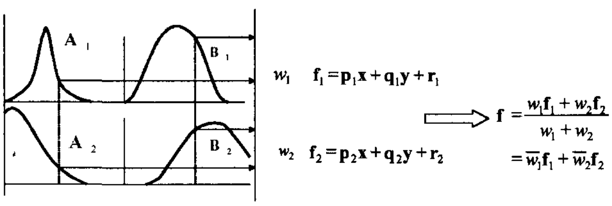

Fig. 4. Concept of TSK Fuzzy Inference System

$$\mu_i = \exp\left(-\frac{1}{2}\frac{(x-c_i)^2}{\sigma^2}\right) \tag{14}$$

Cluster parameters in (14) through self-constructing clustering are directly used in the neuro-fuzzy model as the same clustering parameters and membership function of model. Also overall inferred output are summation of (13). In detail, each $\overline{\omega_i} f_i$ is considered as the independent outputs of sub-model and the performance of clustering. The errors of each fuzzy sub-model are show as (15).

$$e_{j,nerofuzzy} = y - \overline{\omega_i} f_i \tag{15}$$

Using (15), we construct the error weighting function ef_{ij} as (16).

$$ef_{ij} = \exp\left(-\frac{1}{2}\frac{(e_{j,nerofuzzy} - \mu_{ei})^2}{\sigma_{ci}^2}\right) \tag{16}$$

The similarity between clusters are improved by altering from (3) to (17).

$$r_{ij} = r_{ij}(ef_{ij} \cdot (1-t) + t) \tag{17}$$

In the overall model, the self-constructing clustering are used to estimate the number of clusters and optimize the parameters in learning step. The inferred parameters directly used to the neuro-fuzzy model and to learn the consequent parameters by least square method. In the continuous learning process, the proposed model is constructed of the structure identification and the parameter identification simultaneously. We extended the proposed model by clustering based learning instead of the general error derivative based learning as shown in Figure 5.

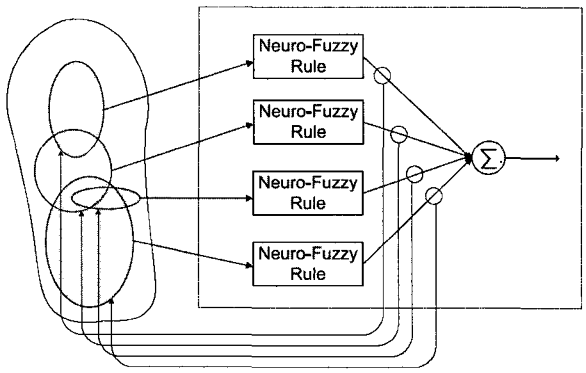

Fig. 5. Proposed structure of learning model

4. Simulation Results

4.1 Example 1 : Box-Jenkins Gas Furnace

Box and Jenkins gas furnace data[2] is conventionally nonlinear time series experiments at sampling with each 9[s] and composed input-output data pairs with 296 data sets. The input is flow of methane and output is the rate of CO_2 and the given data set is composed as $y(t-1)$, $y(t-2)$, $y(t-3)$, $y(t-4)$, $u(t-1)$, $u(t-2)$, $u(t-3)$, $u(t-4)$, $u(t-5)$, $u(t-6)$ at same times.

Using the Jang's input selection method, we choose the training and checking data $y(k+1)=f(y(k),u(k-3))$. Also, we select the training data set from the odd number of data pairs and use the checking data the even number of data set of the whole data. Using the input-output relation, we constructed the TSK neuro-fuzzy model as (18).

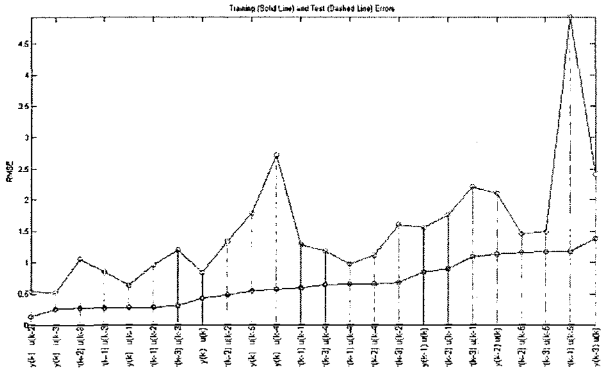

Fig. 6. Input Selection of Jang's method

$$\text{IF } u(k-3) \text{ is } A_i \text{ and } y(k) \text{ is } B_i$$
$$\text{THEN } y(k+1)=p_i u(k-3)+q_i y(k)+r_i \quad (18)$$

(1) Clustering Results

Firstly we set the predetermined threshold ζ to 0.1 and the consequent similarity t to 0.2 and the learning 50th epoch. In this learning, the learned model has 14 clusters and the least mean squared error of training(learning) step 0.1037 and the error of checking step 0.1697. Fig. 7 shows the results of clustering after training steps in premise part.

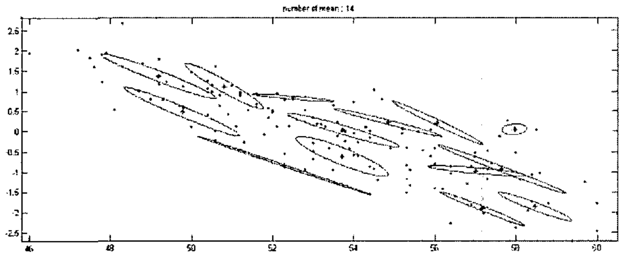

Fig. 7. Clustering Results

(2) Results of clustering based neuro-fuzzy modeing

We show the membership functions in Fig. 8 and Fig. 9.

In Fig. 10 and 11, we show the output of training data and the checking data. As following Fig. inferred outputs are very similar to the given output data.

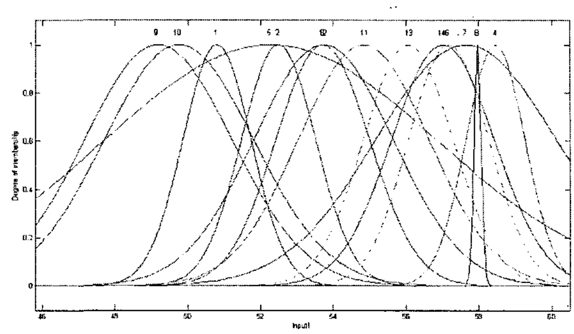

Fig. 8. Membership function of $y(k)$ in Box-Jenkins gas furnace data

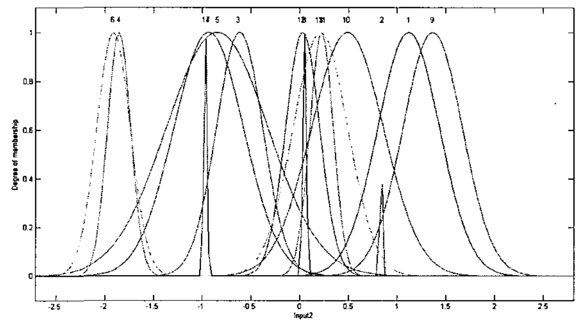

Fig. 9. Membership function of $u(k-3)$ in Box-Jenkins gas furnace data

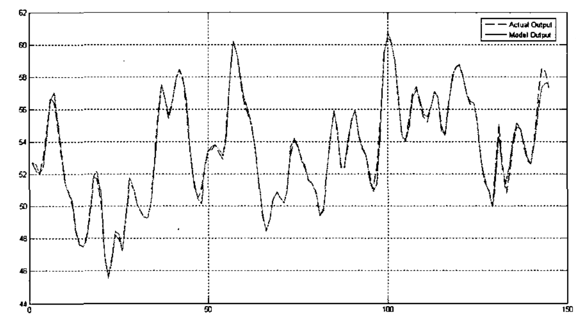

Fig. 10. Compare with actual output and model output of training data

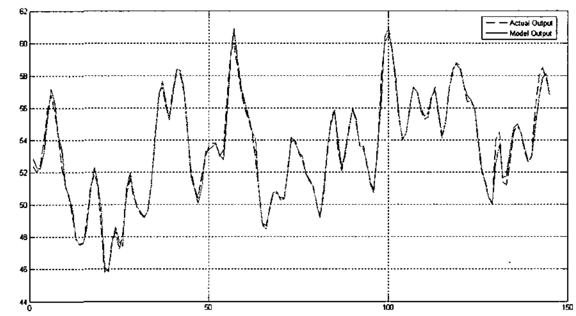

Fig. 11. Compare with actual output and model output of checking data

Also the experimental results of the changes of a predetermined threshold ζ or the consequent similarity t, show the comparison with the previous ones presented in Table 1. Table 1. Compares the results with previous one using RMSE. As shown in Table 1, the proposed system reduces not only MSE with training data, but also MSE with checking data.

Table 2. Compare the results with previous one using RMSE

Method \ Error(RMSE)	rules	Training error	Checking error	Remark
Pedrycz[10]	81	0.320	None	
Xu[10]	25	0.328	None	
Sugeno[10]	2	0.359	None	
Abonyi[10]	16	0.154	None	
Babuska[10]	23	0.124	None	
Jang[6]	4	0.135	0.530	
Oh[18]	4	0.026	0.272	
Proposed ($\zeta=0.2$, $t=0.4$)	9	0.1282	0.1863	50 epoch
Proposed ($\zeta=0.2$, $t=0.4$)	5	0.1267	0.1523	100 epoch
Proposed($\zeta=0.1$, $t=0.2$)	14	0.1037	0.1697	50 epoch

4.2 Example 2 : Iris flower Classification

Fisher's iris classification problem is conventional benchmark problem in pattern recognition. Data set is composed of four inputs and one output for three patterns and is set as the 150th data. Each input four characteristics are sepal length (SL), sepal width (SW), petal length(PL), and petal width(PW). Output is also consisted of three classes iris setosa (class 1), iris veresicolor (class 2), and iris virginica (class 3). Each class has the 50 data sets and the first class linearly derives other two classes easily, but the second class and the last class can not be derived linearly. The input-output relation is represented in (19).

$$y(\text{class}) = f(\text{SL}, \text{SW}, \text{PL}, \text{PW}) \tag{19}$$

(1) Results of clustering

First, we set the predetermined threshold ζ to 0.9 and the *consequent similarity* t *to 0.2 and the learning 100th epoch.* The results yield 5 clusters after learning and Fig. 12 displays with data sets.

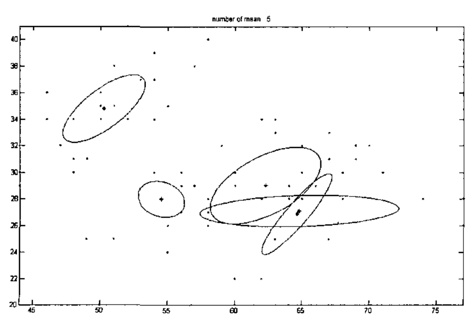

Fig. 12. Clustering result in iris data

(2) Results of clustering based on the neuro-fuzzy modeling

Fig. 13 and 14 show the membership functions of the SW and the PW in the after learned model.

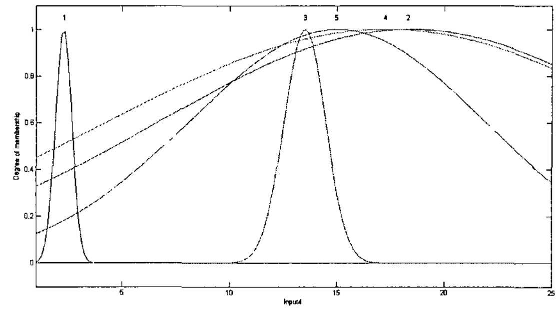

Fig. 13. Membership function of PW in iris data

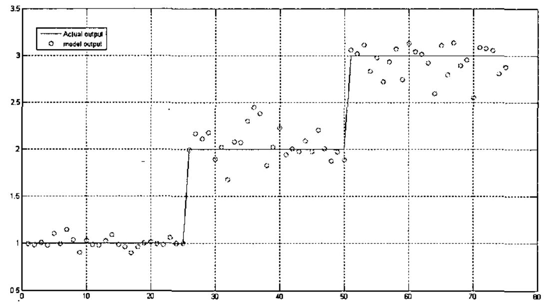

Fig. 14. Compare with actual output and model output of training data

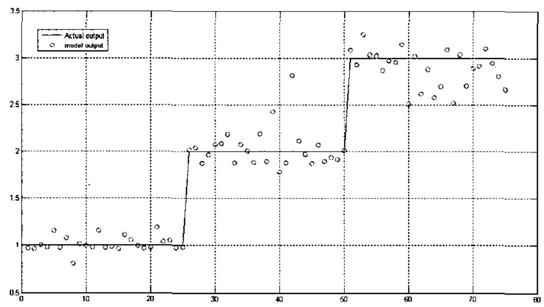

Fig. 15. Compare with actual output and model output of checking data

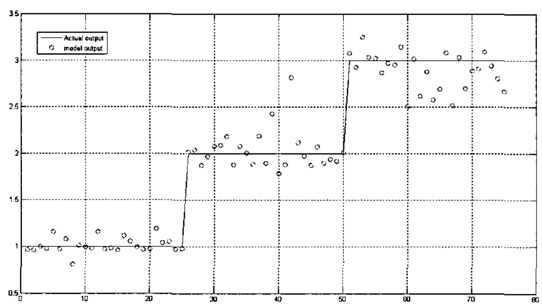

Fig. 16. Compare with actual output and model output of checking data

Also, the results in Table 2 describe the case that the predetermined threshold ζ and the consequent similarity t are varied.

Table 3. Results of classification in iris data

Error(MSE)	Number of rules	Recognition rate
T.P. Hong, C.Y. Lee[12]		95.570
C.H. Chang, S.M. Chen[13]		96.070
T.P. Wu, S.M. Chen[14]		96.210
F.M. Tsai, S.M. Chen[15]		95.833
J.L. Csstro, J.J. Castro-Schez, J.M. Zurita[16]		96.600
T.P. Hong, J.B. Chen[17]		97.333
Proposed($\zeta=0.05$, $t=0.5$)	8	97.333
Proposed($\zeta=0.05$, $t=0.2$)	6	97.333
Proposed($\zeta=0.9$, $t=0.2$)	5	98.67

In the iris classification, the optimized model has no error in classifying the training data but has some error in the checking process, some papers have reported both classified with no error.

5. Conclusions

In this paper, we proposed a new leaning scheme for the neuro-fuzzy model with self-constructing clustering method. Proposed model extends the learning method from with the error derivative for a basis to with the clustering for a basis, and automatically identifies the structures and parameters of the neuro-fuzzy model simultaneously. We extended the learning method of clustering from the unsupervised learning to the supervised learning using the input-output relation as a neuro-fuzzy model. Through simulation, proposed method have the benefits comparing with the previous ones. Next, we can extend the proposed method in a way to reduce computational load, hence we modify the structure of neuro-fuzzy model for improving performance of overall model

References

[1] Lin C.T. and C.S.G. Lee, *Neural Fuzzy Systems : A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall, 1996.

[2] Jang J.S.R., C.T. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing : A Computational Approach to learning and Machine Intelligence*, Prentice Hall, 1997.

[3] Kim S.S., S.S. Kim and J.W. Ryu, "Advanced Neuro-Fuzzy Modeling applied New Clustering Algorithm", *Trans on KIEE*, Vol. 53D, No. 7, pp. 536-543, 2004.

[4] Kim S.S., K.C. Kwak, J.W. Ryu and M.K. June, "Neuro-Fuzzy System Modeling Using GMM and Clustering Method", *Trans of KFIS*, Vol. 12, No. 6, pp. 571-576, 2002.

[5] Jang J.S.R., "ANFIS : Adaptive Network-based Fuzzy Inference System", *IEEE Trans on System, Man, and Cybernetics*, Vol.23, No. 3, pp. 665-685, 1993.

[6] Wong C.C. and C.C. Chen, "A Hybrid Clustering and Gradient Descent Approach for Fuzzy Modeling", *IEEE Trans on Systems, Man, and Cybernetics-Part B : Cybernetics*, Vol. 29, No.6, pp. 686-693, 1999.

[7] Xuan G., W. Zhang, P. Chai, "EM algorithm of Gaussian Mixture Model and Hidden Markov Model", *Image Processing Proceedings, International Conference on*, Vol. 1, pp. 145-148. 2001.

[8] Bilms J., "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," *ICSI TR-97-021*, April 1998.

[9] Yager R.R. and D.P. Filev, "Generation of Fuzzy Rules by Mountain Clustering", *Jounal of Intelligent and Fuzzy System*, Vol.2, pp. 209-219, 1994.

[10] Abonyi J., Robert Babuska, Ferenc Szeifert," Fuzzy Modeling With Multivariate Membership Functions : Gray-Box Identification and Control Design", *IEEE Trans on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 31, No. 5, pp. 755-767, 2001.

[11] Oh S.K. and W. Pedrycz, "Identification of Fuzzy System by Means of an Auto-Tuning Algorithm and Its Application to Nonlinear Systems", Fuzzy Sets and Systems, Vol. 115, pp. 205-230, 2000.

[12] Hong T.P. and C.Y. Lee, "Induction of Fuzzy Rules and Membership Functions from Training Example", *Fuzzy Sets and Systems*, Vol. 84, No. 1, pp. 33-47, 1996.

[13] Chang C.H. and S.M. Chen, "Constructing Membership Functions and Generating Weighted Fuzzy Rules from Training Data", *Proceedings of the 2001 Ninth National Conference on Fuzzy Theory and Its Applications*, Chungli, Taoyuan, Taiwan, Republic of China, pp. 708-713, 2001.

[14] Wu T.P. and S.M. Chen, "A New Method for Constructing Membership Functions and Fuzzy Rules from Training Examples", *IEEE Trans on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 29, No. 1, pp. 24-50, 1999.

[15] Tsai F.M and S.M. Chen, "A New Method for Constructing Membership Functions and Generating Fuzzy Rules for Fuzzy Classification Systems", *Proceeding of 2002 Tenth National Conference on Fuzzy Theory and Its Application*, Hsinchu, Taiwan, Republic of China, 2002.

[16] Csstro J.L, J.J. Castro-Schez and J. M. Zurita, "Learing Maximal Structure Rules in Fuzzy Logic for Knowledge Acquisition in Expert Systems", *Fuzzy Sets and Systems*, Vol. 101, No. 3, pp. 331-342, 1999.

[17] Hong T.P. and J.B. Chen, " Processing Individual Fuzzy Attributes for Fuzzy Rule Induction", *Fuzzy Sets and Systems*, Vol. 112, No. 1, pp. 127-140, 2000.

Jeong-Woong Ryu
He received the B.S. degree in Dept. of Electrical Engineering from Hanyang University, Seoul, Korea, in 1965. He received the M.S. degree in Dept. of Electrical Engineering from Dankook University, Seoul, Korea, in 1976. He received the Ph.D. degree in Dept. of Electronics Engineering from Dankook University, Seoul, Korea, in 1987. He had worked Daejeon Industrial College from 1969 to 1979, where he had been an associate professor. Since 1979, he is a professor in School of Electrical & Computer Engineering, Chungbuk National University, Cheongju, Korea. His research interests include robust control theory, PID, QFT, intelligence system and fuzzy theory.

Chang-Kyu Song
He received the B.S. and M.S. degree in Dept. of Electrical Engineering from Chungbuk National University, Cheongju, Korea, in 1995 and 1997, respectively. He is currently in the doctor's course at the same graduate school. His research interests include image processing, image compression, fuzzy theory, and intelligence system.

Sung-Suk Kim
He received M.S. degree in Dept. of Electrical Engineering from Chungbuk National University, Cheongju, Korea, in 2002 and 2004, respectively. He is currently in the doctor's course at the same graduate school. His research interests include Intelligent Control, Neuro-Fuzzy system, and its applications.

Sung-Soo Kim
He received the B.S. degree from Chungbuk National University, Chungbuk, Korea in 1983, the M. S. degree from University of Arkansas-Fayetteville in 1989, and the Ph. D. degree from University of Central Florida in 1997. At present, he is an Associate Professor at the Department of Electrical and Computer Engineering, Chungbuk National University, since 2001. His current research interests include Neural Networks, Fuzzy Theory, Artificial Intelligence, Intelligence Systems.