Enhanced RBF Network by Using Auto-Turning Method of Learning Rate, Momentum and ART2

Kwang-baek Kim [a] and Jung-wook Moon [b]

[a] *Dept. of Computer Engineering, Silla University*
San 1-1, Kwaebop-dong, Sasang-gu, Busan 617-736, Korea
Tel: +82-51-309-5052, Fax: +82-51-309-5652, E-mail: gbkim@silla.ac.kr

[b] *Dept. of Computer Engineering, Busan National University*
San 30, Jangjung-dong, Kumjung-gu, Busan 609-735, Korea
Tel: +82-51-510-2876, Fax: +82-51-515-2208, E-mail: jwmoon@isel.cs.pusan.ac.kr

Abstract

This paper proposes the enhanced RBF network, which arbitrates learning rate and momentum dynamically by using the fuzzy system, to arbitrate the connected weight effectively between the middle layer of RBF network and the output layer of RBF network. ART2 is applied to as the learning structure between the input layer and the middle layer and the proposed auto-turning method of arbitrating the learning rate as the method of arbitrating the connected weight between the middle layer and the output layer. The enhancement of proposed method in terms of learning speed and convergence is verified as a result of comparing it with the conventional delta-bar-delta algorithm and the RBF network on the basis of the ART2 to evaluate the efficiency of learning of the proposed method.

Keywords:

RBF Network; Learning Rate; ART2; Delta-bar-delta.

1. Introduction

The study of improving the learning time and the generalization ability of the neural network learning algorithm has been performed. As a result of it RBF(Radial Basis Function), which has been used for multivariate analysis and interpolation of statistics, is being used for organizing the neural network model by Brommhead and Low. Then RBF network is proposed now [1]. RBF network, which has the characteristics of quick learning time, generalization and simplification, is being applied to classifying learning data and modeling nonlinear system. RBF network is a feed-forward neural network which is made of input layer, middle layer and output layer. Different algorithm can be applied to each layer because different works are performed in each layer and it can be organized with separated optimization by each layer [2]. The composition of layers is classified to three types. The first type is fixed centers selected at random which has the system of selecting the node of the middle layer randomly from learning data set. The second one is self-organized selection of centers which decides the middle layer according to the form of self-organization and applying the supervised learning to the output layer. The last one is supervised selection of centers which uses the supervised learning for the middle layer and the output layer. Therefore the middle layer of RBF network has a role of clustering. In other words the purpose of this layer is classifying given data set to homogeneous clusters. Homogeneous means that it classifies input data to the same cluster when they are in the fixed radius after the distance is measured between each vector within a cluster in the characteristic space for the input data and otherwise to the different cluster. Clustering within the fixed radius, however, has the weak point of selecting wrong clusters. Therefore the decision of the middle layer has big influence on the overall efficiency of RBF network [3]. Hence enhanced RBF network, which uses ART2 to organize the middle layer efficiently and applies learning rate and the auto-turning method of arbitrating momentum using the fuzzy logic system to the arbitration of the connected weight between the middle layer and the output layer, is proposed in this paper.

2. Related Study

2.1. delta-bar-delta Algorithm

Delta-bar-delta algorithm[4], which improved the quality of backpropagation algorithm, enhances learning quality by arbitrating the learning rate dynamically for each connected weight. It is made of delta and delta-bar to arbitrate learning rate dynamically.

The formula of making delta is as follows. In this expression i means the input layer, j means the middle layer and k means the output layer.

$$\Delta_{ji} = \frac{\sigma E}{\sigma w_{ji}} = -\delta_i x_i \tag{1}$$

$$\Delta_{kj} = \frac{\sigma E}{\sigma w_{kj}} = -\delta_k z_j \tag{2}$$

The formula of making delta-bar is as following.

$$\bar{\Delta}_{ji}(t) = (1-\beta)\cdot\bar{\Delta}_{ji}(t) + \beta\cdot\bar{\Delta}_{ji}(t-1) \tag{3}$$

$$\bar{\Delta}_{kj}(t) = (1-\beta)\cdot\bar{\Delta}_{kj}(t) + \beta\cdot\bar{\Delta}_{kj}(t-1) \tag{4}$$

The value of parameter β in formula (4) is fixed constant between 0 and 1. The change of learning rate for that of delta and delta-bar is as follows. If the change of the connected weight is performed to the same direction in the process of continual learning, the learning rate will increase. At this point delta and delta-bar happen at the same sign and the learning rate increases. Besides if the sign of delta is opposite to that of delta-bar, the learning rate will decrease with the ratio of 1-γ for the present value. The formula of the variable learning rate for each layer is as follows.

$$\begin{aligned} \alpha_{ji}(t+1) &= \alpha_{ji}(t) + k && if\ \bar{\Delta}_{ji}(t-1)\cdot\Delta_{ji}(t) > 0 \\ &= (1-\gamma)\cdot\alpha_{ji}(t) && if\ \bar{\Delta}_{ji}(t-1)\cdot\Delta_{ji}(t) < 0 \\ &= \alpha_{ji}(t) && if\ \bar{\Delta}_{ji}(t-1)\cdot\Delta_{ji}(t) = 0 \end{aligned} \tag{5}$$

$$\begin{aligned} \alpha_{kj}(t+1) &= \alpha_{kj}(t) + k && if\ \bar{\Delta}_{kj}(t-1)\cdot\Delta_{kj}(t) > 0 \\ &= (1-\gamma)\cdot\alpha_{kj}(t) && if\ \bar{\Delta}_{kj}(t-1)\cdot\Delta_{kj}(t) < 0 \\ &= \alpha_{kj}(t) && if\ \bar{\Delta}_{kj}(t-1)\cdot\Delta_{kj}(t) = 0 \end{aligned} \tag{6}$$

2.2 RBF network on the basis of ART2

RBF network on the basis of ART2 algorithm is divided to two stages. Competitive learning is applied to the first stage as the learning structure between the input layer and the middle layer. And that supervised learning is performed between the middle layer and the output layer [5, 6]. Output vector of the middle layer in RBF network on the basis of ART2 is calculated by formula (7) and the node having the least output vector like in formula becomes (8) becomes the winner node.

$$O_j = \sum_{i=1}^{N}\left(\left|x_i - w_{ji}(t)\right|\right) \tag{7}$$

$$O_j^* = Min\{O_j\} \tag{8}$$

where $w_{ij}(t)$ is connected weight value between the input layer and the middle layer.

The winning node of the middle layer in RBF network on the basis of ART2 is the value which has the least difference from the input vector to the output vector of the middle layer and formula (9) is showing the similarity test for the winner node.

$$O_j^* < \rho \tag{9}$$

ρ is vigilance parameter in the formula. It is classified to the same pattern when the output vector is smaller than the vigilance parameter and the different pattern otherwise. The connected in weight is arbitrated to make the homogeneous characteristics of the input pattern reflected on the connected weight when it is classified to the same pattern. The arbitration of ART2 algorithm is as follows.

$$w_{j^*i}(t+1) = \frac{w_{j^*i}(t)\cdot u_n + x_i}{u_n + 1} \tag{10}$$

where u_n means the number of updated patterns in generated clusters. The output vector of the middle layer is normalized by formula (11) and applied to as the input vector of the output layer.

$$z_i = 1 - \frac{O_j}{N} \tag{11}$$

The output vector of the output layer is calculated by formula (12).

$$O_k = f\left(\sum_{j=1}^{M} w_{kj}\cdot z_j\right) \tag{12}$$

$$f(x) = \frac{1}{1+e^{-x}} \tag{13}$$

Error value and error signal value are calculated by comparing the output vector with the target vector. The connected weight is arbitrated by the same way.

$$\delta_k = (T_k - O_k)\cdot O_k\cdot(1-O_k) \tag{14}$$

$$w_{kj}(t+1) = w_{kj}(t) + \alpha\cdot\delta_k\cdot z_j \tag{15}$$

3. Enhanced RBF network

Enhanced RBF network applies ART2 to the learning structure between the input layer and the middle layer and proposed auto-turning method of arbitrating learning rate to the method of arbitrating the connected weight between the middle layer and the output layer. When the absolute value of the difference from the output vector to the target vector for each pattern is below 0.1, it is classified to the accuracy and otherwise to the inaccuracy. Learning rate and momentum are arbitrated dynamically by applying the number of the accuracy and the inaccuracy to the input of the fuzzy logic system.

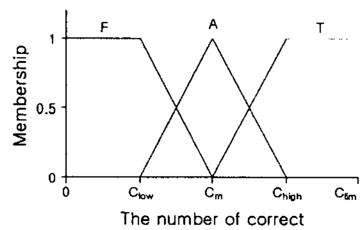

Figure 1 - The membership function to which the accuracy

belongs

Figure 1 is showing the membership function to which the accuracy belongs whereas Figure 2 is showing the membership function to which the inaccuracy belongs.

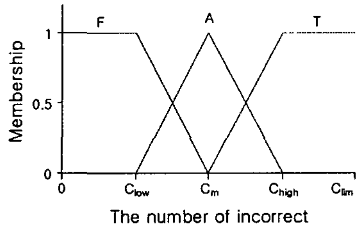

Figure 2 - The membership function to which the inaccuracy belongs

The values of C_{low} and C_{high} are calculated by formula (16) and (17).

$$C_{low} = \log_2(N_i + N_p) \tag{16}$$

$$\begin{pmatrix} N_i : \textit{the number of input nodes} \\ N_p : \textit{the number of patterns} \end{pmatrix}$$

$$C_{high} = C_{lim} + C_{low} \tag{17}$$

F means false, A means average and T means true in Figure 2 and 3 as membership functions. When the rule of controlling fuzzy to arbitrate learning rate is expressed with the form of $if \sim then$, it is as follows.

$R_1 : If\ correct\ is\ F, incorrect\ F\ Then\ \alpha\ is\ B$

$R_2 : If\ correct\ is\ F, incorrect\ A\ Then\ \alpha\ is\ B$

$R_3 : If\ correct\ is\ F, incorrect\ T\ Then\ \alpha\ is\ B$

$R_4 : If\ correct\ is\ A, incorrect\ F\ Then\ \alpha\ is\ M$

$R_5 : If\ correct\ is\ A, incorrect\ A\ Then\ \alpha\ is\ M$

$R_6 : If\ correct\ is\ A, incorrect\ T\ Then\ \alpha\ is\ M$

$R_7 : If\ correct\ is\ T, incorrect\ F\ Then\ \alpha\ is\ S$

$R_8 : If\ correct\ is\ T, incorrect\ A\ Then\ \alpha\ is\ S$

$R_9 : If\ correct\ is\ T, incorrect\ T\ Then\ \alpha\ is\ S$

Figure 3 is showing the output membership function calculating the learning rate, which is going to be applied to learning.

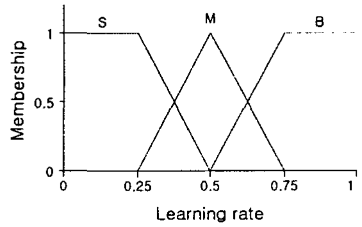

Figure 3 - The membership function of learning rate

S means small, M means medium and B means big in Figure 3 as membership functions. Membership degrees of accuracy and inaccuracy for each membership function are calculated when the accuracy and the inaccuracy are decided as the input value of the fuzzy logic system. The rule of controlling fuzzy is applied to when the membership degree for each membership function. Besides Max_Min method is used for inference. The learning rate, which is going to be used for learning, is calculated by defuzzifier method after the fuzzy inference. Formula (18) is showing the center of gravity, which is used for the defuzzification [7].

$$\alpha = -\frac{\sum \mu(y) \cdot y}{\sum y} \tag{18}$$

Momentum is calculated by formula (19).

$$\mu = \zeta - \alpha \tag{19}$$

ζ is the parameter between 1 and 1.5, which is given empirically.

4. Experiment and analysis of the result

We implemented the experiment with C++ Builder 6.0 on IBM compatible PC in which Intel Pentium-IV CPU and 256MB RAM are mounted.

We analyzed the number of epoch and the convergence by applying 136 number patterns having 10×10 in size, which is abstracted from the citizen registration card, to conventional delta-bar-delta method, RBF network on the basis of ART2 and the learning algorithm proposed in this paper. Figure 4 is showing the pattern of numbers which was used for learning and Table 1 is showing target vectors.

Figure 4 - Pattern of numbers

Table 1 - Target vectors which was used for learning

	0	1	2	3	4	5	6	7	8	9
Target value	0	0	0	0	0	0	0	0	1	1
	0	0	0	0	1	1	1	1	0	0
	0	0	1	1	0	0	1	1	0	0
	0	1	0	1	0	1	0	1	0	1

Table 2 is showing the parameter of each algorithm which was used for the experiment and table 3 is showing the result of learning.

Table 2 - Parameters which were used for learning

Parameter / Learning method	α	ρ	ζ	β	κ	γ
Delta-bar-delta	0.5			0.7	0.005	0.2

RBF network on the basis of ART2	0.5	10				
Proposed method		10	1.0			

means learning rate, ρ means the vigilance parameter of ART2, ζ means the parameter to calculate momentum and , κ, γ means parameters which were fixed by delta-bar-delta algorithm.

The experiment have been performed 10 times under the standard of classifying it to the accuracy when the absolute value of the difference from the actual output vector to the target vector is below ε ($\varepsilon \leq 0.1$) in performing epoch 10000 times.

Table 3 - Comparison the convergence among each algorithm

Result of learning / Learning Method	The number of experiment	The number of success	The number of average epoch
Delta-bar-delta	10	2	2793
RBF network on the basis of ART2	10	10	2710
Proposed method	10	10	1464

The fact that the proposed method is more enhanced than conventional methods in terms of learning speed and convergence is verified in table 3. Moreover the proposed method did not react sensitively to the times of learning and the convergence whereas conventional methods did. As a result of it the efficiency of learning is enhanced.

Figure 5 is showing the graph of the change rate of total sum of square by the number of epoch. Besides the proposed method is showing faster speed of primary convergence than conventional methods in the graph of total sum of square and smaller total sum of square.

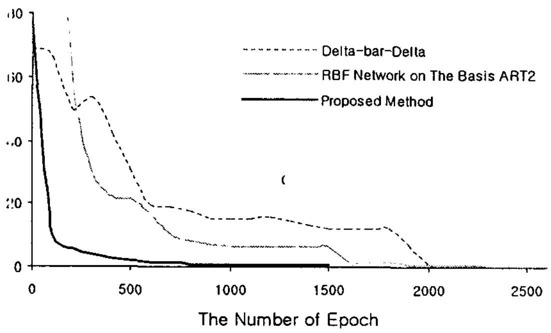

Figure 5 - Graph of total sum of square

4. Conclusions

The learning of RBF network on the basis of ART2 algorithm is divided to two stages. At the first stage the competitive learning is performed between the input layer and the middle layer and at the second stage the supervised learning between the middle layer and the output layer. Enhanced RBF network, which applies ART2 algorithm to the input layer and the middle layer to enhance the efficiency of learning of conventional RBF network on the basis of ART2 and the system of arbitrating the learning rate and momentum automatically by using the fuzzy logic system to arbitrate the weight value efficiently between the middle layer and the output layer, is proposed in this paper. When the absolute value of the difference from the output vector and the target vector is below ε, the proposed learning rate and momentum are classified to the accuracy in terms of the dynamic arbitration and to the inaccuracy otherwise. Then the learning rate is arbitrated dynamically by applying the number of the accuracy and the inaccuracy to the fuzzy logic system and the efficiency of learning is enhanced by the dynamically arbitrated rate.

The fact that the proposed method did not react sensitively to the times of learning and the convergence whereas conventional methods did and the total sum of square has decreased remarkably than conventional methods is verified by the experiment of applying it to the classification of number patterns, which were abstracted from the citizen registration card. Therefore the efficiency of learning in the proposed method is concluded to be enhanced.

The study of the method to generate the optimized middle layer by enhancing the efficiency of ART2 algorithm will be the subject of study in the future.

References

[1] M. Watanabe, K. Kuwata and R. Katayma (1994), "Adaptive Tree-Structured Self Generating Radial Basis Function and its Application to Nonlinear Identification Problem," *Proceedings of IIZUKA*, pp.167-170.

[2] J. Lo (1998), "Multi-layer Perceptrons and Radial Basis Functions are Universal Robust Approximators," *Proceedings of IJCNN*, Vol. 2, pp.1311-1314.

[3] C. Panchapakesan, D. Ralph and M. Palaniswami (1998), "Effects of Moving the Centers in an RBF Network," *Proceedings of IJCNN*, Vol. 2, pp.1256 -1260.

[4] R. A. Jacobs (1998), "Increased rates of convergence through learning rate adaptation," *IEEE Transactions on Neural Networks*, Vol. 1, No. 4, pp. 295-308.

[5] J. U. Ryu, T. K. Kim, K. W. Kim (2002), "The Passport Recognition by Using Enhanced RBF Neural Networks," *Proceedings of Intelligent Information Systems*, pp.529-534.

[6] K. B. Kim, S. W. Jang and C. K. Kim (2003), "Recognition of Car License Plate by Using Dynamical Thresh-olding Method and Enhanced Neural Networks," *Lecture Notes in Computer Science*, LNCS 2756, pp. 309-319.

[7] M. Jamshidi, N. Vadiee, T. J. Ross (1993), *Fuzzy Logic and Control*, Prentice-Hall.