

論文96-33A-11-17

# 역전파 신경회로망의 수렴속도 개선을 위한 학습파라메타 설정에 관한 연구

## (On the Configuration of Learning Parameter to Enhance Convergence Speed of Back Propagation Neural network)

洪 鳳 和 \*, 李 昇 柱 \*\*, 趙 源 敬 \*\*\*

(Bong Wha Hong, Seung Joo Lee, and Won Kyung Cho)

### 요 약

역전파 알고리즘의 학습속도를 향상 시키기 위하여 연결강도 갱신에 관여하는 학습파라메타를 가변적으로 갱신하는 방법을 제안하였다. 제안한 방법은 출력층의 목표값이 실제 출력값과 오차가 많이 날 수록 학습계수를 크게 변경하여 신경회로망의 학습과정에서 학습 횟수를 줄여준다. 반면 출력층의 출력값이 목표값에 접근하면 학습계수의 변동폭을 작게함으로써 신경회로망의 학습을 효율적으로 수행할 수 있도록 해준다. 제안된 방법은 X-OR 및 3-페리티 문제의 학습을 통하여 그 효용성을 확인하였다.

### Abstract

In this paper, the method for improving the speed of convergence and learning rate of back propagation algorithms is proposed which update the learning rate parameter and momentum term for each weight by generated error, changely. the output layer of neural network generates a high value in the case that output value is far from the desired value, and generates a low value in the opposite case. this method decreases the iteration number and is able to learning effectively. The effectiveness of proposed method is verified through the simulation of X-OR and 3-Parity problem.

### I. 서 론

인공지능의 한 분야로서, 인간의 뇌 기능을 가진 로보트나 컴퓨터에 대한 연구가 진행되어 왔으며, 이와 같은 연구분야 중에서 인간의 뇌를 단순화하여 신경모델로 모델링한 신경회로망 연구분야가 대두되었다.

이러한 신경회로망은 신경단위인 뉴런의 특성을 갖는 기본소자로 구성하여 아날로그 형태로 처리하거나 디지털 형태로 처리하고자 하였으며, 패턴인식이나 영상 처리, 적응제어등 여러 분야에 연구의 관점을 갖고 진행되고 있다. [3] [4] [5]

신경회로망의 알고리즘으로는 역전파 알고리즘이 널리 이용된다. 역전파 알고리즘은 미분의 반복규칙을 여러번 반복적으로 적용하여 확률 근사치를 유도해내는 알고리즘으로서 결합강도 공간(Weight Space)의 결합점(Synapse)에서 오차의 표면(Error Space)이 기울기를 따라가는(Gradient Descent) 방법으로서, 연결가중치를 층 별로 오차신호에 따라서 바꾼다. [1] [2] [3] [4] [5]

\* 正會員, 慶熙大學校 電子工學科  
(Dept. of Electronic Eng., Kyunghee Univ.)

\*\* 正會員, 淸田 專門大學 電子科  
(Dept. of Electronic, Hye Jeon Junior)

\*\*\* 正會員 慶熙大學校 電子工學科  
(Dept. of Electronic Eng., Kyunghee Univ.)

接受日字:1995年9月31日, 수정완료일: 1996年11月11日

역전파 신경회로망의 연결강도들이 임의의 작은값으로 초기화 된 후 학습패턴이 입력되면 전방향 계산에 의해 출력노드 값들이 계산된다. 그 다음, 출력층의 가중치들은 출력노드의 오차신호에 따라 변경되는데, 출력노드의 오차 신호는 목표값과 실제값의 차이에 시그모이드 활성화 함수의 기울기가 곱해진 값이다. 전층의 연결강도들은 출력노드의 오차신호가 연결강도를 통하여 역전파된 신호에 시그모이드 활성화 함수의 기울기가 곱해진 신호에 따라 갱신된다. 이러한 방법으로 모든 전층의 연결강도들이 갱신된다. 이와 같이 역전파 알고리즘은 전방향 계산과 역방향 연결강도 갱신의 반복에 의해 신경회로망을 학습시킨다. 그렇지만, 이 방법은 반복학습으로 인한 과도한 학습시간이 요구되는 문제점과 기울기 감소법(Gradient Decendent)을 사용하여 오차를 최소화 시키기 때문에 종종 국소해(local minimum)로 떨어지는 문제점들이 야기된다.<sup>[11] - [17]</sup>

위에서 언급한 역전파 알고리즘은 반복적인 연산과 학습을 수행시 연결강도의 갱신과정에 있어 학습 파라메타값을 일정한 상수값으로 고정하여 일률적으로 가중치 갱신에 관여하는 불합리성을 지니고 있다. 이 문제를 해결하기 위하여 Jacobs는 제한된 공간상에서 국소 기울기 정보를 이용하여 여러가지의 학습률을 부가하는 DBD(Delta Bar Delta)를 발표하였다.<sup>[8]</sup> 이 알고리즘은 학습속도를 증가시키는데 효과적이지만, 인위적으로 관성계수를 부가함에 의해 종종, 신경회로망을 발산하게 만드는 문제가 발생한다. 또한, Minai 와 Williams는 DBD알고리즘을 개선하여 EDBD(Extended Delta bar Delta) 알고리즘을 제안하였다.<sup>[9]</sup> 이 방법은 표준 DBD알고리즘에 비하여 시스템의 안정도 및 학습속도는 향상되었으나, 많은 부가적인 계산이 필요하고 학습파라메타에 관여하는 상수값들의 설정에 많은 시간이 소요되는 문제점을 갖는다.

한편, Yao Zhang은 에너지 함수를 이용하여 학습계수를 적응적으로 변경하는 알고리즘을 제안하였다.<sup>[10]</sup> 이 알고리즘은 위의 방법들보다 계산 방법이 단순하고 효과적이지만, 학습파라메타 갱신에 관여하는 상수값들을 각 층마다 각기 다르게 설정해야 하는 문제점 및 학습계수에 민감한 신경망일 경우, 발산하게 될 가능성이 높다.

본 논문에서는 역전파 알고리즘의 연결강도 갱신과정에 영향을 미치는 학습계수와 관성계수를 오차의 크기에 따라 적응적으로 변환시킴으로 새로운 파라메타

의 부가없이 시스템의 안정도를 향상시키고 학습속도를 향상시킬 수 있는 알고리즘을 제안하였다.<sup>[11] [12]</sup> II장에서는 DBD 알고리즘, EDBD알고리즘 및 적응 알고리즘(ABP)을 설명하고, 제안된 알고리즘과 그에 따른 역전파 알고리즘을 설명하였다. III장에서 이 3가지 방법을 XOR 및 3-Parity문제의 학습 모의실험을 통하여 비교분석하였으며, IV장에서 결론을 맺었다.

## II. 학습 파라메타의 설정

그림 1과 같이 각 층이  $N_i$ 개의 노드를 가지고 L층으로 구성된 역전파 신경회로망을 고려하자.

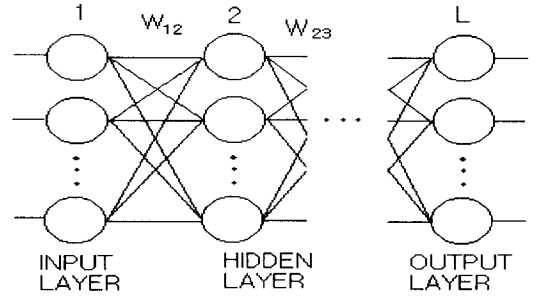


그림 1. 역전파 신경회로망  
Fig. 1. Backpropagation neural networks.

k층에서, k+1층으로 신호가 전달될 경우, (k+1)층 노드의 입력( $u_j^{k+1}$ )은 k층의 시그모이드 함수의 출력( $a_i^k$ )과 연결강도의 내적으로서 구하며, 그 과정은 다음과 같이 기술할 수 있다.

$$u_j^{k+1} = \sum_{p=1}^{N_k} W_{pj}^k a_i^k \quad (p=i, j) \tag{1}$$

식(1)의 노드 내부에 있어서 시그모이드 함수에 의해 출력되는 값은 식(2)와 같이 기술할 수 있다.

$$a_i^k = \frac{1}{1 + \exp(-u_j^k)} \tag{2}$$

실제출력  $a_i$  와 목표출력  $T_i$ 와의 차이를 구하기 위하여 평균자승오차(LMS)알고리즘을 적용하면 식(3)과 같다.

$$E = \frac{1}{2} \sum_{i=1}^{N'} (T_i^k - a_i^k)^2 \tag{3}$$

역전파 신경회로망은 오차 E를 최소화 하기 위하여 은닉층의 가중치에 대한 미분을 변경시키면서 학습을

행한다. 입력과 혼련패턴에 의한 학습과정은 가중치를 순차적으로 변경하면서 수행하며, 그 연결은 학습 알고리즘에 기본을 둔다. 학습과정에 있어서, 각 층의 노드에서는 식 (4)와 같이 오차가 발생한다.

$$\delta_j^{k+1} = a_i^{k+1} (1 - a_i^{k+1}) \sigma_i^k \quad (4)$$

k층의 발생한 오차는 k+1층의 오차와 연결강도의 내적으로 식 (5)와 같이 기술할 수 있다.

$$\sigma_i^k = \sum_{j=1}^l W_{ij}^{k+1} \delta_j^{k+1} (j = i, j) \quad (5)$$

단, 출력층의 오차  $\sigma_i^l$ 는 식(6)과 같다.

$$\sigma_i^l = a_i^l - T_i^l \quad (6)$$

식 (4)와 (5)을 입력층에서 출력층까지 순차적으로 반복하여 중간층에서의 오차를 계산한다. k 층과 k+1 층 노드간의 연결강도  $W_{ij}^k$ 에 대한 변화량  $\Delta W_{ij}^k$ 는 식 (7)과 같다.

$$W_{ij}(t+1) = W_{ij}(t) + \eta \delta_j^{k+1} a_i^k + \alpha \Delta W_{ij}(t-1) \quad (7)$$

식 (7)에서 t는 반복횟수를 나타내며, 관성계수( $\alpha$ ), 학습계수( $\eta$ )는 고정된 상수값으로서 1회의 학습에 따른 변화량을 계산하기 위하여 이용되며, 학습계수와 관성계수는 가중치 갱신시 중요한 역할을 한다. 그러나, 기존에는 이들 파라미터 값을 고정시킴으로서 오차값이 크던 작던간에 일정한 고정된 비율에 의하여 가중치가 갱신되는 불합리성을 지니고 있다. 이 문제를 해결하기 위하여 DBD알고리즘이 다음과 같이 제안 되었다.

$$\eta_{ij}(t+1) = \eta_{ij}(t) + \Delta \eta_{ij}(t) \quad (8)$$

윗 식에서 t는 반복횟수를 나타내며, i, j는 각각 해당 층과 노드를 나타낸다. 식 (9)는 식 (10)과 (11)에 생성된 기울기 정보량에 의해 3가지 상태의 값으로 표현 되는 학습계수를 나타낸다.

$$\Delta \eta_{ij}(t) = \begin{cases} k & \text{if } \overline{\delta_{ij}}(t-1)\delta_{ij}(t) > 0 \\ -\phi \eta_{ij}(t) & \text{if } \overline{\delta_{ij}}(t-1)\delta_{ij}(t) < 0 \\ 0 & \text{그 외} \end{cases} \quad (9)$$

식 (10)은 오차공간에서의 기울기의 변화량을 나타내며, 이 기울기의 변화량에 따라 학습률을 식 (9)와 같이 갱신한다.

$$\delta_{ij}(t) = \frac{\partial E(t)}{\partial W_{ij}(t)} \quad (10)$$

$$\overline{\delta_{ij}}(t) = (1 - \theta)\delta_{ij}(t) + \theta \overline{\delta_{ij}}(t-1) \quad (11)$$

윗 식에서 알수 있듯이 DBD알고리즘은 국소 기울기의 정보에 따라 학습계수를 변경하여 학습속도를 향상시켰지만, k,  $\phi$ ,  $\theta$ 값이 부적절할 경우, 신경회로망을 발산하게 만들거나 학습속도를 저하시키는 요인이 발생한다. 이러한 문제를 보완하고 향상시킨 알고리즘이 EDBD알고리즘이다.

$$\Delta W_{ij}(t) = -\eta_{ij}(t) \frac{\partial E}{\partial W_{ij}(t)} + \mu_{ij}(t) \Delta W_{ij}(t-1) \quad (12)$$

EDBD 알고리즘은 (14)와 같이 표준 DBD 알고리즘이 간과하였던 관성계수항을 부가하였다. 또한, 학습계수와 관성계수의 상한치와 하한치를 정하여 신경회로망의 안정도를 어느정도 향상시켰다.

$$\eta_{ij}(t+1) = \text{MIN}[\eta_{\max}, \eta_{ij}(t) + \Delta \eta_{ij}(t)] \quad (13)$$

$$\mu_{ij}(t+1) = \text{MIN}[\mu_{\max}, \mu_{ij}(t) + \Delta \mu_{ij}(t)] \quad (14)$$

DBD알고리즘이 국소기울기가 큰 부분에 대하여 상수 값 k를 일정하게 증가 시키고, 평탄한 영역에서는 k값을 감소시킨것에 비하여, EDBD알고리즘은 역으로 평탄한 영역에서는 학습계수를 지수함수적으로 증가시키고, 국소 기울기가 급한영역에서는 지수함수적으로 감소시킴으로서 학습효율을 향상시켰다.

$$\Delta \eta_{ij}(t) = \begin{cases} \left[ \begin{matrix} k_l \exp(-\gamma_l \overline{\delta_{ij}}(t)) \\ -\phi \eta_{ij}(t) \\ 0 \end{matrix} \right] \Delta \mu_{ij}(t) \\ \left[ \begin{matrix} k_m \exp(-\gamma_m \overline{\delta_{ij}}(t)) & \text{if } \overline{\delta_{ij}}(t-1)\delta_{ij}(t) > 0 \\ -\phi_n \mu_{ij}(t) & \text{if } \overline{\delta_{ij}}(t-1)\delta_{ij}(t) < 0 \\ 0 & \text{그 외} \end{matrix} \right] \end{cases} \quad (15)$$

그러나, EDBD알고리즘 역시, 학습계수와 관성계수를 갱신하기 위하여 부가적인 계산이 필요하며, 학습계수와 관성계수를 3가지 상태로 제한함으로써 학습효율이 유한하다는 단점이 발생하며, 학습파라메타 각 계수값의 설정문제가 제기된다. 또한, 식 (16) 과 같이 오류함수를 이용하여 적응적인 방법으로 학습계수를 적응적으로 갱신하는 알고리즘이 제안되었다.

$$\begin{cases} \Delta \eta_{ij}(n+1) = -\varepsilon \frac{\partial E_p}{\partial \eta_{ij}} + \beta \Delta \eta_{ij}(n) & \frac{\partial E_p}{\partial \eta_{ij}} = -\left(\frac{\partial E_p}{\partial W_{ij}}\right)^2 \\ \Delta \eta_{jk}(n+1) = -\mu \frac{\partial E_p}{\partial \eta_{jk}} + \gamma \Delta \eta_{jk}(n) & \frac{\partial E_p}{\partial \eta_{jk}} = -\left(\frac{\partial E_p}{\partial W_{jk}}\right)^2 \end{cases} \quad (16)$$

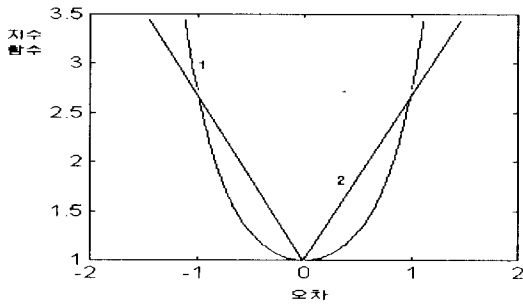
식 (16)에서 알수 있듯이 DBD 및 EDBD 알고리즘에

비하여 학습효율 및 계산량을 줄일 수 있었으나, 각 층마다 학습파라메타에 관여하는 상수값을 서로 다르게 선택하여야 하는 문제점이 발생한다. 또한, 관성계수를 배제함으로써 학습효율의 극대화를 기대할 수 없으며, 신경회로망의 안정성이 학습파라메타에 민감하지 않은 응용분야에만 적용되는 단점이 있다.

이러한 관점에서 역전과 학습을 위한 학습파라메타 설정 방법을 제안하였다. 신경회로망에 입력이 주어질 경우, 전방향 처리과정을 거쳐 출력층에서 식 (7)과 같이 오차를 산출하며, 식 (8)의 에너지를 최소화하기 위하여 학습과정을 반복한다. 이 경우, 발생된 오차는 양의 오차와 음의 오차로 표현된다. 신경회로망에 발생한 오차를 지수함수적으로 증가 또는 감소 함수형태로 나타내면 식 (17)과 같다.

$$h(\delta_{ij}) = A \exp(\delta_{ij}) \quad (17)$$

그러나, 식 (17)에 있어 오차가 음으로 증가될 경우, "0"로 근접하는 문제가 발생한다.



( 1 :  $\exp(\delta^2)$  , 2:  $\exp(|\delta|)$  ) 단  $\delta$  : 오차 )

그림 2. 오차에 따른 지수함수값의 변화

Fig. 2. The change value of exponential function for error.

따라서, 그림 2에 나타낸 바와 같이 음의 오차를 양의 오차가 갖는 성질로 하기 위한 방법으로 식 (18)과 같이 오차의 절대치를 이용하거나 오차를 제곱하여 나타낼 수 있다.

$$\begin{cases} h(\delta_{ij}^2) = A \exp(\delta_{ij}^2) & \delta_{ij} > 1, \delta_{ij} < -1 \\ h(|\delta_{ij}|) = A \exp(|\delta_{ij}|) & -1 \leq \delta_{ij} \leq 1 \end{cases} \quad (18)$$

$A$  : 상수

식 (18)에서 발생한 오차에 의한 두 식의 크기를 비교하면 표 1과 같다.

표 1에서 알 수 있듯이 발생한 오차가 1보다 크거나 -1보다 작을 경우, 오차의 제곱에 대한 지수함수값이 오

차의 절대치 함수값보다 크므로써 연결강도의 갱신시 연결강도의 변동량이 커진다. 역으로, 오차가 -1 과 1 사이의 값을 가질 경우에는 오차의 제곱에 대한 지수 함수값이 오차의 절대치에 대한 지수 함수값보다 작게 된다. 따라서, 연결강도의 갱신시, 연결강도의 변동폭을 작게 함으로써 시스템의 안정도를 향상시킬 수 있으므로 오차의 제곱에 대한 지수함수값을 학습파라메타로 이용하는것이 더욱 효율적이다.

표 1. 오차에 따른 절대치 함수값과 오차를 제곱한 함수값의 비교

Table 1. Comparing Absolute value and square value of error.

발생한 오차 ( $\delta$ )	대소 비교
$-1 < \delta < 1$	$E_{pow} < E_{abs}$
$\delta = \pm 1$	$E_{pow} = E_{abs}$
$\delta > 1, \delta < -1$	$E_{pow} > E_{abs}$

식(17)을 학습계수에 적용하여 기술하면 다음과 같다.

$$\eta_{ij}(n+1) = \eta_{ij}(n) + \Delta\eta_{ij}(n) \quad (19)$$

$$\begin{aligned} \Delta\eta_{ij}(n) &= \begin{pmatrix} h(\delta_{ij}^2) & \delta_{ij} > 1, \delta_{ij} < -1 \\ h(|\delta_{ij}|) & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \\ &= A \begin{pmatrix} \exp(\delta_{ij}^2) & \delta_{ij} > 1, \delta_{ij} < -1 \\ \exp(|\delta_{ij}|) & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \end{aligned} \quad (20)$$

또한, 식 (7)에서 학습계수와 관성계수는 반비례 관계가 성립되므로 관성계수에 대응된 관계식은 다음과 같이 나타낼 수 있다.

$$\alpha_{ij}(n+1) = \alpha_{ij}(n) + \Delta\alpha_{ij}(n) \quad (21)$$

$$\Delta\alpha_{ij}(n) = \begin{pmatrix} \frac{1}{h(\delta_{ij}^2)} & \delta_{ij} > 1, \delta_{ij} < -1 \\ \frac{1}{h(|\delta_{ij}|)} & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} = B \begin{pmatrix} \frac{1}{\exp(\delta_{ij}^2)} & \delta_{ij} > 1, \delta_{ij} < -1 \\ \frac{1}{\exp(|\delta_{ij}|)} & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \quad (22)$$

따라서, 역전과 알고리즘의 연결강도 갱신과정은 식 (20), (22)를 식 (7)에 대입하여 얻을 수 있다.

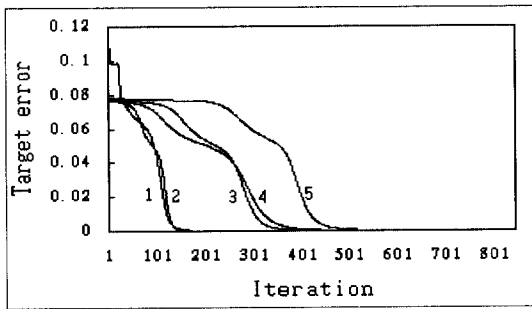
$$\begin{aligned} \Delta W_{ij}(n+1) &= \eta_{ij}(n) \delta_{ij} + \alpha_{ij}(n) W_{ij}(n) \\ &= \begin{pmatrix} h(\delta_{ij}^2) \delta_{ij} + \frac{1}{h(\delta_{ij}^2)} W_{ij}(n) & \delta_{ij} > 1, \delta_{ij} < -1 \\ h(|\delta_{ij}|) \delta_{ij} + \frac{1}{h(|\delta_{ij}|)} W_{ij}(n) & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \\ &= \begin{pmatrix} A \exp(\delta_{ij}^2) \delta_{ij} + \frac{1}{A \exp(\delta_{ij}^2)} W_{ij}(n) & \delta_{ij} > 1, \delta_{ij} < -1 \\ A \exp(|\delta_{ij}|) \delta_{ij} + \frac{1}{A \exp(|\delta_{ij}|)} W_{ij}(n) & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \end{aligned} \quad (23)$$

III. 모의실험 및 평가

본 논문에서 제안한 알고리즘의 타당성을 검증하기 위하여 X-OR 및 3-패리티 문제에 적용하여 보았으며, 고정적인 상수값을 이용하는 고전적인 역전파알고리즘과 기존에 제안된 방법(DBD알고리즘, EDBD알고리즘, ABP알고리즘)과의 수렴속도를 비교분석 하였다.

실험에 사용한 신경회로망의 구조는 입력층, 은닉층 그리고 출력층을 갖는 삼층구조의 역전파 신경회로망을 사용했으며, 각 층의 노드는 입력층 수에 따라 결정된다. 또한, 층과 층 사이의 연결선은 모두 이어져 있으며 가중치는 -0.5 ~ 0.5까지 이산적으로 발생시켜서 실험하였다.

본 실험에서 수렴을 위한 제한조건은 최대오차 한계를  $10^{-5}$ 로 하였고 최대 반복횟수로 5000번으로 제한하였다.



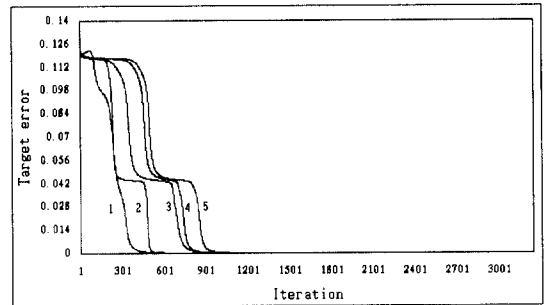
(1:제안된 방법, 2:ABP방법 3:DBD방법, 4:EDBD방법, 5: 표준 BP)

그림 3. XOR 문제에 대한 역전파 모델의 수렴과정  
Fig. 3. The convergence of BP model for XOR problem.

그림 3은 X-OR에서의 오차수렴 과정을 나타낸다. DBD알고리즘과 EDBE알고리즘의 실험은 참고문헌 [9]에 기술한 값( $k=k_1=0.095$ ,  $\phi=\phi_1=0.1$ ,  $k_m=0.1$ ,  $\phi_m=0.3$ ,  $\gamma_m=1.0$ )으로 하였으며, ABP알고리즘은 비교적 빠르게 수렴하는 상수값( $\eta_i\_init=0.9$ ,  $\eta_k\_init=0.9$ ,  $\alpha=0.7$ ,  $\varepsilon=1.0$ ,  $\beta=1.0$ ,  $\nu=1.0$ ,  $\gamma=1.0$ )으로 하여 실험하였다. 실험결과, 표준적인 역전파 알고리즘은 843회의 반복횟수를 나타내었으며, DBD알고리즘은 649회의 반복횟수를 나타내었다. 또한, EDBD알고리즘 및 ABP알고리즘은 각각 639회, 342회의 반복횟수를 나타내었고 제안된 알고리즘은 239회의 반복횟수를 보임으로써, 제안된 알고리즘이 기존의 방법에 비하여 안정되고 적은 반복

횟수를 나타낸다.

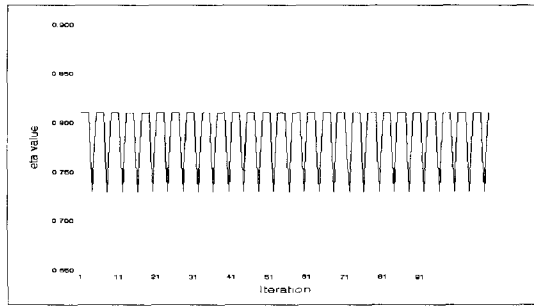
그림 4는 3-패리티 문제에 대한 역전파 알고리즘과 오차 수렴과정을 나타내며, 실험에 이용한 파라메타 집합은 XOR와 같은 값을 사용하였다.



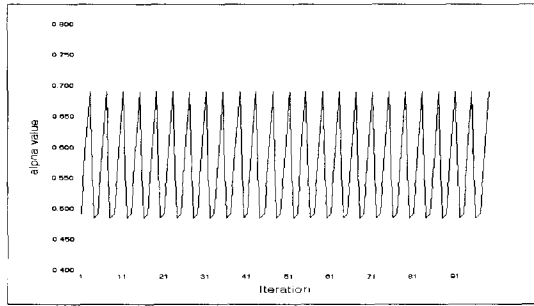
(1:제안된 방법, 2:ABP방법, 3:DBD방법, 4:EDBD방법, 5:표준 BP)

그림 4. 3-패리티 문제에 대한 역전파 모델의 수렴과정  
Fig. 4. The convergence of BP model for 3-parity problem.

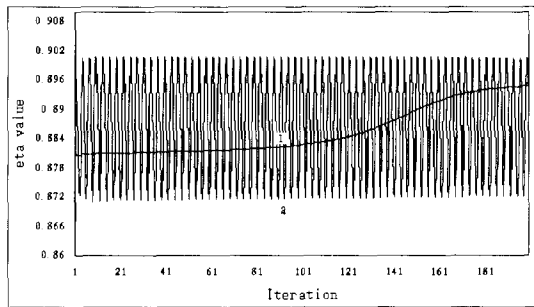
실험결과, 3-패리티 문제에 있어 표준적인 역전파 모델은 3246회의 반복횟수를 나타내었고, DBD알고리즘의 경우는 2751회의 반복횟수를 나타내었다. 또한, EDBD알고리즘 및 ABP알고리즘은 각각 2683회, 2662의 반복횟수를 나타내었고, 제안된 알고리즘이 1163회의 반복횟수를 보임으로써 3-패리티 문제에 있어서도 기존의 방법에 비하여 오차수렴이 빠르고 안정됨을 보였다. 그림 5는 역전파 알고리즘의 학습방법에 따른 학습계수( $\eta$ )값의 변동사항을 나타낸다. 그림 5(a), (b), (c), (d)에서 알 수 있듯이 DBD알고리즘과 EDBD알고리즘은 학습계수( $\eta$ )값을 가변적으로 변경하여 학습효율을 높일 수 있었으나, 제한된 상태에서 학습계수를 설정해야 하는 문제점이 발생한다. ABP알고리즘은 DBD 및 EDBD알고리즘에 비하여 어느정도 안정도 및 학습효율이 향상되었으나, 각층마다 서로다른 상수파라메타를 설정해야 하는 문제 및 초기 학습파라메타들의 설정이 부적절한 경우, 수렴하지 못하고 발산하는 경우가 존재한다. 또한, 관성계수값을 배제함으로써 학습효율의 극대화를 기대할 수 없는 단점과 학습계수에 민감하지 않은 신경회로망의 분야에만 적용되는 단점이 있다. 반면, 제안된 방법은 학습파라메타값을 선형적으로 증감함으로써 시스템의 안정도를 향상시킬 수 있으며, 학습효율의 극대화를 이룰 수 있다.



(a)

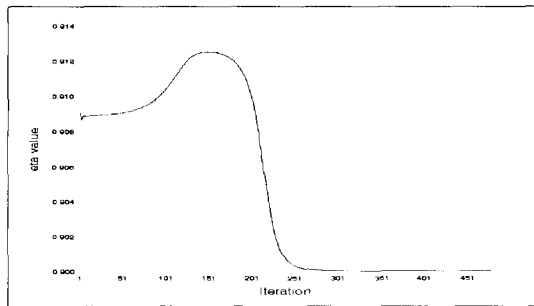


(b)



(1:  $\eta_j$ 의 변화량 2:  $\eta_{jk}$ 의 변화량)

(c)

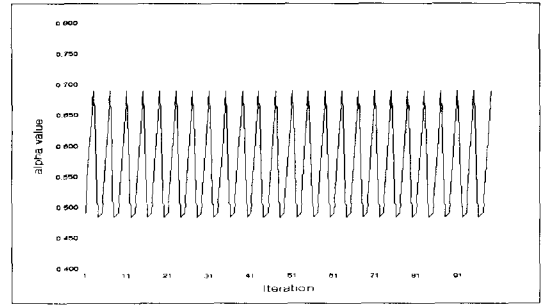


(d)

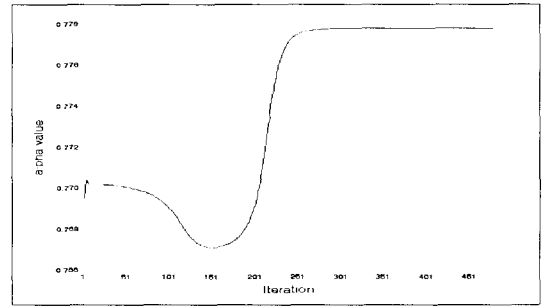
그림 5. 역전과 알고리즘의 (a) DBD, (b)EDBD, (c) ABP알고리즘 (d) 제안된 학습방법에 따른 학습계수( $\eta$ )의 변화

Fig. 5. (a) the DBD method, (b) EDBD method, (c) ABP method, and (d) Proposed method for the change of learning rate of the back propagation algorithms.

그림 6 (a)는 역전과 알고리즘의 EDBD방법에 따른 관성계수( $\alpha$ )값의 변동상태를 나타내며, 그림 6 (b)는 역전과 알고리즘에 제안된 방법을 적용하여 학습할 경우의 관성계수의 변화량을 나타낸다.



(a)



(b)

그림 6. 역전과 알고리즘의 (a)EDBD, (b) 제안된 학습방법에 따른 관성계수( $\alpha$ )의 변화

Fig. 6. (a) EDBD method, and (b) Proposed method for the change of momentum rate of the back propagation algorithms.

그림 6에서 알 수 있듯이 제안된 알고리즘이 EDBD알고리즘에 비하여 관성계수값을 선형적으로 가변시킴으로써 시스템의 안정도를 향상시킬 수 있으며, 학습을 효율적으로 수행 할 수 있음을 고찰하였다.

#### IV. 결론

본 논문에서는 역전과 알고리즘에 의한 학습시 부적절한 학습파라메타에 의한 학습의 지연현상을 줄이고 시스템의 안정도를 향상시킬 수 있는 학습파라메타 설정방법을 제안 하였다. 이 방법은 연결강도 갱신과정시 사용되는 학습파라메타값을 발생한 오차에 따라 가변적으로 변화시킨다. 출력값이 목표값과 가까우면 작은 학습계수를 발생시켜 연결강도의 변동폭을 작게하여

불필요한 학습을 줄이고 시스템의 안정도를 향상시킨다. 제안한 학습파라메타 설정방법은 X-OR 및 3-페리티 문제의 모의실험을 통하여 고정 상수값을 학습파라메타로 이용하는 표준적인 역전파 알고리즘과 기울기 정보를 이용하는 DBD, EDBD 및 적응(ABP)방법과 비교하였다. 모의 실험결과 표준적인 역전파 알고리즘에 대한 X-OR에서의 오차수렴 과정은 고정적인 학습파라메타를 이용함으로써, 불필요한 학습을 과다하게 수행함으로써 반복횟수가 증가되는 문제점이 나타났다. DBD, EDBD 및 ABP방법은 표준적인 역전파 알고리즘에 의한 학습에 비하여 반복횟수를 줄일 수 있었지만 연결강도의 갱신에 관여하는 파라메타들의 초기치를 부적절하게 선택하면 일반화 성능은 오히려 좋지 못하였다. 제안한 방법은 오차에 따라 학습계수와 관성계수를 선형적으로 변경시키면서 학습을 효율적으로 수행하였으며, X-OR 및 3-페리티 문제에 있어, 반복횟수를 2~3배정도 줄일 수 있었다. 특히, 시스템의 안정도를 고려하면 제안한 방법이 더 우수했으며, 부가적인 초기값 계산을 하지 않고 효율적으로 학습을 수행하였다.

앞으로의 연구방향은 본 논문에서 제안한 알고리즘을 필기체 문자인식의 적용과 학습 파라메타의 초기값을 초기 발생오차에 따라서 적응적으로 발생하게 하는 연구가 필요하다.

#### 참 고 문 헌

- [1] D.E Rumelhart, J.L McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vols.1 and 2, MIT Press, Cambridge, MA., 1986.
- [2] *DARPA Neural Network study*, AFCEA international press, 1988.
- [3] Yoh-Han Pao, *Adaptive Pattern Recognition and Neural Networks*. Addison Wesley pp. 113-139, 1989.
- [4] James A. Freeman, David M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques*. Addison-Wesley pp. 89-125, 1991.
- [5] G. A. Carpenter, "Neural Network Models for Pattern Recognition and Associative Memory", *Neural Networks*, vol. 2, pp. 243-257, 1989.
- [6] Yoshio Hirose, Koichi Yamashita and Shimpei Hijiya, "Backpropagation Algorithm Which Varies the Number of Hidden Units", *Neural Networks*, vol., 4, pp. 61-66, 1991.
- [7] Jacek M. Zurada, *Introduction to Artificial Neural System*. West, pp. 163-248, 1992.
- [8] R.A Jacobs, "Increased Rates of Convergence through Learning Rate Adaptation", *Neural Networks*, vol. 1, pp. 295-307, 1988.
- [9] Ali A. Minai and Ronald D. Williams, "Acceleration of back-propagation through learning rate momentum adaptation", *International Joint Conference on Neural Network*, vol. no.1, Nagoya Japan, pp 676-679, Oct 1993.
- [10] Yao Zhang, "Updating learning rates for back-propagation network", *International Joint Conference on Neural Networks*, vol. no.1 Nagoya Japan, pp 569-572, Oct. 1993.
- [11] 홍봉화, 탁형근, 조원경 "역전파 학습알고리즘의 최적화를 위한 파라메타에 관한 연구", 대한전자공학회 추계종합 학술대회 논문집, vol. 15, no. 1, pp. 725-727 1992
- [12] 홍봉화, 조원경, "역전파 신경회로망의 속도개선을 위한 학습파라메타 설정에 관한 연구", 대한전자공학회 인공지능, 신경망 및 퍼지시스템 학술 발표대회 논문집, vol. 1, pp. 105-112. 1995

## 저 자 소 개



洪 鳳 和(正會員)

1962년 10월 25일생. 1987년 경희대학교 전자공학과(공학사). 1992년 경희대학교 전자공학과 졸업(공학석사). 1996년 12월 현재 경희대학교 전자공학과 박사과정. 주관심 분야는 컴퓨터구조, 병렬처리, 신경

회로망, VLSI 설계등임.

趙 源 敬(正會員) 第 29卷 B編 第 2號 參照



李 昇 柱(正會員)

1956년 3월 16일생. 1979년 경희대학교 전자공학과(공학사). 1985년 경희대학교 전자공학과 졸업(공학석사). 1996년 12월 현재 경희대학교 전자공학과 박사과정. 1996년 현재 혜전 전문대학 전자과 조교수.

주관심 분야는 컴퓨터구조 신경망 및 퍼지이론 VLSI 설계등임