

논문 99-8-1-07

## 다층 구조 신경회로망의 학습 속도 향상을 위한 활성화 함수의 변화

李秉度\*, 李敏鎬\*\*

Variation of activation functions for accelerating the learning speed  
of the multilayer neural network

Byungdo Lee\*, Minho Lee\*\*

## 요 약

이 논문에서는 오차 역전파 학습 알고리즘의 학습 속도를 향상시키기 위한 새로운 학습 방법을 제안한다. 제안하고자 하는 방법은 시그모이드 형태를 갖는 신경회로망의 활성화 함수(activation function) 자체에 고차항(higher order)을 적절히 이용하여 초기 학습 단계에서 발생할 수 있는 조기 포화(premature saturation) 현상을 계산량의 큰 증가 없이 효과적으로 대처할 수 있다. 고차항을 이용함으로써 은닉층 활성화 함수의 도함수가 작은 값으로 감소함에 따라 신경망의 연결 강도를 학습시키는 학습율은 적응적으로 큰 값을 갖게 된다. 또한, 은닉층에 고차항을 이용하는 제안한 방법에 모멘텀(momentum) 학습 알고리즘을 결합하는 새로운 hybrid 학습 방법을 제안한다. 컴퓨터 모의 실험을 통해 제안하고자 하는 학습 방법과 기존의 방법들과의 학습 속도 성능을 비교한다.

## Abstract

In this paper, an enhanced learning method is proposed for improving the learning speed of the error back propagation learning algorithm. In order to cope with the premature saturation phenomenon at the initial learning stage, a variation scheme of active functions is introduced by using higher order functions, which does not need much increase of computation load. It naturally changes the learning rate of inter-connection weights to a large value as the derivative of sigmoid function abnormally decrease to a small value during the learning epoch. Also, we suggest the hybrid learning method incorporated the proposed method with the momentum training algorithm. Computer simulation results show that the proposed learning algorithm outperforms the conventional methods such as momentum and delta-bar-delta algorithms.

## 1. 서론

측정 대상의 물리량이나 화학량을 선택적으로 포착하여 유용한 신호(주로 전기적 신호)로 변환·출력하는 장치인 센서(sensor)는 기본적으로 우수한 감도(sensitivity), 선택도(selectivity), 안정도(stability) 및

복귀도(reversibility) 등의 기본요건(basic requirements)과 높은 기능성, 적용성, 규격성, 생산성, 보존성 등 다양한 부대요건(subsidiary requirements)들을 갖추어야 한다. 최근에 보다 나은 성능의 센서를 개발하기 위해 인간의 정보 처리 방식을 공학적으로 모델링한 인공 신경회로망을 이용하고자 하는 연구가 있어 왔다.<sup>1-3</sup> 한 예로, 인간의 후각 기능을 갖는 odour-sensing 시스템에서 수정발진기(quartz-resonator) 센서를 배열시켜 진동 주파수(oscillation frequency)의 증감

\* 한국해양대 전기공학과 (Dept. of Electrical Eng., Korea Maritime Univ.)

\*\* 경북대학교 센서기술연구소 (Sensor Technology Research Center, Kyungpook National Univ.)

<접수일자 : 1998년 8월 30일>

을 파악하고, 주파수에 따른 패턴들을 인식하여 에탄올 농도(ethanol concentration)를 측정하는 시스템이 있다.<sup>1</sup> 복잡한 비선형 패턴의 분류 및 인식을 위해 오차 역전파 학습 알고리즘을 갖는 신경 회로망 모델이 이용되어 우수한 패턴 인식기로 이용되었다.<sup>1</sup> 또한, 광섬유(optical fiber) 스트레인 센서의 출력 특성 조정(calibration)과정에 있어서, 스트레인 센서에 본질적으로 내재한 비선형 특성을 처리하기 위해 다층 구조 신경회로망을 이용하는 연구 결과도 소개되었다.<sup>3</sup> 센서의 특성 보정 및 인식기로 다양하게 사용되어온 대부분의 신경회로망 모델은 다층 구조를 갖는 신경회로망 모델로서 연결 가중치들(interconnection weights)의 학습 알고리즘으로 오차 역전파 학습 알고리즘(error back-propagation learning algorithm)이 흔히 사용되고 있다.<sup>4-6</sup>

오차 역전파 학습 알고리즘은 가중치 공간(weight space)에서 오차표면(error surface)의 경사도(gradient)에 의존한다는 점에서 급강하 기법(steepest descent technique)의 1차 근사(first order approximation)라고 할 수 있다. 그러나, 이 학습 방법은 신경회로망의 초기 학습 과정에서 일어날 수 있는 조기 포화(premature saturation) 현상에 의해 지역 최소점(local minima)에 빠질 수 있으며, 학습 수렴율은 상대적으로 느린 경향이 있다.<sup>4-7</sup> 지금까지 오차 역전파 학습 알고리즘(error backpropagation learning algorithm)의 수렴 속도(convergence speed)를 향상시키기 위해 많은 연구가 있어왔다. Conjugate gradients와 Quasi-Newton 알고리즘이 그 한 예이며, 이 알고리즘은 2차 경사도(second order gradient)를 사용했지만 계산량이 상대적으로 많고,<sup>8-10</sup> 특히 대규모 실제 문제의 응용에서는 학습에 걸리는 시간이 너무 많다. 계산량의 부담을 줄이기 위한 방법으로 다른 근사 방법이 소개되었지만, 역행렬을 구하는 대신에 이용된 대각 헤시안 근사 모델(diagonal Hessian approximation model)들의 정확성에 학습 과정의 성공 여부가 달려있다.<sup>9</sup> 학습 속도를 향상하기 위한 또다른 방법은 "Quickprop"가 있다.<sup>11</sup> 이것은 곡률 변화량의 측정과 같이 오차 함수의 두 연속적인 경사도간의 차이를 사용하고, 이러한 정보는 step size를 변화시키는데 사용된다. Quickprop 알고리즘이 학습 시간을 상당히 감소시킨다고 알려진 이래로, step size를 적응적으로 변화시키는 몇몇의 알고리즘들이 소개되었다. 이들의 예가 델타-바-델타 학습 알고리즘과

differential step size back-propagation (DSBP) 알고리즘들이다.<sup>12-14</sup> 이것은 학습율에 변화를 주어 평평한 학습 곡선상에서 학습 속도를 증가시키는데 필요한 것이라 알려져 있다.

본 논문에서는, 계산량의 큰 손실없이 오차 역전파 학습 알고리즘의 조기 포화 현상을 효과적으로 대처하기 위한 새로운 학습 방법을 제안한다. 제안하고자 하는 방법은 신경회로망 학습의 초기 과정에서 비정상적으로 증가된 상호결합 가중치(weight)가 있으면, 활성화 함수 출력을 자동적으로 고차 활성화 함수의 출력으로 대체한다. 이것은 자연스럽게 급경사 강하법(steepest descent)에 기초한 학습 알고리즘의 도함수에 의해 보다 큰 학습을 값을 적응적으로 발생시키게 되어 조기 포화 현상을 벗어나게 할 수 있다. 또한, 기존의 모멘텀 학습 알고리즘에 제안한 고차 활성화 함수를 결합하는 새로운 hybrid 학습 알고리즘을 도입하고 컴퓨터를 이용한 모의 실험으로 제안한 방법의 성능을 보인다.

## II. 다층 구조 신경회로망의 학습 알고리즘

다음 그림은 역전파 알고리즘을 수행하기 위한 다층 구조 퍼셉트론(multilayer perceptron)의 망 구조를 나타낸다.<sup>15-18</sup>

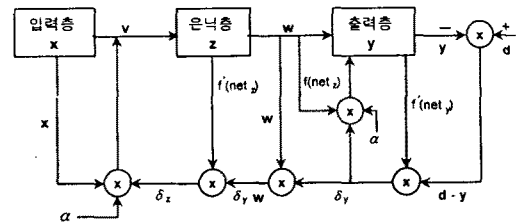


그림 1. 오차 역전파 알고리즘을 이용한 다층 퍼셉트론  
Fig. 1. Multilayer perceptron using error back propagation algorithm

위 그림에서  $x$ 는 입력 패턴 (input pattern) 벡터를 나타내며  $y (= f(\text{net}_y))$ 는 출력 패턴 (output pattern) 벡터이며,  $d$ 는 학습하고자 하는 목적 패턴 (target pattern) 벡터를 의미한다.  $v$ 는 1층,  $w$ 는 2층의 상호결합 가중치 벡터이며,  $\delta_z$ 는 은닉층(hidden layer)의 오차 신호이고  $\delta_y$ 는 출력층(output layer)의

오차 신호이다. 또한,  $net_z$ 는 입력 패턴과 1층 상호결합 가중치의 곱에 대한 합 ( $= \sum_i \mathbf{x} \mathbf{v}$ )이며,  $net_y$ 는 은닉층과 2층 상호결합 가중치의 곱에 대한 합 ( $= \sum \mathbf{z} \mathbf{w}$ ,  $\mathbf{z} = f(net_z)$ )을 나타낸다. 그리고  $f'(net_z)$ ,  $f'(net_y)$ 는 은닉층의 출력과 출력층의 출력에 대한 시그모이드 형태를 갖는 활성화 함수의 도함수(derivative)이다. 일반적인 오차 역전과 학습 알고리즘의 오차 신호는 식 (1)과 같고, 1층 및 2층의 가중치 변화값(weight variation)과 가중치 갱신(weight update)은 식 (2)와 같다.

$$\begin{aligned} \delta_z &= -\frac{\partial E}{\partial z_j} [z_j]' \\ &= -\frac{\partial E}{\partial z_j} f'(net_j) \end{aligned} \quad (1)$$

$$\begin{aligned} \Delta \mathbf{v}^k &= \alpha \delta_z \mathbf{x}^k, \mathbf{v}^{k+1} = \mathbf{v}^k + \alpha \delta_z \mathbf{x}^k \\ \Delta \mathbf{w}^k &= \alpha \delta_y \mathbf{z}^k, \mathbf{w}^{k+1} = \mathbf{w}^k + \alpha \delta_y \mathbf{z}^k \end{aligned} \quad (2)$$

식 (1)과 (2)에서 보는 바와 같이 학습 과정의 초기에서 비정상적으로 증가된 연결 가중치들은 오차 역전과 학습 알고리즘에서 이용되는 활성화 함수의 도함수들을 아주 작은 값으로 만들어 지역 최소점에 빠지게 되고, 따라서 학습 속도가 상당히 저하되는 결과를 초래할 수 있다.<sup>7</sup> 오차 역전과 학습 알고리즘의 학습 속도를 개선하기 위해 소개된 대표적인 몇가지 방법들을 소개하면 다음과 같다.

### 2.1 Momentum algorithm

일반적인 역전과 알고리즘에서 오프-라인상의 학습 속도를 향상시키기 위한 한 방법으로 모멘텀을 이용하는 방법이 있다.<sup>9</sup> 이러한 모멘텀 알고리즘은  $k$ 번째 학습 단계에서 은닉층과 출력층간의 가중치 변화값  $\Delta \mathbf{v}^k$ 와 입력층과 은닉층간의 가중치 변화값  $\Delta \mathbf{w}^k$ 을 다음과 같은 식에 의해 구한다.

$$\begin{aligned} \Delta \mathbf{v}^k &= \alpha \delta_z \mathbf{x}^k + \beta \Delta \mathbf{v}^{k-1} \\ \Delta \mathbf{w}^k &= \alpha \delta_y \mathbf{z}^k + \beta \Delta \mathbf{w}^{k-1} \end{aligned} \quad (3)$$

여기서  $\alpha$ ,  $\beta$ 는 모멘텀 상수이다. 다음은  $k+1$ 번째 학습 단계에서 모멘텀 규칙을 이용하여 가중치를

갱신하는 식이다.

$$\begin{aligned} \mathbf{v}^{k+1} &= \mathbf{v}^k + \alpha \delta_z \mathbf{x}^k + \beta \Delta \mathbf{v}^{k-1} \\ \mathbf{w}^{k+1} &= \mathbf{w}^k + \alpha \delta_y \mathbf{z}^k + \beta \Delta \mathbf{w}^{k-1} \end{aligned} \quad (4)$$

여기서  $\mathbf{v}^{k+1}$ 는 은닉층과 출력층간의 가중치 갱신을 의미하고,  $\mathbf{w}^{k+1}$ 는 입력층과 은닉층간의 가중치 갱신이다.  $k+1$ 번째의 연결 가중치 갱신에서  $k$ 번째의 연결 가중치에  $k-1$ 번째의 연결 가중치의 정보를 추가로 이용하므로써 학습 속도를 개선할 수 있지만, 시그모이드 활성화 함수의 도함수가 작은 영역에서는 기존의 오차 역전과 학습 방법의 한계를 크게 개선하기는 힘들다.

### 2.2 Delta-bar-Delta ( $\Delta - \bar{\Delta}$ ) algorithm

일반적인 역전과 알고리즘에서 오프-라인상의 학습 속도를 향상시키기 위한 또다른 방법의 하나로는 델타-바-델타 알고리즘이 있다.<sup>8</sup> 출력층과 은닉층간의 가중치 변화값은 다음과 같이 되고,

$$\Delta^k_{\mathbf{w}} \equiv -\delta_y \mathbf{z}^k \quad (5)$$

$$\bar{\Delta}^k_{\mathbf{w}} \equiv (1-\zeta)\Delta^k_{\mathbf{w}} + \zeta \bar{\Delta}^{k-1}_{\mathbf{w}}$$

$$\Delta^k_{\mathbf{v}} \equiv -\delta_z \mathbf{x}^k \quad (6)$$

$$\bar{\Delta}^k_{\mathbf{v}} \equiv (1-\zeta)\Delta^k_{\mathbf{v}} + \zeta \bar{\Delta}^{k-1}_{\mathbf{v}}$$

$k+1$ 번째 학습 단계에서 새로운 학습율은 다음의 방정식으로 주어진다.

$$a^{k+1} = \begin{cases} a^k + \nu & ; \bar{\Delta}^{k-1} \cdot \Delta > 0 \\ (1-\gamma)a^k & ; \bar{\Delta}^{k-1} \cdot \Delta < 0 \\ a^k & ; \bar{\Delta}^{k-1} \cdot \Delta = 0 \end{cases} \quad (7)$$

여기서  $\zeta$ ,  $\nu$ ,  $\gamma$ 는 임의의 양의 상수이다.  $k+1$ 번째 학습 단계에서 델타-바-델타 규칙을 사용한 가중치의 갱신은

$$\begin{aligned} \mathbf{w}^{k+1} &= \mathbf{w}^k + \Delta \mathbf{w}^k \\ &= \mathbf{w}^k + a^{k+1} \delta_y \mathbf{z}^k \end{aligned} \quad (8)$$

$$\begin{aligned} \mathbf{v}^{k+1} &= \mathbf{v}^k + \Delta \mathbf{v}^k \\ &= \mathbf{v}^k + \alpha^{k+1} \delta_z \mathbf{x}^k \end{aligned}$$

와 같이 된다.

이 방법은 학습율의 크기를 적응적으로 바꾸어 학습 속도를 개선할 수 있을 뿐만 아니라, 조기 포화 현상에 의한 지역 최소화 문제를 어느 정도 극복할 수 있지만, 임의의 크기에 의해 학습율을 갱신하므로, 설정된 상수 값의 크기에 따라 학습 과정에 진동 현상이 일어날 수 있으며, 주어진 문제에 적합한 최적 상수 값을 찾기가 어렵다.

### III. 고차항에 의한 활성화 함수의 변화

이 절에서는 계산량의 큰 손실없이 학습 속도를 향상시키는데 효과적인 새로운 학습율의 적응적 설계 방법을 설명한다. 오차 역전파 알고리즘에서의 수렴 속도가 느린 이유는 주로 조기 포화 현상에 기인한다.<sup>7</sup> 만약 뉴런의  $net_j$  값,  $net_j = \sum_i x_i v_{ji}$ 가 비정상적으로 시그모이드 함수의 큰 값으로 커진다면,  $f'(net_j)$ 는 작아진다. 여기서  $x_i$ 는 입력 패턴이고,  $v_{ji}$ 는 1층 상호결합 가중치이다. 그러므로, 상호결합 가중치의 변화율은 감소한다. 합당하지 않는 포화(saturation)에 대처하기 위해 소개된, DSBP 알고리즘은 역전파 학습 알고리즘에서  $f'(\cdot)$ 을 제거하고, 다른 알고리즘들에는 역전파 학습 과정 동안  $f'(\cdot)$ 가 임계치(threshold value)보다 적을 때  $f'(\cdot)$ 의 값 대신에 상수값(constant value)을 사용한다. 그러나, 이들의 방법들은 방법상 인위적인 경향이 있으며, 또한 학습의 최종 단계에서 진동(oscillation) 현상을 내포하게 된다.

본 논문에서는 합당하지 않는 조기 포화에 대처하기 위한 새로운 학습 방법을 제안한다. 만약 출력 오차가 원하는 만큼 감소되지 않는 상황에서 활성화 함수의 미분값이 임계치보다 작다면, 기존의 시그모이드 함수  $f(\cdot)$  대신에  $f^3(\cdot)$ 와 같은 고차 활성화 함수(higher order activation function)값을 사용한다. 여기서 임계치라함은 시그모이드 함수값에서 조기 포화현상이 일어날 확률이 큰 값의 한계치를 의미한다. 본 논문에서는 오차의 정확도가 미리 정한 목적 함수에 의해 계산된 오차값이 원하는 정도의 90%이내에 들지

않은 상태에서 오차의 변화분이 현재 오차 크기의 10%이내에서 바뀌지 않으면, 이때 각 은닉층 뉴런들의 출력값이 0.9 이상의 값을 갖는 은닉층 뉴런들을 조기 포화 현상에 빠져 있다고 가정한다. 따라서, 임계치 0.9 보다 큰 값이 각 은닉층 시그모이드 함수의 결과값으로 나타날 경우 각 은닉층의 시그모이드 함수는 제안하는 고차 활성화 함수를 사용한다. 제안된 신경회로망의 학습 절차는 다음과 같다. 우선 오차 함수(error function)를 다음 식 (9)와 같이 정의한다.

$$E = \frac{1}{2} \sum_{k=1}^m (d_k - y_k)^2 \quad (9)$$

여기서  $m$ 은 목적하는 학습 패턴의 수이다.  $d_k$ 는 학습하고자 하는 목적 패턴값이고  $y_k$ 는 신경망의 출력이다. 뉴런의 가중합과 출력  $z_j$ 는

$$net_j = \sum_i x_i v_{ji}, \quad z_j = [f(net_j)] \quad (10)$$

이고, 여기서  $x_i$ 는 입력 뉴런 벡터이며,  $w_{ji}$ 는  $i$ 번째 입력 뉴런(input neuron)과  $j$ 번째 은닉 뉴런(hidden neuron) 사이의 가중치를 의미하고  $f$ 는 시그모이드 함수이다. 만약 오차(error)  $E$ 가 원하는 값보다 크고,  $\frac{\partial E}{\partial w_{ji}}$ 가 임계치보다 작다면, 뉴런 출력은 다음 식으로 바뀐다.

$$net_j = \sum_i x_i v_{ji}, \quad z_j = [f^3(net_j)]^3 \quad (11)$$

위 식과 같은 활성화 함수를 사용한 학습 알고리즘은 다음과 같다. 식 (11)에서의  $z_j$ 와 앞 단층의 출력 사이에 연결된 가중치로 전파되는 오차 신호는

$$\begin{aligned} \delta_z &= -\frac{\partial E}{\partial z_j} [z_j]' \\ &= -3[f^3(net_j)]^2 \frac{\partial E}{\partial z_j} f'(net_j) \end{aligned} \quad (12)$$

가 된다. 식 (12)에서 역전파되는 오차 신호를 사용하여 가중치의 갱신은 다음과 같다.

$$\begin{aligned}
 v_{ji}^{k+1} &= v_{ji}^k + \Delta v_{ji}^k \\
 &= v_{ji}^k + \alpha \delta_z x_{ki}^k
 \end{aligned}
 \tag{13}$$

식 (12)에서 보여지는 것처럼, 오차 역전파 학습 알고리즘을 이용하는 다층 구조 신경회로망의 학습 초기 과정에서 조기 포화 현상이 발생하면, 고정된 상수값  $\alpha$ 를 갖는 학습율은 식 (13)에서 나타낸 것처럼 제안된 알고리즘에 의해  $3f(net_j)^2 \alpha$ 의 형태로 적응적으로 바뀐다. 따라서, 조기 포화 현상에 의해 지역 최소점에 빠진 연결 가중치들을 효과적으로 학습시킬 수 있을 것으로 기대된다.

제안한 학습 방법은 학습 속도를 향상시키기 위하여 기존의 모멘텀 학습 방법에 제안된 고차 활성화 함수를 결합하는 새로운 형태의 hybrid 학습 방법을 이용한다. 제안하고자 하는 hybrid 방법은 기존의 모멘텀 학습 알고리즘의 장점을 이용하고, 학습율의 크기 변화분의 설정을 고차항을 이용해서 결정하는 방법이다.

모멘텀 학습 알고리즘에 고차 활성화 함수를 이용하는 Hybrid 학습 방법은 다음과 같다.  $k+1$  번째 학습 단계에서 조기 포화 현상을 갖는 경우의 새로운 학습율은 hybrid 방법에 의하여 가중치의 갱신은

$$\begin{aligned}
 v^{k+1} &= v^k + \alpha \delta_z x^k + \beta \Delta v^{k-1} \\
 w^{k+1} &= w^k + \alpha \delta_y z^k + \beta \Delta w^{k-1}
 \end{aligned}
 \tag{14}$$

와 같이 된다. 식(14)는 기존의 모멘텀 알고리즘인 식 (4)와 같으나, 조기 포화 현상이 발생할 경우  $z^k$ 와  $\delta_z$ 는 식 (11)과 (12)로 변환하여 학습하는 동안 일어날 수 있는 조기 포화에 적응적으로 대처하게 된다.

#### IV. 컴퓨터 모의 실험

기존의 학습 알고리즘과 제안된 학습 알고리즘의 성능을 비교·평가하기 위해 XOR 문제와 기존의 오차 역전파 학습 방법으로 비교적 학습시키기가 어려운 4 비트 패리티 검사(parity check) 문제를 이용한 컴퓨터 모의 실험을 수행한다. 표 1은 XOR 문제 및 4 비트 패리티 검사 문제로서 모의 실험을 위한 입력 패턴과 학습하고자 하는 목적 패턴을 나타낸다.

표 1. XOR 및 4 비트 패리티 문제

Table 1. XOR and 4 bit parity problem

$x_1$	$x_2$	$y^k$	$x_1$	$x_2$	$x_3$	$x_4$	$y^k$
1	1	-0.9	1	1	1	1	-0.9
1	-1	0.9	1	1	1	-1	0.9
-1	1	0.9	1	1	-1	1	0.9
-1	-1	-0.9	1	1	-1	-1	-0.9
			1	-1	1	1	0.9
			1	-1	1	-1	-0.9
			1	-1	-1	1	-0.9
			1	-1	-1	-1	0.9
			-1	1	1	1	0.9
			-1	1	1	-1	-0.9
			-1	1	-1	1	-0.9
			-1	1	-1	-1	0.9
			-1	-1	1	1	-0.9
			-1	-1	1	-1	0.9
			-1	-1	-1	1	0.9
			-1	-1	-1	-1	-0.9

표 2. 모의 실험을 위한 파라미터 값

Table 2. Parameter values in the simulation

Problem	은닉층 뉴런수		constant		
			$\eta$	$\sigma$	$\beta$
XOR	3	GBP	0.83	0.42	
		MBP	0.86	0.42	0.90
		NMBP	0.70	0.56	0.90
	4	GBP	0.88	0.80	
		MBP	0.90	0.80	0.90
		NMBP	0.90	0.80	0.90
	5	GBP	0.65	0.68	
		MBP	0.76	0.68	0.90
		NMBP	0.76	0.68	0.90
	6	GBP	0.59	0.68	
		MBP	0.59	0.76	0.90
		NMBP	1.19	0.76	0.90
4 bit Parity	5	GBP	0.57	0.54	
		MBP	0.69	0.50	0.80
		NMBP	0.69	0.50	0.80
	6	GBP	0.60	0.60	
		MBP	0.42	0.60	0.80
		NMBP	0.52	0.60	0.80
	7	GBP	학습안됨		
		MBP	0.68	0.46	0.80
		NMBP	0.73	0.48	0.80
	8	GBP	0.61	0.40	
		MBP	0.68	0.42	0.80
		NMBP	0.72	0.42	0.80

표 2는 모의 실험에 사용된 파라미터 값을 표시하며, GBP(General Back-Propagation)는 일반적인 역전파 알고리즘을 사용한 것이고, MBP (Momentum Back-Propagation)는 모멘텀 알고리즘을 나타내고, NMBP (New MBP)는 제안하고자 하는 것으로서 기존의 모멘텀 알고리즘에 은닉층 고차 활성화 함수를 적용한 것이다. 또한, 표 2의 파라미터 값은 여러 번의 시행 착오에서 얻은 신경망이 최적의 학습속도를 갖도록 설정된 값이다. 다층 신경회로망의 구조는 1개의 은닉층을 사용한다.<sup>18</sup> 학습 속도의 성능을 비교하기 위한 모의 실험은 여러 가지의 초기 조건을 가지고 행하였으며, 그 중 가장 좋은 학습 결과를 주는 초기 조건에 대해 기존의 방법인 일반적인 역전파 알고리즘, 모멘텀 알고리즘과 제안된 hybrid 학습 알고리즘의 학습 성능을 그림 2와 3에 나타내었다.

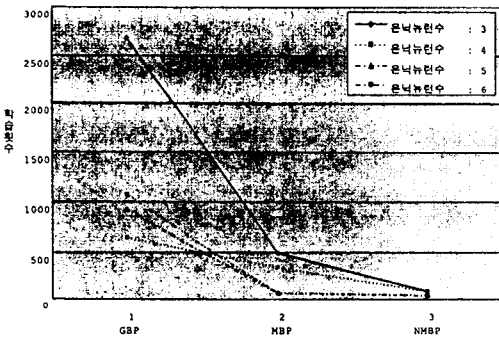


그림 2. XOR 문제에 대한 은닉층 뉴런의 변화에 따른 오차 수렴의 비교

Fig. 2. Comparison of error convergence of various number of hidden neurons in the XOR problem.

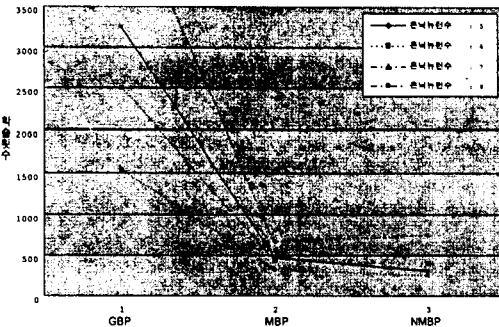


그림 3. 4 비트 패리티 문제에 대한 은닉층 뉴런의 변화에 따른 오차 수렴의 비교

Fig. 3. Comparison of error convergence of various number of hidden neurons in the 4 bit parity problem.

다음은 제안된 방법의 일반화(generalization) 측면에서의 성능 검토를 위해 신경망의 입력값에 대한 출력값의 비교를 표 3, 4를 통해 나타낸다. 표 3에서 보인 결과는 XOR 문제에 대해 은닉층 뉴런수가 3개인 신경망을 이용한 경우이고, 표 4는 패리티 문제에 대해 은닉층 뉴런수가 6개인 경우이다.

표 3. XOR 문제에 대한 신경망의 테스트 데이터 비교  
Table 3. Comparison of test data for neural network in the XOR problem

Input		Output		
$x_1$	$x_2$	GBP	MBP	NMBP
-0.9	-0.9	-0.844712	-0.863938	-0.876548
-0.9	0.8	0.837645	0.816783	0.853436
0.8	-0.8	0.788037	0.776637	0.765852
0.7	0.8	-0.735306	-0.746930	-0.768941

표 4. 4 비트 패리티 문제에 대한 신경망 테스트 데이터 비교

Table 4. Comparison of test data for neural network in the 4 bit parity problem

Input				Output		
$x_1$	$x_2$	$x_3$	$x_4$	GBP	MBP	NMBP
-0.9	-0.9	-0.9	-0.9	-0.891573	-0.738292	-0.891685
-0.9	-0.9	-0.8	0.8	-0.895693	-0.855460	0.909782
-0.8	-0.8	0.9	-0.9	0.887023	0.866754	0.917439
-0.8	-0.8	0.8	0.8	-0.843290	-0.617253	-0.794929
-0.9	0.9	-0.9	-0.9	0.843291	0.661523	0.815238
-0.9	0.8	-0.9	0.8	-0.875798	-0.805823	-0.910231
-0.8	0.9	0.8	-0.9	-0.863133	-0.779825	-0.902318
-0.8	0.8	0.8	0.8	0.890402	0.804321	0.879206
0.7	-0.7	-0.7	-0.7	0.643943	0.703079	0.824479
1.0	-0.9	-1.0	0.9	-0.894467	-0.897743	-0.908768
0.8	-0.8	0.7	-0.7	-0.626314	-0.403015	-0.543506
0.7	-0.7	0.8	0.8	0.845493	0.748585	0.869887
0.9	0.9	-0.7	-0.7	-0.836702	-0.789744	-0.864275
0.7	0.8	-0.9	0.9	0.467352	0.640144	0.796424
0.7	0.7	0.8	-0.9	0.471534	0.624667	0.774074
0.9	0.8	0.7	0.7	-0.898462	-0.784061	-0.877356

그림 4는 신경망의 매개변수(parameter)들을 변화시킬 때 제안하는 방법이 학습 속도의 개선면에서 일반성을 유지하는지를 검증하기 위해 기존의 학습 방법들과

제안한 방법의 오차 수렴 속도를 비교한 것이다.

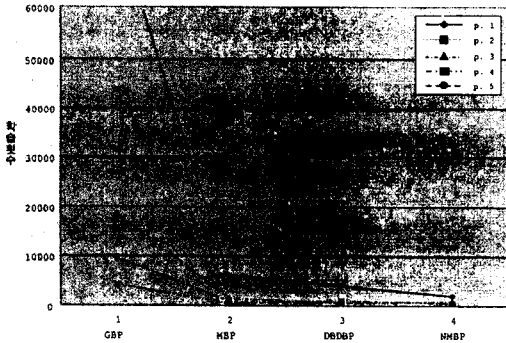


그림 4. 4 비트 패리티 문제에 대한 파라미터 변화에 따른 오차 수렴의 비교

Fig. 4. Comparison of error convergence of various parameter in the 4 bit parity problem.

그림 4의 DBDBP는 델타-바-델타 학습 알고리즘을 나타낸다. 사용된 학습 변수들은 각각의 학습 방법들에서 여러 번의 시행착오로부터 얻은 최적의 변수들이다. 그리고, 그림 4에서 나타낸 p.1 ~ p.5는 사용된 신경망의 매개 변수들의 값을 나타낸다.

표 5. 그림 4의 실험을 위한 파라미터

Table 5. Parameter for Fig. 4.

	은닉 뉴런수	Parameter					
		$\eta$	$\sigma$	$\beta$	$\zeta$	$\nu$	$\gamma$
p. 1	6	0.4	0.3	0.8	0.32	0.05	0.096
p. 2	6	0.6	0.5	0.8	0.28	0.05	0.097
p. 3	6	0.7	0.5	0.8	0.32	0.05	0.060
p. 4	5	0.5	0.5	0.8	0.34	0.06	0.100
p. 5	5	0.5	0.7	0.8	0.32	0.05	0.060

위의 결과에서 알 수 있는 것처럼, 제안된 hybrid 학습 알고리즘이 기존의 알고리즘에 비해 학습속도가 향상됨을 그림 2와 3, 4를 통해 확인할 수가 있고, 표 3과 4에서 보는 것처럼 일반화 측면에서도 기존의 알고리즘에 비해 큰 차이가 없다.

### V. 결론

본 논문에서는, 기존의 오차 역전파 학습 알고리즘의 학습 속도를 향상시키는 새로운 학습 알고리즘을 제안

했다. 고차 활성화 함수(higher activation functions)는 계산량의 큰 손실없이 기존의 학습 방법에 약간의 변화를 줌으로써 신경회로망의 학습 초기에 발생하는 조기 포화 현상을 효과적으로 대처한다. 기존의 델타-바-델타 방법에 고차 활성화 함수를 갖는 학습 알고리즘을 결합하는 hybrid 방법은 학습 시간을 상당히 감소시키는 결과를 주었다. 더 나아가 신경망의 학습 속도(training speed)와 일반화 성능(generalization performance)을 향상시키기 위해 은닉 묘사(hidden representation)에 대해 새로운 표준(criterion)을 세우는 것이 필요하고 연구중에 있다.

### 참고 문헌

- [1] T. Nakamoto, A. Fukuda and T. Moriizumi, "Perfume and favour identification by odour-sensing system using quartz-resonator sensor array and neural-network pattern recognition," *Sensors and Actuators B*, vol. 10, pp. 85-90, 1993.
- [2] Eugeniusz Porada, Mario Beaulieu, and Tinko A. Eftimov, "Automatic Calibration of a Fiber-Optic Strain Sensor Using a Self-Learning System," *IEEE transactions on instrumentation and measurement*, vol. 43, No. 2, pp. 341-346, APRIL 1994.
- [3] S. M. Yang and G. S. Lee, "Vibration Control of Smart Structures by Using Neural Networks," *Journal of Dynamic Systems, Measurement, and Control*, vol.119, pp. 34-39, MARCH 1997.
- [4] Philip D. Wasserman(1989), *Neural Computing, Theory and practice.*, Van Nostrand Reinhold, New York, pp. 43-60.
- [5] Philip D. Wasserman(1993), *Advanced methods in neural computing*, Van Nostrand Reinhold, New York, pp. 97-118.
- [6] Simon Haykin(1994), *Neural Networks A Comprehensive Foundation*, Macmillan, Toronto, pp. 138-229.
- [7] Y. J. Lee, S. h. Oh, M. W. Kim, "An analysis of premature saturation in backpropagation

- learning," *Neural Networks*, vol. 6, pp.719-728, 1993.
- [8] A. R. Webb, David Lowe, M. D. Bedworth (1988), A Comparison of Nonlinear Optimisation Strategies for Feedforward Adaptive Layered Royal Signals and Radar Establishment Memorandum 4157.
- [9] Sue Becker, Yann le Cun(1988), "Improving the Convergence of Back-Propagation Learning with Second Order Methods," *Proc. 1988 Summer School on Connectionist Model*, Carnegie-mellon Univ., pp. 22-37.
- [10] J. S. R. Jang, C. T. Sun and E. Mizutani(1977), *Neuro-Fuzzy and Soft Computing*, Prentice Hall.
- [11] Scott E. Fahlman(1988), "Fast-Learning Variations on Back-propagation:An Empirical Study," *Proc. 1988 Summer School on Connectionist Model*, Carnegie-Mellon Univ., pp. 38-51.
- [12] J. R. Chen and P. Mars(1990), "Stepsize Variation Methods for Accelerating the Back-Propagation Algorithm," *International Joint Conference on Neural Networks*, vol. 1, pp.601-604, 1990.
- [13] S. Haykin(1986), *Adaptive Filter Theory Englewood Cliffs, NJ*, Prentice-Hall.
- [14] Robert A. Jacobs(1988), "Increased Rates of Convergence Through Learning Rate Adaptation Neural Network," vol. 1, pp. 195-307.
- [15] Seung-Joon Lee and Dong-Jo Park(1993), "New accelerated learning algorithm motivated for novel shape of error surfaces for multilayer feedforward neural networks," *International Joint Conference on Neural Networks*, vol. 1, pp. 553-556, 1993.
- [16] Peter GECZY and Shiro USUI(1997), "Learning Performance Measures for MLP networks," *IEEE International Conference on Neural Networks*, vol. 3, pp. 1845-1850.
- [17] Yao Zhang(1993), "Updating Learning Rates for Backpropagation Network," *International Joint Conference on Neural Networks*, vol. 1, pp. 569-572, 1993.
- [18] Marcus Pfister and Raul Rojas(1993), "Speeding up Backpropagation - A Comparison of Orthogonal Techniques," *International Joint Conference on Neural Networks*, vol. 1, pp. 517-522, 1993.

---

 著 者 紹 介
 

---

## 이병도(Byungdo, Lee)

1966년 11월 14일생

1997년 2월 부경대학교 제어계측공학과(학사), 1997년 3월~1998년 2월 한국해양대학교 전기공학과 석사학위 취득 예정, 1985년~1991년 현대전자 근무, 주관심분야 : 신경회로망 모델링, 소음신호처리 및 신호 분리 등

E-mail : vipslee@kmucc.kmaritime.ac.kr



## 이민호(Minho, Lee)

1965년 12월 19일생

1988년 2월 경북대학교 공과대학 전자공학과(학사), 1992년 3월 한국과학기술원 전기 및 전자공학과(석사), 1995년 8월 한국과학기술원 전기 및 전자공학과(박사)

1995년 9월 ~ 1997년 9월 한국해양대학교 전기공학과 전임강사, 1997년 10월 ~ 1998년 2월 한국해양대학교 전기공학과 조교수, 1998년 3월 ~ 경북대학교 센서기술연구소 전임강사, 주관심분야 : 신경회로망, 뉴로-퍼지 회로망, 지능제어, 신호처리 센서시스템