

論文94-31B-4-17

# 확률적 근사법과 후향전과 알고리즘을 이용한 다층신경망의 학습성능 개선

## (Improving the Training Performance of Multilayer Neural Network by Using Stochastic Approximation and Backpropagation Algorithm)

趙鏞鉉\*, 崔興文\*\*

(Yong Hyun Cho and Heung Moon Choi)

### 要約

본 논문에서는 확률적 근사법과 후향전과 알고리즘을 조합함으로써 다층신경망의 학습성능을 개선 할 수 있는 효율적 방법을 제안하였다. 제안된 방법에서는 학습에 확률적 근사법과 후향전과 알고리즘을 혼합사용한 전역최적화 기법을 적용함으로써 학습성능을 최대한 개선할 수 있도록 하였다. 확률적 근사법은 국소최소점을 벗어나 전역최적점에 치우친 근사점을 결정해 주는 기능을 하도록 하고, 이점을초기값으로 하여 결정론적 기울기 하강법인 후향전과 알고리즘을 적용함으로써 빠른 수렴속도로도 전역최적점으로서의 수렴확률을 높였다. 제안된 방법에서 은닉층과 출력층의 학습율과 모멘트를 각 층의 뉴런활성화분포의 표준편차에 직접 적응 조정되도록 함으로써 학습속도를 더욱 개선하였다. 제안된 방법을 패리티 검사와 패턴 분류에 적용하여 그 타당성과 성능을 확인한 결과 제안된 방법은 초기값을 무작위로 설정하는 기존의 후향전과 알고리즘에 비해 전반적인 학습성능이 우수하며, Baba의 MROM과 Sun의 방법보다 학습속도가 빠름을 확인할 수 있었다. 또한 학습율과 모멘트를 적응 조정함에 의해서 학습속도를 약 20% 정도 더 개선할 수 있었다.

### Abstract

This paper proposes an efficient method for improving the training performance of the neural network by using a hybrid of a stochastic approximation and a backpropagation algorithm. The proposed method improves the performance of the training by applying a global optimization method which is a hybrid of a stochastic approximation and a backpropagation algorithm. The approximate initial point for fast global optimization is estimated first by applying the stochastic approximation, and then the backpropagation algorithm, which is the fast gradient descent method, is applied for a high speed global optimization. And further speed-up of training is made possible by adjusting the training parameters of each of the output and the hidden layer adaptively to the standard deviation of the neuron output of each layer. The proposed method has been applied to the parity checking and the pattern classification, and the simulation results show that the performance of the proposed method is superior to that of the backpropagation, the Baba's MROM, and the Sun's method with randomized initial point settings. The results of adaptive adjusting of the training parameters show that the proposed method further improves the convergence speed about 20% in training

\* 正會員, 嶺南專門大學 電子科  
(Dept. of Electronics, Yeungnam Junior College)

\*\* 正會員, 慶北大學校 電子工學科

(Dept. of Elec. Eng., Kyungpook Nat'l Univ.)

接受日字: 1993年 8月 9日

## I. 서론

최근 학습을 통하여 패턴의 인식<sup>[1,2]</sup>과 같이 복잡하고 계산집약적인 과제들을 능률적으로 해결하기 위하여 대규모 병렬성을 가진 신경망을 이용하고자 하는 연구가 많이 진행되고 있다.<sup>[3-15]</sup>

신경망을 학습에 이용할 때는 인식이나 분류의 능력을 높이기 위하여 입력층과 출력층 사이에 한개 이상의 은닉층(hidden layer)을 갖는 다층(multilayer)구조를 흔히 사용한다.<sup>[1]</sup> 이때 흔히 쓰이는 후향전과(backpropagation) 알고리즘은 기울기하강(gradient descent)법에 해당하는 결정론적(deterministic) 알고리즘으로 최적해로의 수렴속도는 빠르나 국소최적해를 만났을 때 이를 벗어 나기가 어렵다.<sup>[6,7]</sup> 이러한 후향전과 알고리즘에서는 학습율(learning rate)과 모멘트(momentum) 등의 학습 파라미터, 초기의 연결가중치(synapse weight), 또는 신경망의 구조 등에 따라 그 성능이 달라진다.

Rumelhart 등<sup>[7]</sup>은 후향전과 알고리즘에서 연결가중치를 경신(update)할 때에 학습 파라미터를 변화시켜 그 학습성능을 개선하였으며, 이때 학습율과 모멘트는 경험적으로 설정하였다. Pedone 등<sup>[8]</sup>은 은닉층만의 뉴런 활성화분포의 표준편차에 따라 은닉층과 출력층 각각의 연결가중치 경신에 이용되는 학습율과 모멘트를 적응 조정하여 학습속도를 개선하였다. 또한 Chen 등<sup>[9]</sup>은 삼층 전향(feedforward) 신경망에서 선형대수학에 근거하여 초기의 연결가중치를 적절히 설정함으로써 학습속도를 개선하였으나, 이때 국소최적해로 수렴될 가능성은 여전히 남아 있다.

Hirose 등<sup>[10]</sup>은 은닉층의 뉴런수를 에너지함수의 증감에 적응 조정하였고, Wang 등<sup>[11]</sup>은 자기성장학습(self growing learning) 알고리즘을 제안하여 전역최적해로의 수렴확률을 개선하였다. 그러나 이들 방법에서는 은닉층 뉴런의 추가 및 삭제에 따른 알고리즘 및 시스템의 복잡도가 증가되었다.

한편, Baba<sup>[12]</sup>는 후향전과 알고리즘의 오차함수를 최소화하기 위하여 수정된 무작위 최적화방법(modified random optimization method: MROM)을 제안하였으며 이를 패턴의 분류에 이용하였다. 이 방법에서는 연결가중치의 절대값이 한정된 범위내에 있을 때에만 전역최적해로 수렴되며, 두개 이상의 은닉층을 필요로 하고 가우스 분포함수의 편차에 따라 그 수렴속도가 크게 달라진다. 또한 Sun 등<sup>[13]</sup>은 Baba의 MROM에서 가우스 분포함수의 분산을 오차함수의 변화에 대해 적응 조정함으로써 은닉층을 한 개로 줄일 수 있었으나, 초기 분산값에 따

라서 최적해로의 수렴속도가 변화하였다.

이제까지의 연구에서는 전역최적해로의 수렴확률을 개선하거나 또는 학습속도를 개선하기 위한 연구 중에 하나에 역점을 두었을 뿐, 그 두가지를 동시에 만족시키기 위한 연구는 찾아보기 힘들다. 또한 전역최적해로의 수렴확률이나 또는 학습속도의 개선을 위하여 취한 방법으로는 주로 학습율과 모멘트의 조정이나 은닉층 뉴런수와 같은 신경망의 구조 변경등을 사용하였다. 그리고 초기 연결가중치는 학습성능을 결정하는 중요한 요소이나 대부분의 연구에서는 이를 경험적으로 또는 무작위로 설정하고 있으며, 초기 연결가중치를 전역최적해 가까이로 설정함으로써 전역최적해로의 수렴확률과 학습속도를 동시에 개선하려는 체계적인 방법은 거의 찾아볼 수 없다.<sup>[14]</sup> 또한 학습 파라미터의 적응조정을 통해 학습속도를 개선하는데 있어서, 은닉층과 출력층 연결가중치 경신에 있어서 공히 은닉층뉴런의 활성화분포에 대해서만 학습 파라미터를 적응 조정하였으나, 후향전과 알고리즘의 오차 전파기법으로 보아 출력층 연결가중치 경신은 출력층 뉴런의 활성화분포에 직접 적응 조정되도록 하는 것이 좀더 효과적일 것으로 기대된다.

본 논문에서는 확률적 근사법(stochastic approximation)<sup>[15]</sup>과 후향전과 알고리즘을 조합함으로써 다층신경망의 학습성능을 개선할 수 있는 효율적인 방법을 제안하였다. 제안된 방법에서는 확률적 근사법과 후향전과 알고리즘을 혼합 사용한 전역최적화 기법을 적용함으로써 학습성능을 최대한 개선할 수 있도록 하였다. 확률적 근사법은 국소최소점을 벗어나 전역 최적점에 치우친 근사적인 초기치를 결정해주는 기능을 하도록 하고, 이점을 초기값으로 하여 결정론적 기울기 하강법인 후향전과 알고리즘을 적용함으로써 빠른 수렴속도로 전역최적화가 가능하도록 하였다. 제안된 방법에서 은닉층과 출력층의 학습율과 모멘트를 각각 자기 층의 뉴런 활성화분포의 표준편차에 직접 적응 조정되도록 함으로써 학습속도를 더욱 개선하였다. 제안된 방법을 패리티 검사와 패턴 분류에 각각 적용 시뮬레이션하여 초기값을 무작위로 설정하는 기존의 후향전과 알고리즘, Baba의 MROM 및 Sun의 방법을 이용한 결과와 비교 고찰하였다.

## II. 확률적 근사법과 후향전과 알고리즘을 이용한 효율적 학습

함수 평활화(function smoothing) 기능을 갖는 확률적 근사법은 다극점함수(multiextremal function)의 전역최소점(global minimum)을 찾는

최적화 알고리즘이다.<sup>[15]</sup> 이때 최적화는 다극점함수  $f(x)$ 에 평활화 함수(smoothing function)  $h(\gamma)$ 를 상승적분(convolution)하여 평활된 함수(smoothed function)  $f(x, \beta)$ 를 구하고,  $h(\gamma)$ 의 분산을 제어하는 파라미터  $\beta$ 를 순차적으로 감소시켜 가며 근사화를 반복함으로써 이루어진다. 여기서  $x$ 는 상태변수이고  $\gamma$ 는 무작위교란벡터(random perturbation vector)이다. 평활화는 다극점함수를 단일극점함수로 변형하기 위한 목적으로 상승적분을 근사화하여 사용하며 평활화 함수  $h(\gamma)$ 를 이용하여 상태  $x$  주위의 제한된 공간에 대해 다극점함수  $f(x)$ 를 평균하여 구한다.  $\gamma$ 를 생성하는 평활화 함수  $h(\gamma)$ 로는 가우스, 균일, 또는 코오시 확률밀도함수(probability density function)가 이용된다.

예로서 연속이고 미분 가능한 2 개의 다른 최소점을 가지는 함수

$$f(x) = x^4 - 16x^2 + 5x \quad (1)$$

에서 평활된 함수  $f(x, \beta)$ 는

$$f(x, \beta) = (1/2)[f(x + \beta\gamma) + f(x - \beta\gamma)] \quad (2)$$

와 같이 상태  $x$  점을 중심으로  $\pm\beta\gamma$  범위 내에서의 평균치를 근사값으로 사용한다.

그림 1은 분산제어 파라미터  $\beta$ 의 순차값(sequence)을  $\{\beta\} = \{5.0, 4.0, 3.0, 2.0, 1.0, 0.001\}$ 로, 평

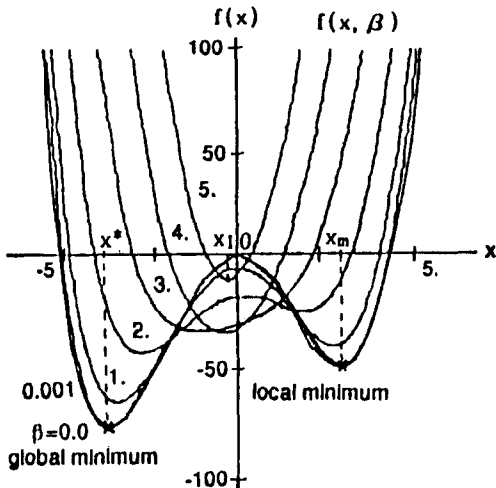


그림 1. 분산제어 파라미터  $\beta$ 의 변화에 대한 평활된 함수  $f(x, \beta)$

Fig. 1. Smoothed function  $f(x, \beta)$  to dispersion control parameter  $\beta$ .

활화 함수  $h(\gamma)$ 를 균일 확률밀도함수로 하고, 초기 상태를  $x_0 = 4.0$ 으로 할 때에, 각  $\beta$  값에 대한  $f(x, \beta)$ 의 변화를 나타낸 것이다. 그림에서 보는 바와 같이 확률적 근사법을 이용하여 전역최적해를 구하는데는  $\beta$ 의 순차값  $\{\beta\}$ 와, 각  $\beta$  값에 대한 제어 파라미터들을 적정설정해야 하는 번거로움이 있을 뿐 만 아니라 알고리즘의 확률적 동작으로 수렴속도가 느리다는 등의 문제점이 있다. 한편 그림에서  $\beta$  값이 5.0, 4.0, 그리고 3.0일 때를 보면 이 정도의 큰  $\beta$  값에 대해서는  $f(x, \beta)$ 가 단봉특성을 가짐을 알 수 있다. 이때 각  $f(x, \beta)$ 의 최소점은 원래 다극점함수  $f(x)$ 의 전역최적해에 대한 근사값이다. 그러므로 초기에 큰  $\beta$  값으로 평활화를 통한 확률적 근사법을 적용한다면 다극점함수  $f(x)$ 의 전역최적해로의 수렴이 용이한 초기상태를 구할 수 있을 것으로 보인다. 또한  $\beta$ 의 값이 클수록 함수  $f(x)$ 의 전역최소점  $x^*$ 와 평활된 함수  $f(x, \beta)$ 가 최소가 되는 상태  $X_i$ 사이의 거리  $D_{SA}$ 는 증가한다. 그러나 상태  $X_i$ 는 국소최소점  $x_m$ 보다는 전역최소점  $x^*$  쪽 경사면에 가깝게 위치한다. 따라서 이  $X_i$ 를 초기상태로 하여 결정론적 방법으로 수렴속도가 빠른 기울기 하강법을 적용한다면 직접 전역최소점  $x^*$ 로 빠르게 수렴시킬 수 있을 것이다.

따라서 학습을 위한 다층신경망에서 연결가중치벡터  $W$ 와 학습패턴  $p$ 에 대한 평균자승 오차함수(mean square error function)  $E_p(W)$ 의 최소값을 찾는 과정은 확률적 근사법으로 상태벡터  $x$ 에 대한 비용함수  $f(x)$ 의 최소값을 찾는 과정으로 대응시킬 수 있다. 그러므로 다층신경망의 학습에 확률적 근사법을 이용하여 평활된 평균자승 오차함수의 값이 최소인 연결가중치를 구하고, 구해진 연결가중치를 초기값으로 하여 기울기하강법의 빠른 수렴특성을 갖는 후향전파 알고리즘을 이용하면 빠른 시간내에 전역최소의 평균자승 오차를 보장하는 학습을 시킬 수 있음을 알 수 있다.

먼저, 다층신경망에 확률적 근사법을 적용하기 위해서 에너지함수  $E(W)$  및 학습패턴  $p$ 에 대한 평균자승 오차함수  $E_p(W)$ 를 각각

$$E(W) = \sum_p E_p(W) \quad (3)$$

$$E_p(W) = (1/2) \sum_i (x_{ip}(W) - d_{ip})^2 \quad (4)$$

로 정의한다. 여기서  $x_{ip}(W)$ 와  $d_{ip}$ 는 각각  $p$  번째 패턴에 대한 출력층 뉴런  $i$ 의 실제 출력과 원하는 출력이다. 또한 평균자승 오차함수  $E_p(W)$ 의 평활된 함수  $E_p(W, \beta)$  및 평활된 함수의 기울기  $\nabla_w E_p(W, \beta)$ 는

각각

$$\begin{aligned}
 E_p(W, \beta) &= (1/2) [E_p(W + \beta\gamma) + E_p(W - \beta\gamma)] \\
 &= (1/4) \sum_i [2d_{ip}^2 - 2d_{ip}(x_{ip}(W + \beta\gamma) \\
 &\quad + x_{ip}(W - \beta\gamma) + x_{ip}^2(W + \beta\gamma) + x_{ip}^2(W - \beta\gamma))] \\
 \nabla_w E_p(W, \beta) &= (1/4) \sum_i [-2d_{ip}(\nabla_w x_{ip}(W + \beta\gamma) \\
 &\quad + \nabla_w x_{ip}(W - \beta\gamma)) + \nabla_w x_{ip}^2(W + \beta\gamma) \\
 &\quad + \nabla_w x_{ip}^2(W - \beta\gamma)] \quad (5)
 \end{aligned}$$

로 구할 수 있다. 이때 출력층과 은닉층의 평활된 함수의 기울기는 각각 다음과 같이 계산된다. 즉, 출력층 연결가중치 경신시 평활된 함수의 기울기는

$$\begin{aligned}
 \nabla_w E_p(W, \beta) &= (1/2) \sum_i \left[ (f(\sum_j (W_{ij} + \beta\gamma) y_{jp}(W))) \right. \\
 &\quad - d_{ip}) f(\sum_j (W_{ij} + \beta\gamma) y_{jp}(W)) (1 \\
 &\quad - f(\sum_j (W_{ij} + \beta\gamma) y_{jp}(W))) + (f(\sum_j (W_{ij} \\
 &\quad - \beta\gamma) y_{jp}(W))) - d_{ip}) f(\sum_j (W_{ij} \\
 &\quad - \beta\gamma) y_{jp}(W)) (1 - f(\sum_j (W_{ij} \\
 &\quad - \beta\gamma) y_{jp}(W))) \left. \right] y_{jp}(W) \quad (6)
 \end{aligned}$$

와 같으며, 은닉층 연결가중치 경신시 평활된 함수의 기울기는

$$\begin{aligned}
 \nabla_w E_p(W, \beta) &= (1/2) \sum_i \left[ (f(\sum_j W_{ij} f(\sum_m (W_{jm} - \beta\gamma) o_{mp}))) \right. \\
 &\quad - d_{ip}) f(\sum_j W_{ij} f(\sum_m (W_{jm} - \beta\gamma) o_{mp})) (1 \\
 &\quad - f(\sum_j W_{ij} f(\sum_m (W_{jm} - \beta\gamma) o_{mp}))) \sum_j W_{ij} f(\sum_m (W_{jm} \\
 &\quad + \beta\gamma) o_{mp})) + (f(\sum_j W_{ij} f(\sum_m (W_{jm} - \beta\gamma) o_{mp}))) \\
 &\quad - f(\sum_j W_{ij} f(\sum_m (W_{jm} - \beta\gamma) o_{mp}))) \sum_j W_{ij} f(\sum_m (W_{jm} \\
 &\quad - \beta\gamma) o_{mp})) \left. \right] o_{mp} (1 - f(\sum_m (W_{jm} + \beta\gamma) o_{mp})) \quad (7)
 \end{aligned}$$

과 같이 계산된다. 여기서 i, j 및 m은 각각 출력층, 은닉층 및 입력층의 뉴런번호이다.  $y_{jp}(W)$ 는 입력패턴 p에 대한 은닉층 뉴런 j의 출력이며,  $o_{mp}$ 는 입력패턴 p에 대한 입력층이나 혹은 은닉층이 여러개이면 그 아래층 뉴런 m의 출력이다.

따라서 다층신경망에서 확률적 근사법과 후향전파 알고리즘을 이용하여 학습성능을 개선하기 위해 제안한 학습 알고리즘은 다음과 같이 정리될 수 있다.

단계 1 : 다층신경망의 에너지함수  $E(W)$  및 학습 패턴 p에 대한 평균자승 오차함수  $E_p(W)$ 를 정의한다.

단계 2 : 출력층 및 은닉층의 평활된 함수  $E_p(W, \beta)$ 와 그 기울기  $\nabla_w E_p(W, \beta)$ 를 각각 계산한다.

단계 3 : 분산제어 파라미터  $\beta$ 를 크게 하여 출력층과 은닉층의 평활된 함수의 기울기를 계산하여 확률적 근사법을 수행한다.

단계 4 : 단계 3에서 구해진 각 층의 연결가중치를 초기값으로 하여 후향전파 알고리즘을 수행한다.

단계 5 : 후향전파 알고리즘의 수행으로 구해지는 연결가중치 및 오차 에너지 함수가 허용치 이하이면 종료하고, 아니면 단계 3으로 간다.

이와 같이 제안된 방법에서는 확률적 근사법을 이용하여 초기 연결가중치를 설정하고 설정된 연결가중치를 초기값으로 하여 후향전파 알고리즘을 사용한다. 따라서 기존의 후향전파 알고리즘이 그 초기값에 따라 국소최적해에 수렴되는 문제를 해결할 수 있으며, 기울기 하강법의 빠른 수렴속도도 살릴 수 있다. 만일 확률적 근사법과 후향전파 알고리즘을 적용하여 구해지는 해가 확률적 근사법의 1회 적용에 의한 초기치 설정에 따라 국소최적해로 빠질 경우에는 알고리즘 단계 3에서 단계 5까지가 반복되어 결국 전역최적해에 도달하게 된다.

한편, 기존의 후향전파 알고리즘에서는 학습 파라미터중 학습율  $\eta$ 와 모멘트  $\alpha$ 는 초기에 설정된 값이 변하지 않고 학습과정 동안에 동일한 값으로 유지되며, 빠른 학습속도를 얻기 위한 적절한 상호관계를 가지는  $\eta$ 와  $\alpha$  값의 설정이 어렵다. 따라서 연결가중치의 경신과정에서 그 수렴특성에 영향을 미치는 인수를 찾아  $\eta$ 와  $\alpha$ 를 조정하면 그 특성을 개선할 수 있을 것이다. 즉, 초기에 무작위로 설정한  $\eta$ 와  $\alpha$ 를 후향전파 알고리즘의 초기단계에서는 크게하고 점차 전역최적해에 가까워질수록 감소시키면 전역최적해로의 수렴속도를 개선시킬 수 있을 것이다. 후향전파 알고리즘에서 연결가중치경신에 영향을 미치는 인수는 평균자승 오차함수  $E_p(W)$ 의 기울기  $\nabla_w E_p(W)$ 로  $E_p(W)$ 의 미분값이다. 여기서 그 미분값은 각 층 뉴런 활성화분포의 변화량에 따라 달라지며 그 변화량은 다시 각 층뉴런 활성화분포의 표준편차  $\sigma$ 에 의존한다. 그러므로 상기 제안된 방법의 단계 4에서 구해진 각 층의 뉴런 활성화 분포에 대한 표준편차

$\sigma$  를 계산하고, 학습을  $\eta$  와 모멘트  $\alpha$  를 계산된 표준편차  $\sigma$  에 따라 다음과 같이 적응 조정한다. 즉,

$$\begin{aligned} \eta(k) &= \eta_{\text{old}} + \sigma(k) \\ \alpha(k) &= \alpha_{\text{old}} + \sigma(k) \end{aligned} \quad (8)$$

로 한다. 여기서  $\eta$  와  $\alpha$  는 각각 초기학습율과 초기 모멘트이며,  $k$  는 반복횟수이다. 학습의 초기단계에서는 초기학습율  $\eta$  와 초기모멘트  $\alpha$  보다 큰 값의  $\eta$  와  $\alpha$  를 이용하고, 학습이 반복될 수록 이들은 적응적으로 점차 감소하게 된다. 이때 출력층 학습 파라미터는 출력층 뉴런 활성화분포의 표준편차에, 은닉층 학습 파라미터는 은닉층 뉴런 활성화분포의 표준편차에 각각 독립적으로 적응 조정되도록 함으로써 좀 더 후향전과 알고리즘의 오차 전파기법을 직접적으로 활용할 수 있도록 하였다. 그러나 출력층 뉴런의 수가 1 개인 경우에 대해서는 1 개 변수인 경우에 해당되어  $\sigma$  가 0이 되므로 기존과 같이 출력층 학습 파라미터의 경신에 은닉층 뉴런 활성화 분포의 표준편차를 그대로 이용하면 된다.

### III. 응용 예 및 시뮬레이션 결과고찰

제안된 방법의 타당성과 학습성능을 확인하기 위한 응용실험 대상으로 패리티 검사와 패턴 분류를 택하였다. 패리티 검사는 가장 유사한 패턴들이 서로 다른 출력들을 요구하는 특징을 가진 여러개의 국소최적해가 존재하는 문제 중의 하나이다. 이 문제는 출력층 뉴런의 수가 1 개인 구조의 신경망을 이용하며, XOR 문제는 2 비트 패리티 검사의 대표적인 문제에 속한다. 또한 패턴 분류는 학습을 통하여 불안정한 입력패턴을 분류해 내는 특징을 가진 문제로서 영상이나 음성 등의 분류에 널리 응용되고 있다. 이 문제에서는 출력층 뉴런의 수가 학습패턴의 갯수와 동일한 구조의 신경망이 이용된다. 이 두문제들은 신경망을 이용한 학습에서 알고리즘의 성능을 평가하는 대표적인 문제로 이용되고 있다.<sup>7</sup> 따라서 제안된 방법들이 이들 문제에 각각 적용하여 그 타당성을 확인하였으며, 무작위로 설정한 초기의 연결가중치를 이용하는 기존의 후향전과 알고리즘, Baba의 MROM 및 Sun 방법의 결과와 그 성능을 비교 고찰하였다.

학습에 이용된 다층신경망의 구조로는 층 사이의 뉴런간에 완전한 연결을 갖는 입력층, 은닉층 및 출력층으로 구성된 삼층 구조를 택하였다. 초기 연결가중치들의 범위는 -0.5 와 +0.5 사이의 값으로 하였으며 무작위씨앗(random seed)의 변경으로 연결가중

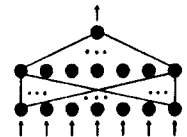
치들을 변화시킬 수 있도록 하였다. 알고리즘의 종료는 계산반복 횟수가 13000 번 이상이거나, 전체 오차함수의 값이 주어진 값 PEV(permissible error value) 이하일 때로 하였다. 여기서 계산 반복횟수(number of iterations)는 모든 입력패턴이 연결가중치를 경신하기 위해서 한 번씩 다 이용된 때를 1 회로 하였다. 특히 후향전과 알고리즘이나 제안된 방법에서 초기 학습율  $\eta$  와 초기 모멘트  $\alpha$  는 각각 그 조합으로 실험한 결과 중에서 가장 우수한 조합의 값으로 설정하였다. 제안된 방법에서 분산제어 파라미터  $\beta$  의 값은 3.0 으로 하여 확률적 근사법을 1 회 적용하였으며, 무작위 교란벡터  $\gamma$  를 생성하는 평활화 함수  $h(\gamma)$  로는 균일 확률밀도함수를 이용하였다. 앞으로의 설명에서 제안된 방법 1은 확률적 근사법과 후향전과 알고리즘을 조합한 방법을 말하며, 제안된 방법 2는 방법 1에서 후향전과 알고리즘의 학습 파라미터를 적응 조정한 경우를 말한다.

#### 1. 패리티 검사

패리티 검사는 각 학습패턴 내에 포함된 "1" 의 갯수에 따라 출력을 결정하는 문제로서, 실험에서는 일부분(partial) 및 우수(even) 패리티로 "1" 의 갯수가 짝수 개이면 "1", 홀수 개이면 "0" 이 출력되도록 학습시켰다.

실험은 학습패턴의 수와 크기가 각각 7과 10인 그림 2에서와 같은 문제를 대상으로 하였으며, 허용 오차함수 PEV는 0.0001로 하였다.

Training patterns	Desired outputs
0000000	1
0100000	0
0000010	0
1010000	1
0000101	1
0101000	1
0001010	1
1010100	0
0010101	0
1111111	0



(a)

(b)

그림 2. 10 개의 7 비트 패턴에 대한 패리티 검사 문제의 (a) 학습패턴과 (b) 신경망의 구조  
Fig. 2. The training patterns for 10 patterns of (a) 7-bit parity check and (b) it's neural network structure.

그림 3은 그림 2의 학습 문제에서 무작위씨앗을 8. 초기 학습율  $\eta$  와 초기 모멘트  $\alpha$  를 각각 0.5와 0.9

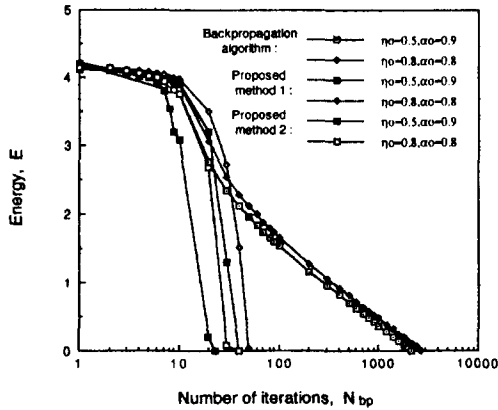


그림 3. 초기 학습율  $\eta_0$ 와 초기 모멘트  $\alpha_0$ 에 따른 10 개의 7 비트 패턴에 대한 패리티 검사 문제의 에너지함수 E

Fig. 3. Energy function E of the parity check for the 10 of the 7-bit patterns for each initial learning rate  $\eta_0$  and momentum  $\alpha_0$ .

및 0.8과 0.8로 하여 후향전과 알고리즘, 제안된 방법을 각각 적용할 때 반복횟수  $N_{bp}$ 에 따른 에너지함수 E의 변화를 나타낸 것이다. 그림에서 보는 바와 같이 후향전과 알고리즘, 제안된 방법 1 및 방법 2의 수렴

속도는 초기 학습율 및 초기 모멘트에 의존한다. 이는 제안된 방법들도 학습 알고리즘으로 후향전과 알고리즘을 이용하기 때문이다. 또한 제안된 방법들에서는 후향전과 알고리즘보다 최적해로의 수렴속도가 빠르며, 제안된 방법 2의 수렴속도는 제안된 방법 1의 수렴속도보다도 더 빠름을 알 수 있다. 이는 제안된 방법 2에서는 학습 파라미터를 적응 조정하면서 학습시키기 때문이다

표 1은 그림 2의 패리티 검사 문제에 대해 학습 파라미터와 무작위씨앗을 변화시키면서 실험한 결과들을 나타낸 것이다. 그 결과는 각 학습 파라미터에 대해 100 개의 서로 다른 초기 연결가중치로 실험한 평균값이다. 표에서  $N_s$ ,  $N_m$  및  $N_j$ 는 각각 확률적 근사법, Baba의 MROM 및 Sun 방법의 반복횟수이며,  $t_{bp}$ ,  $t_m$  및  $t_j$ 는  $N_{bp}$ ,  $N_m$  및  $N_j$ 에 각각 소요된 CPU 시간이다.  $t_{p1}$ 과  $t_{p2}$ 는 각각 제안된 방법 1과 2에서의  $N_{bp}$ 와  $N_s$ 에 각각 소요된 CPU 시간의 합이다. E는 종료시의 에너지함수 값이다. 또한 Baba의 MROM과 Sun 방법의 결과에 있는  $h_{var}$ ,  $m_{var}$  및  $o_{var}$ 은 각각 첫 번째와 두 번째은닉층 및 출력층의 연결가중치를 계산할 때에 요구되는 가우스 교란함수의 초기 분산값들이다.

표에서 나타난 바와 같이 제안된 방법 1과 2는 후향전과 알고리즘보다 그 수렴속도가 빠름을 알 수 있다. Baba의 MROM에서는 각 층의 연결가중치를 경신할 때마다 가우스 분포의 무작위교란을 인가하는

표 1. 10 개의 7 비트 패턴에 대한 패리티 검사의 실험결과.

Table 1. Results of the parity check for 10 of the 7-bit patterns.

$\eta_0, \alpha_0$	BP algorithm			Proposed method 1			Proposed method 2		
	$N_{bp}$	$E_t$	$t_{bp}$	$N_s, N_{bp}$	$E_t$	$t_{p1}$	$N_s, N_{bp}$	$E_t$	$t_{p2}$
1.0, 0.0	11839	$E < 0.0001$	4672.0	3, 182	$E < 0.0001$	74.3	3, 155	$E < 0.0001$	70.5
0.8, 0.8	2698	$E < 0.0001$	1083.5	2, 52	$E < 0.0001$	28.0	2, 41	$E < 0.0001$	26.8
0.3, 0.7	11196	$E < 0.0001$	6404.5	1, 182	$E < 0.0001$	107.5	1, 125	$E < 0.0001$	84.8
0.5, 0.9	2162	$E < 0.0001$	792.6	2, 41	$E < 0.0001$	21.8	2, 28	$E < 0.0001$	17.2

$h_{var}=m_{var}$ $=o_{var}$	Baba's MROM			$h_{var},$ $o_{var}$	Sun's method		
	$N_m$	$E_t$	$t_m$		$N_j$	$E_t$	$t_j$
0.05	496	$E < 0.0001$	674	0.1, 0.1	152	$E < 0.0001$	172
0.1	263	$E < 0.0001$	493	0.5, 0.5	146	$E < 0.0001$	165
0.25	185	$E < 0.0001$	252	1.0, 1.0	161	$E < 0.0001$	182
0.5	213	$E < 0.0001$	289	1.75, 1.75	131	$E < 0.0001$	148
0.75	216	$E < 0.0001$	294	2.5, 1.5	135	$E < 0.0001$	153
1.0	161	$E < 0.0001$	218	3.0, 1.75	202	$E < 0.0001$	229

$t_j$  : CPU time in [sec]

계산이요구되므로 최적해로의 수렴시간이 오래 걸리며, 그 수렴속도는 초기 분산값에 의존한다.

Sun의 방법에서는 1 개의 은닉층을 가진 구조로도 전역최적해로의 수렴이 가능하나 여기서도 무작위교란을 인가하는 계산이 요구되어 그 수렴속도가 늦음을 알 수 있다. 제안된 방법 2에서는 학습 중에적응적으로 학습 파라미터를 조정하는 계산시간이 추가 요구됨에도 불구하고 제안된 방법 1보다는 그수렴속도가 약 1.2 배 정도 개선되었다. 표에서 나타난 바와 같이 제안된 방법 1의 수렴속도는 후향전과 알고리즘에 비해 약 56 배 정도 개선되었으며, 또한 Baba의 MROM이나 Sun 방법의 수렴속도에 비해서는 각각 약 6.4 배 및 약 3 배 정도 개선되었다

표 2에서는 20 개의 7 비트 패턴에 대한 패리티 검사에 있어서 후향전과 알고리즘과 제안된 방법 1을 각각 적용하여 무작위씨앗을 다르게 하여 100 번씩 시도한 실험결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 각각 평균  $\bar{x}$ 와 표준편차  $\sigma$ 를 나타낸 것이다. 여기서 초기 학습율  $\eta_0$ 와 초기 모멘트  $\alpha_0$ 는 각각 0.5와 0.9로 하였다. 표에서 보는 바와 같이 제안된 방법 1에서는 후향전과 알고리즘에 비하여 최적해로의 수렴확률이 약 2.2 배 정도 개선되었으며, 수렴속도에서는 약 16.8 배 정도 개선되었다. 또한 제안된 방법의 표준편차는 후향전과 알고리즘의 표준편차보다 약 2 배 정도적은 값이다. 이는 후향전과 알고리즘에 비해 제안된 방법은 그 성능이 초기 연결가중치에 덜 의존함을 보여준다.

표 2. 20 개의 7 비트 패턴에 대한 패리티 검사의 100 번 시도에 따른 실험결과

Table 2. Results for the 20 of the 7-bit patterns parity check of 100 trials.

	BP algorithm		Proposed method	
	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
N <sub>bp</sub>	7662.5	800.9	457.2	348.5
t <sub>bp</sub>	7018.9	733.6	418.8	319.2
P <sub>r</sub>	45%		100%	

$\bar{x}$  : Mean,  $\sigma$  : Standard deviation

P<sub>r</sub>: Convergence ratio

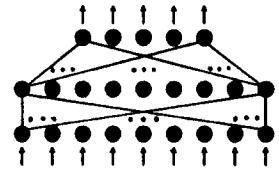
### 2. 패턴 분류

패턴 분류에서는 각 학습패턴에 대해 해당되는 출력 뉴런만 "1"이 되고, 나머지 출력 뉴런들은 "0"을 출력하도록 학습시킨다. 실험에서는 첫 번째 학습패턴에 대

해 첫 번째 출력 뉴런만 "1"을출력하고 나머지 출력 뉴런은 "0"을 출력하는 즉, 각 패턴의 입력순서와 같은 번호에 해당되는 출력뉴런만 반응하도록 학습시켰다.

실험은 그림 4에서와 같이 학습패턴의 크기가 9 비트인 5 개의 패턴에 대하여 실시 하였으며, 학습의 종료는 그 반복횟수가 13000 번 이상이거나, 오차함수 값이 설정된 오차값 PEV = 0.08 이하일 때, 혹은 각 출력 뉴런들이 원하는 출력을 나타낼 때로 하였다.

Training patterns
111001001(7)
100100111(2)
111100111(7)
111111111(8)
111101111(8)



(a)

(b)

그림 4. (a) 5 개의 9 비트 학습패턴과 (b) 이 패턴 분류 문제를 풀기 위한 신경망의 구조

Fig. 4. 5 of the 9-bit training patterns (a) and it's neural network structure (b) for pattern classification.

그림 5는 초기 학습율  $\eta_0$ 와 모멘트  $\alpha_0$ 가 0.3과 0.7일 때에 무작위씨앗을 각각 0과 8의 두 가지에 대하여 실험한 에너지함수의 변화를 나타낸 것이다. 그

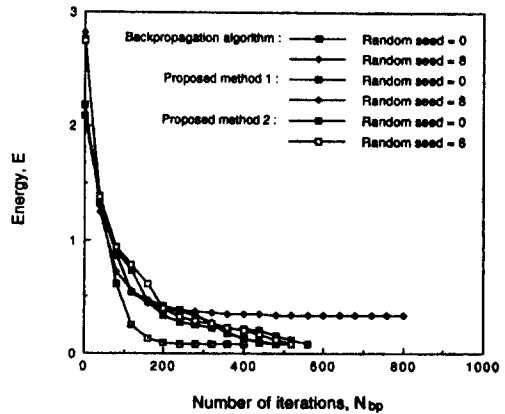


그림 5. 무작위씨앗의 변화에 따른 5 개의 9 비트 패턴에 대한 패턴 분류 문제의 에너지함수 E (초기학습율  $\eta_0 = 0.3$ , 초기모멘트  $\alpha_0 = 0.7$ )

Fig. 5. Energy function E of pattern classification to random seed variation at the learning rate  $\eta_0 = 0.3$  and the initial momentum  $\alpha_0 = 0.7$ .

림에서 보면 무작위씨앗이 8인 경우는 제안된 방법들이 후향전과 알고리즘보다도 수렴속도가 빠르며, 특히 제안된 방법 2의 수렴속도는 제안된 방법 1의 수렴속도보다도 더 빠름을 확인할 수 있다. 이 결과는 무작위씨앗의 단 두 가지에 대해서만 얻은 것으로 평균적인 특징차이는 보기 어렵다.

표 3은 3 개의 학습 파라미터 조합 각각에 대해 무작위씨앗을 4 가지로 변화시켜 가며 실험한 결과를 나타낸 것이며, \*는 제약조건을 만족하지 못하는 결과이다. 표에서 나타난 것과 같이 후향전과 알고리즘에서는 3 개의 학습 파라미터 모두에 대해 무작위씨앗이 3일 때는 그 종료조건을 만족하지 못하였으므로 원하는 출력을 얻을 수 없었다. 이와 같은 상황이 발생하는 것은 초기 연결가중치가 국소최적해 근처에 설정된 경우로, 기존의 후향전과 알고리즘은 그 동작 특성상 가장 가까운 국소 최적해로 수렴하기 때문으로 추측된다. 또한 무작위씨앗이 0과 5인 경우에는 학습율과 모멘트에 따라 학습시간이 줄어들고 패턴을 분류해 낼 수 있는 최적해로도 수렴될 수 있어, 학습 파라미터가 학습성능에 크게 영향을 미침을 확인할 수 있다. 그러나 제안된 방법 1과 2에서는 4 개의 무작위씨앗 모두에 대하여 패턴을 분류해 낼 수 있는

최적해로 수렴되었다. 표에서는 제안된 방법 1과 2가 후향전과 알고리즘에 비해 최적해로의 수렴확률은 약 2.4 배 정도 개선되었고, 그 수렴속도는 각각 약 2.7 배 정도와 약 3.2 배 정도 개선되었다. 또한 제안된 방법 1보다는 학습 파라미터의 적응 조정이 가능한 제안된 방법 2가 평균적으로 약 1.2 배 정도 더 빠른 수렴속도를 갖는다.

IV. 결론

본 논문에서는 확률적 근사법과 후향전과 알고리즘을 조합함으로써 다층신경망의 학습성능을 개선할 수 있는 효율적 방법을 제안하였다. 제안된 방법에서는 학습에 확률적 근사법과 후향전과 알고리즘을 혼합사용한 전역최적화 기법을 적용함으로써 학습성능을 최대한 개선할 수 있도록 하였다. 확률적 근사법은 국소최소점을 벗어나 전역최적점에 치우친 근사점을 결정해 주는 기능을 하도록 하고, 이점을 초기값으로 하여 결정론적 기물이 하강법인 후향전과 알고리즘을 적용함으로써 빠른 수렴속도로 전역최적화가 가능하도록 하였다. 제안된 방법에서 은닉층과 출력층의 학습율과 모멘트를 각 층의 뉴런 활성화분포의 표준편

표 3. 5 개의 9 비트 패턴에 대한 패턴 분류의 실험결과

Table 3. Results of the pattern classification for 5 of the 9-bit patterns

$\eta_0,$ $\alpha_0$	Random seed	BP algorithm		Proposed method 1		Proposed method 2	
		$N_{bp}$	$E_t$	$N_s, N_{bp}$	$E_t$	$N_s, N_{bp}$	$E_t$
0.3, 0.7	0	400	0.0795	1, 560	0.0793	1, 520	0.0782
	3	420	0.3233*	8, 460	0.2019	8, 380	0.1937
	5	540	0.3260*	4, 620	0.2122	4, 540	0.1528
	8	820	0.3343	4, 539	0.0798	4, 511	0.0794
0.6, 0.9	0	1520	0.8742*	8, 573	0.0797	8, 568	0.0792
	3	420	1.2252*	8, 260	0.1824	8, 260	0.1970
	5	1660	0.4201	4, 380	0.2024	4, 360	0.2074
	8	2020	0.4621	6, 540	0.2987	6, 560	0.2988
0.8, 0.8	0	1120	0.4328	4, 447	0.0793	4, 263	0.0779
	3	420	0.6960*	4, 440	0.1880	4, 300	0.1878
	5	1100	0.4259	4, 360	0.2055	4, 360	0.2053
	8	2020	0.4685	4, 540	0.3046	4, 220	0.3034

$t_1$  : CPU time in [sec]



차에 직접 적응 조정되도록 함으로써 학습속도를 더욱 개선하였다.

제안된 방법을 20 개의 7 비트 및 10 개의 7 비트 패턴에 대한 패리티 검사와 5 개의 9 비트 패턴에 대한 패턴 분류에 각각 적용하여 그 타당성과 성능을 확인하였다. 20 개의 7 비트 패턴 패리티검사에 있어서 기존의 후향전파 알고리즘에 비해 최적해로의 수렴확률이 약 2 배 정도 개선되었고, 수렴속도는 약 17 배 정도 개선되었다. 또한 10 개의 7 비트 패턴에 대한 패리티 검사에서 제안된 방법은 Baba나 Sun의 방법보다 최적해로의 수렴속도가 약 6 배 및 3 배 정도 개선되었다. 패턴 분류에서도 제안된 방법은 후향전파 알고리즘보다 우수한 학습성능을 가짐을 확인할 수 있다. 특히, 은닉층과 출력층의 학습율과 모멘트를 각 층의 뉴런 활성화분포의 표준편차에 직접 적응 조정하여 다시 약 20% 수렴속도의 개선을 더 얻을 수 있었다.

향후 제안된 혼성방법을 고속병렬 시스템으로 구현하여 규모가 큰 문제에 대한 신속한 최적학습에 대하여 원활한 연구 수단을 마련하는 일이 남아있다.

#### 參 考 文 獻

- [1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation numerical method*, Prentice-Hall, London, pp.1-50, 1989.
- [2] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, John Wiley & Sons, Inc., London, pp. 10-43, 1973.
- [3] J. T. Tou and R. C. Gonzalez, *Pattern recognition principles*, Addison Wesley Pub. Co., London, pp. 158-242, 1974.
- [4] S. Akaho and S. Amari, On the capacity of three-layer networks, "International Joint Conference on Neural Networks, San Diego, vol. 3, pp. 1-6, June 1990.
- [5] A. J. Maren, C. T. Harston, and R. M. Pap, *Handbook of neural computing application*, Academic Press, pp. 107-250, 1990.
- [6] J. A. Freeman and D. M. Skapura, *Neural networks: algorithms, Applications, and programming techniques*, Addison Wesley, London, pp.89-168, 1991.
- [7] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, MIT press, Cambridge, MA., pp. 282-362, 1986.
- [8] R. Pedone and D. Parisi, Learning the learning parameters," *International Joint Conference on Neural Networks*, Singapore, vol.3, pp.2033-2037, Nov. 1991.
- [9] C. H. Chen and R. S. Nutter, Improving the training speed of three-layer feedforward neural nets by optimal estimation of the initial weights," *International Joint Conference on Neural Networks*, Singapore, vol.3, pp.2063-2068, Nov.1991.
- [10] Y. Hirose, K. Yamashita, and S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," *Neural Networks*, vol. 4, no.1, pp. 61-66, 1991.
- [11] S. D. Wang and C. H. Hsu, "A self growing learning algorithm for determining the appropriate number of hidden units," *International Joint Conference on Neural Networks*, Singapore, vol. 2, pp. 1098-1104, Nov. 1991.
- [12] N. Baba, "A new approach for finding the global minimum of error function of neural networks," *Neural Networks*, vol. 2, no. 5, pp. 367-373, 1989.
- [13] J. Sun, W. I. Grosky, and M. H. Hassoun, "A fast algorithm for finding global minima of error functions in layered neural networks," *Proceedings of International Joint Conference on Neural Networks*, Washington D.C., vol. 1, PP. 585 - 588, Jan. 1990.
- [14] Y. H. Cho and H. M. Choi, Improving the training performances of the multilayer neural networks by SAS-based optimal estimation of initial weights," *JTC-CSCC*, Kyungju, pp.

- 475-478, July 1992.
- [15] M. A. Styblinski and T. S. Tang,  
 "Experiments in nonconvex optimization  
 : stochastic approximation with

function smoothing and simulated  
 annealing." *Neural Net works*, vol. 3,  
 no. 4. pp. 467-483, 1990.

---

著 者 紹 介

趙 鏞 鉉(正會員) 第 30卷 B編 第 8號 參照  
 현재 영남전문대학 전자과 교수

崔 興 文(正會員) 第 30卷 B編 第 8號 參照  
 현재 경북대학교 전자공학과 교수