

論文96-33B-4-22

# 후향전과 알고리즘과 동적터널링 시스템을 조합한 다층신경망의 새로운 학습방법

## (A New Training Method of Multilayer Neural Networks Using a Hybrid of Backpropagation Algorithm and Dynamic Tunneling System)

趙 鏞 鉉 \*

(Yong Hyun Cho)

## 요 약

본 논문에서는 후향전과 알고리즘과 동적터널링 시스템을 조합 사용함으로써 다층신경망의 학습성능을 개선할 수 있는 효율적인 방법을 제안하였다. 제안된 방법에서는 빠른 수렴속도로 전역최적화가 가능하도록 하기 위하여 기울기하강법의 후향전과 알고리즘을 사용하고, 국소최적해를 만났을 때 이를 벗어난 다른 연결가중치의 설정을 위하여 터널링 기능을 가지는 결정론적 방법의 동적터널링 시스템을 적용함으로써 빠른 수렴속도로 전역최적해로의 수렴확률을 높였다. 제안된 방법을 패턴인식과 패턴분류에 적용하여 그 타당성과 성능을 확인한 결과 제안된 방법은 초기값을 무작위로 설정하는 기존의 후향전과 알고리즘에 비해 최적해로의 수렴확률과 그 수렴속도가 우수함을 확인할 수 있었다.

## Abstract

This paper proposes an efficient method for improving the training performance of the neural network using a hybrid of backpropagation algorithm and dynamic tunneling system. The backpropagation algorithm, which is the fast gradient descent method, is applied for high speed optimization. The dynamic tunneling system, which is the deterministic method with a tunneling phenomenon, is applied for global optimization. Converging to the local minima by using the backpropagation algorithm, the approximate initial point for escaping the local minima is estimated by applying the dynamic tunneling system. The proposed method has been applied to the parity check and the pattern classification, and the simulation results show that the performance of proposed method is superior to that of backpropagation algorithm with randomized initial point settings.

## I. 서 론

최근 대규모 병렬분산 처리능력을 갖는 신경망을 이용하여 패턴인식<sup>11,12)</sup> 등과 같은 복잡하고 계산집약적인 과제들을 능률적으로 풀기위한 연구가 활발하게 진행되고 있다<sup>13,12)</sup>.

신경망을 학습에 이용할 때에는 인식이나 분류의 능력을 높이기 위하여 한 개이상의 은닉층(hidden layer)을 가지는 다층(multilayer) 구조를 흔히 이용한다<sup>13)</sup>. 이때 널리 이용되는 후향전과 알고리즘(back propagation algorithm)은 기울기하강(gradient descent)법에 기초한 결정론적(deterministic) 방법으로 최적해로의 수렴속도는 빠르나 국소최적해를 만났을 때 이를 벗어나기가 어렵다<sup>14-16)</sup>. 이러한 후향전과 알고리즘에서는 학습률(learning rate)과 모멘트(mo-

\* 正會員, 嶺南專門大學 電子科

(Dept. of Electronics, Yeungnam Junior College)

接受日字:1995년11月9日, 수정완료일:1996년3월21日

mentum) 등의 학습파라미터, 초기 연결가중치(synapse weight), 또는 신경망의 구조 등에 따라 그 성능이 달라진다.

Rumelhart 등<sup>[6]</sup>은 후향전과 알고리즘에서 연결가중치를 경신할 때에 학습률과 모멘트를 변화시켜 그 학습성능을 개선하였으며, 이 때에 학습률과 모멘트는 경험적으로 설정하였다. Jacobs<sup>[7]</sup>는 경험적으로 학습률을 적응조정함으로써 학습속도를 개선시키는 delta-bar-delta 알고리즘을 제안하였다. 한편, Chen 등<sup>[8]</sup>은 삼층전향 신경망에서 선형대수학에 근거하여 초기의 연결가중치를 설정함으로써 학습속도를 개선하였다. 그러나 이 연구들에서는 국소최적해로의 수렴 가능성을 여전히 남아있다.

Hirose 등<sup>[9]</sup>은 은닉층의 뉴런수를 에너지함수의 증가에 적응조정하였고, Wang 등<sup>[10]</sup>은 자기성장학습(self growing learning) 알고리즘을 제안하여 전역최적해로의 수렴확률을 개선하였다. 이들 방법에서는 은닉층 뉴런의 추가나 삭제에 따른 알고리즘 및 시스템의 복잡도가 증가하였다.

한편, Baba<sup>[11]</sup>는 수정된 무작위 최적화 방법(modified random optimization method)을 제안하여 패턴분류에 이용하였다. 이 방법에서는 연결가중치의 절대값이 한정된 범위 내에 있을 때만 전역최적해로 수렴되며, 두개 이상의 은닉층을 필요로 하고 가우스 분포함수의 편차에 따라서 그 수렴속도가 크게 달라진다. Cho 등<sup>[12]</sup>은 확률적 근사법과 후향전과 알고리즘을 혼합 사용함으로써 학습성능을 개선할 수 있도록 하였다. 이 방법에서는 확률적 근사법을 이용하여 전역최소점에 치우친 근사 연결가중치를 결정하고, 이 가중치를 초기값으로 하여 후향전과 알고리즘을 적용함으로써 빠른 수렴속도로 전역최적화가 가능하도록 하였다.

이제까지의 연구에서는 전역최적해로의 수렴확률을 개선하거나 또는 학습속도를 개선하기 위한 연구 중에서 어느 하나에 주로 역점을 두었을 뿐, 그 두가지를 동시에 만족시키기 위한 연구는 찾아보기가 힘들다. 또한 전역최적해로의 수렴확률이나 학습속도의 개선을 위하여 취한 방법으로 학습률과 모멘트를 조정하거나 초기 연결가중치를 최적해 가까이에 설정하려는 시도 및 은닉층 뉴런수와 같은 신경망의 구조변경 등을 사용하였다. 그러나 학습 중에 국소최적해를 만났을 때 이를 벗어나므로써 전역최적해로의 수렴확률과 학습속도를 동시에 개선하려는 체계적인 연구는 거의 찾아보

기 힘들다.

본 논문에서는 후향전과 알고리즘과 동적터널링 시스템(dynamic tunneling system)을 조합 사용함으로써 다층신경망의 학습성능을 개선할 수 있는 효율적인 방법을 제안하였다. 제안된 방법에서는 빠른 수렴속도로 전역최적화가 가능하도록 하기 위하여 기울기하강법의 후향전과 알고리즘을 사용하고, 국소최적해를 만났을 때 이를 벗어난 다른 연결가중치의 설정을 위하여 동적터널링 시스템을 적용하였다. 제안된 방법을 패리티검사(parity check)와 패턴분류(pattern classification) 문제에 각각 적용하여 초기값을 무작위로 설정하는 기존의 후향전과 알고리즘을 이용한 결과와 비교 고찰하였다.

## II. 후향전과 알고리즘과 동적터널링 시스템을 이용한 효율적 학습

동적 흐름(flow)과 터널링 기능을 갖는 동적터널링 알고리즘은 다극점함수(multiextremal function)의 전역최소점을 찾는 전역최적화 알고리즘이다<sup>[13]</sup>. 최적화는 다극점함수  $f(\mathbf{x})$ 에 동적최적화 시스템(dynamic optimization system)을 수행하여 국소최소점  $\mathbf{x}^*$ 를 구하고, 구해진  $\mathbf{x}^*$ 에 교란  $\varepsilon$ 을 추가한 다음 동적터널링 시스템을 수행하여  $f(\mathbf{x}^*) \geq f(\mathbf{x}^0)$ 인 새로운 시작점  $\mathbf{x}^0$ 를 구하는 두 시스템을 반복 적용함으로써 이루어진다. 이때 교란  $\varepsilon$ 의 값은 무작위로나 혹은 결정론적으로 설정된다. 여기서 다극점함수  $f(\mathbf{x})$ 의 국소최소점을 구하는 동적최적화 시스템은

$$\frac{dx_i}{dt} = - \frac{\partial f}{\partial x_i} \quad (1)$$

와 같다. 이 과정은 임의의 초기값으로부터 국소최소점  $\mathbf{x}^*$ 를 찾는 기울기하강법이다. 또한 구해진 국소최소점  $\mathbf{x}^*$ 에 교란  $\varepsilon$ 을 추가하여 시작하는 동적 흐름과 터널링 기능의 동적터널링 시스템은

$$\frac{dx_i}{dt} = - \frac{\partial f / \partial x_i}{\left[ (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \right]^{\lambda}} - k f'(\mathbf{x}^*) \frac{\partial f}{\partial x_i} \quad (2)$$

이다. 여기서 T는 전치를 의미하고  $f'(\mathbf{x})$ 는  $f'(\mathbf{x}) = f(\mathbf{x}) \cdot f(\mathbf{x}^*)$ 로 정의된다. k는 터널링 패널티(tunneling penalty)이고,  $\lambda$ 는  $\partial f / \partial \mathbf{x}$ 의 영(zero)점  $\mathbf{x}^*$ 의 차수

보다 작거나 같은 값이다. 또한  $f^*(z)$ 는 piecewise-linear 함수인 heaviside 함수로 정의한다. 즉,

$$f^*(z) = \begin{cases} z, & \text{if } z > 0 \\ 0, & \text{if } z < 0 \end{cases} \quad (3)$$

이다.

한편, 그림 1은 연속이고 미분 가능한 6 개의 최소점을 가지는 다극점함수  $f(x)$ 에 동적최적화와 동적터널링 시스템을 반복 적용함으로써 전역최소점을 찾는 동적터널링 알고리즘의 기하학적 해석을 나타낸 것이다.

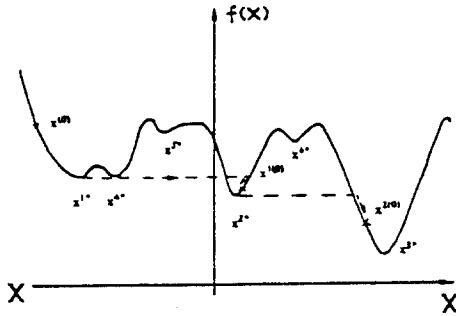


그림 1. 동적터널링 알고리즘의 기하학적 해석  
Fig. 1. The geometric concept of the dynamic tunneling algorithm.

그림에서 보는 바와같이 임의의 값  $x^{(0)}$ 를 초기값으로 하여 함수  $f(x)$ 에 기울기하강법의 동적최적화 시스템을 적용하면 국소최소점  $x^{1*}$ 에 도달하고,  $x^{1*}$ 에 교란  $\epsilon$ 을 추가한 초기값으로 동적터널링 시스템을 적용하면  $x^{(1)}$ 에 수렴하게 된다. 또한  $x^{(1)}$ 를 초기값으로 동적최적화 시스템을 적용하면  $x^{2*}$ 에 도달하고, 동적터널링 시스템에 의하여  $x^{2(0)}$ 를 구하며, 다시  $x^{2(0)}$ 를 초기값으로 하여 동적최적화 시스템에 의해서 전역최소점인  $x^{3*}$ 를 얻을 수 있다. 여기서  $x^{4*}$ 와  $x^{5*}$  그리고  $x^{6*}$ 는 터널링 현상에 의하여 제거된다.

따라서 학습을 위한 다층신경망에서 연결가중치 벡터  $w$ 와 학습패턴  $p$ 에 대한 평균자승 오차함수  $Ep(w)$ 의 최소값을 찾는 과정은 동적터널링 알고리즘으로 상태벡터  $x$ 에 대한 함수  $f(x)$ 의 최소값을 찾는 과정으로 대응시킬 수 있다. 그러므로 다층신경망의 학습에 동적최적화 시스템으로 기울기하강법의 후향전과 알고리즘을 적용하여 평균자승 오차함수의 값이 최소인 연결가중치를 구하고, 구해진 연결가중치에 교란을 추가한 연결가중치를 초기값으로 하여 터널링 기능을 가지는 동적터널링 시스템을 적용하면 빠른 시간내에 전역최소

의 평균자승 오차를 보장하는 학습을 시킬 수 있음을 알 수 있다.

먼저, 다층신경망에서 전체 오차함수  $E(w)$  및 학습 패턴  $p$ 에 대한 평균자승 오차함수  $Ep(w)$ 를 각각

$$E(w) = \sum_p Ep(w)$$

$$Ep(w) = (1/2) \sum_k (d_{pk} - o_{pk})^2 \quad (4)$$

로 정의한다. 여기서  $d_{pk}$ 와  $o_{pk}$ 는 각각  $p$ 번째 패턴에 대한 출력층 뉴런  $k$ 의 원하는 출력과 실제 출력이다. 따라서 평균자승 오차함수  $Ep(w)$ 에 동적터널링 시스템을 적용할 때  $p$ 번째 패턴에 대한 연결가중치  $w$ 의 변화는

$$w(t+1) = w(t) + \eta \Delta_p w(t) + \alpha \Delta_p w(t-1)$$

$$\Delta_p w = - \frac{\partial Ep(w)}{\partial w} \left[ \frac{1}{[(w - w^*)^T (w - w^*)]^2} - k Ep^*(Ep^*(w)) \right] \quad (5)$$

로 구할 수 있다. 이는 연결가중치의 경신식과 더불어서 동적터널링 시스템에 의한 변화량을 나타낸 것이다. 여기서  $\eta$ 와  $\alpha$ 는 각각 학습률(learning rate)과 모멘트(momentum)이고,  $Ep^*(w) = Ep(w) - Ep(w^*)$ 이다. 또한  $Ep^*(z)$ 는  $z > 0$ 이면  $z$ 이고,  $z < 0$ 이면 0의 값을 갖는다. 그리고 출력층과 은닉층 뉴런의 출력이 s자형(sigmoidal) 특성을 가질 때에 동적터널링 시스템에 의한 각 층의 연결가중치 변화는 각각 다음과 같이 계산된다. 즉, 출력층 연결가중치 경신시 연결가중치의 변화는

$$\begin{aligned} \Delta_p w_{kj} = & - \frac{\partial Ep(w)}{\partial w_{kj}} \left[ \frac{1}{(w - w^*)^T (w - w^*)^2} - k Ep^*(Ep^*(w)) \right] \\ & - \frac{\partial Ep(w)}{\partial w_{kj}} = (d_{pk} - o_{pk}) o_{pk} (1 - o_{pk}) i_{pj} \\ & \left| (w_{11} - w_{11}^*)^2 + (w_{12} - w_{12}^*)^2 \right. \\ & \left. + \dots + (w_{kj} - w_{kj}^*)^2 + \dots \right]^{-2} \end{aligned} \quad (6)$$

와 같으며, 은닉층 연결가중치 경신시 연결가중치의 변화는

$$\begin{aligned} \Delta_p w_{ji} = & - \frac{\partial Ep(w)}{\partial w_{ji}} \left[ \frac{1}{(w - w^*)^T (w - w^*)^2} - k Ep^*(Ep^*(w)) \right] \\ & - \frac{\partial Ep(w)}{\partial w_{ji}} = (1 - i_{pj}) i_{pj} x_j \sum_k (d_{pk} - o_{pk}) (1 - o_{pk}) w_{kj} \\ & \left| (w_{11} - w_{11}^*)^2 + (w_{12} - w_{12}^*)^2 + \dots \right. \\ & \left. + (w_{ji} - w_{ji}^*)^2 + \dots \right]^{-2} \end{aligned} \quad (7)$$

와 같이 계산된다. 여기서  $i, j$  및  $k$ 는 각각 입력층, 은닉층 및 출력층의 뉴런번호이다.  $i_{in}$ 는  $p$ 번째 패턴에 대한 은닉층 뉴런  $j$ 의 출력이며,  $x_{pi}$ 는  $p$ 번째 패턴에 대한 입력층 뉴런  $i$ 의 출력이다.

따라서 다층신경망에서 후향전파 알고리즘과 동적터널링 시스템을 이용하여 학습성능을 개선하기 위해 제안한 학습 알고리즘은 다음과 같다.

단계 1: 다층신경망의 전체 오차함수  $E(\mathbf{w})$  및 학습 패턴  $p$ 에 대한 평균자승 오차함수  $E_p(\mathbf{w})$ 를 각각 정의한다.

단계 2: 출력층과 은닉층의 초기 연결가중치를 각각 무작위로 발생시킨다.

단계 3: 후향전파 알고리즘을 수행하여 계산되는 전체 오차함수의 값이 허용치 이하이면 학습을 종료하고, 그렇지 않고 전체 오차함수의 값이 허용치 이상으로 알고리즘의 반복에 따라 변하지 않으면 후향전파 알고리즘을 종료한다.

단계 4: 단계 3에서 계산된 연결가중치에 교란  $\epsilon$ 을 추가하여 동적터널링 시스템을 수행한다.

단계 5: 동적터널링 시스템의 수행으로 계산된 전체 오차함수의 값이 허용치 이하이면 학습을 종료하고, 그렇지 않고 계산된 전체 오차함수의 값이 단계 3의 후향전파 알고리즘에서 계산된 오차함수 값 이하이면 단계 3으로 간다.

여기서 학습 알고리즘의 종료는 전체 오차함수의 값이 이미 설정된 허용치 이하일 때로 한다. 이와같이 제안된 방법에서는 후향전파 알고리즘을 수행함으로써 국소최적해로 수렴될 때에, 교란을 가한 연결가중치를 초기값으로 하여 터널링 기능을 가진 동적터널링 시스템을 실행하게 된다. 따라서 기존의 후향전파 알고리즘이 학습파라미터나 신경망의 구조, 그리고 학습파라미터 등과 같은 요인에 따라 국소최적해로 수렴되는 문제를 해결할 수 있으며, 기울기하강법의 빠른 수렴속도도 그대로 살릴 수 있다. 만약 후향전파 알고리즘과 동적터널링 시스템의 1 회 적용에 의하여 국소최적해로 수렴될 경우에는 제안된 방법의 단계 3에서 단계 5까지의 과정을 반복함으로써 결국 전역최적해로 수렴될 수 있다.

### III. 응용 예 및 시뮬레이션 결과고찰

제안된 방법의 타당성과 학습성능을 확인하기 위한 응용실험 대상으로 패리티검사와 패턴분류를 택하였다. 패리티검사는 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 패리티검사의 대표적인 문제에 속한다. 또한 패턴 분류는 학습을 통하여 불완전한 입력패턴을 분류해 내는 특징을 가진 문제로서 영상이나 음성 분류에 널리 응용되고 있다. 이 문제에서는 출력층 뉴런의 수가 학습패턴의 갯수와 동일한 구조의 신경망이 이용된다. 이 두 문제들은 신경망을 이용한 학습에서 알고리즘의 성능을 평가하는 대표적인 문제로 이용되고 있다<sup>16)</sup>. 따라서 제안된 방법을 이들 문제에 각각 적용하여 그 타당성을 확인하였으며, 무작위로 설정한 초기의 연결가중치를 이용하는 기존의 후향전파 알고리즘의 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 랜덤 시드(random seed)의 변경으로 연결가중치들을 변화시킬 수 있도록 하였다. 알고리즘의 종료는 계산 반복횟수(number of iterations)가 20000 번 이상이거나, 전체 오차함수의 값이 주어진 값 PEV(permissible error value) 이하일 때로 하였다. 여기서 계산 반복횟수는 모든 입력패턴이 연결가중치를 경신하기 위해서 한 번씩 다 이용될 때를 1 회로 하였다. 특히 후향전파 알고리즘이나 제안된 방법에서 학습률  $\eta$ 와 모멘트  $\alpha$ 는 각각 그 조합으로 실험한 결과 중에서 가장 우수한 조합의 값으로 설정하였다.

#### 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 크기가 7 비트인 20개의 패턴을 대상으로 하였으며, 허용오차함수 PEV는 0.0001로 하

였다. 또한 신경망의 구조로 입력층과 은닉층의 뉴런 수는 각각 7 개로 하고 출력층의 뉴런 수는 1 개로 하였다.

그림 2는 랜덤시드를 10, 학습률  $\eta$ 와 모멘트  $\alpha$ 를 각각 0.3과 0.7 및 0.8과 0.8로 하여 후향전파 알고리즘과 제안된 방법을 적용할 때의 반복횟수  $N$ 에 따른 전체 오차함수  $E$ 의 변화를 나타낸 것이다.

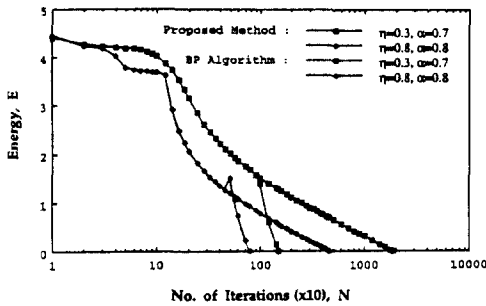


그림 2. 학습률  $\eta$ 와 모멘트  $\alpha$ 에 따른 20 개의 7 비트 패턴에 대한 패리티검사 문제의 오차함수  $E$

Fig. 2. Error function  $E$  of the parity check for the 20 of the 7-bit patterns for each learning rate  $\eta$  and momentum  $\alpha$ .

그림에서 보는 바와 같이 후향전파 알고리즘보다 제안된 방법을 적용할 경우에 최적해로의 수렴속도는 더욱 개선되었고 후향전파 알고리즘과 제안된 방법 둘다 그 수렴속도는 학습률과 모멘트에 의존한다. 이는 제안된 방법에서도 학습 알고리즘으로 후향전파 알고리즘을 이용하기 때문이다. 또한 후향전파 알고리즘에서는 오차함수의 변화가 작은 범위에 수렴될 때에 그 수렴속도가 느려지나 제안된 방법에서는 이 범위에 터널링 효과를 가지는 동적터널링 시스템을 적용함으로써 최적해로의 수렴속도가 빠름을 알 수 있다.

표 1은 20 개의 7 비트 패턴에 대한 패리티검사에 있어서 후향전파 알고리즘과 제안된 방법을 각각 적용하여 랜덤시드를 다르게 하여 100 번씩 시도한 결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 각각 평균  $\bar{x}$ 와 표준편차  $\sigma$ 를 나타낸 것이다. 표에서  $N_{bp}$ ,  $N_{pm}$ , 그리고  $N_{ch}$ 는 각각 후향전파 알고리즘, 제안된 방법, 그리고 동적터널링 시스템의 반복횟수이며,  $t_{bp}$ 와  $t_{pm}$ 은 각각 후향전파 알고리즘과 제안된 방법에 소요된 CPU 시간이다. 특히  $t_{pm}$ 은 후향전파 알고리즘과 동적터널링 시스템에 각각 소요된 CPU 시간의 합이다. 표에서 보는 바와같이 후향전파 알고리즘과 제

안된 방법은 학습률과 모멘트의 값에 따라 최적해로의 수렴확률 및 그 수렴속도가 매우 의존됨을 알 수 있다. 제안된 방법은 후향전파 알고리즘에 비하여 반복횟수의 경우에는 약 6.3 배에서 약 8.8 배 정도 개선되었으며, 학습시간의 경우에는 약 5.2 배에서 약 7.6 배 정도 개선되었다. 또한 제안된 방법은 기존의 후향전파 알고리즘에 비하여 학습률  $\eta$ 와 모멘트  $\alpha$ 의 값이 각각  $\eta = 0.8$ 과  $\alpha = 0.8$  및  $\eta = 0.5$ 와  $\eta = 0.9$ 인 경우에는 비슷한 최적해로의 수렴확률을 보여주나  $\eta = 1.0$ 과  $\alpha = 0.0$  및  $\eta = 0.3$ 과  $\alpha = 0.7$ 일 경우에는 각각 약 10 배 정도 및 약 11.1 배 정도 그 수렴확률이 개선되었다. 여기에서 제안된 방법은 4 가지 학습파라미터의 조합 각각에 대하여 100 번씩 시도한 결과 모두 최적해로 수렴되어 국소최적해로의 수렴문제를 해결할 수 있음을 알 수 있다. 한편 제안된 방법에서 반복횟수에 대한 표준편차는 후향전파 알고리즘의 표준편차에 비하여  $\eta = 1.0$ 과  $\alpha = 0.0$  및  $\eta = 0.3$ 과  $\alpha = 0.7$ 의 경우에는 각각 약 5.0 배 및 약 1.1 배 정도 큰 값이며,  $\eta = 0.8$ 과  $\alpha = 0.8$  및  $\eta = 0.5$ 와  $\alpha = 0.9$ 의 경우에는 각각 약 3.5 배 및 약 2.3 배 정도 적은 값이다. 또한 학습시간에 대한 표준편차는  $\eta = 1.0$ 과  $\alpha = 0.0$ 의 경우를 제외하고는 약 1.6 배에서 약 2.4 배 정도 적은 값이다. 이는 후향전파 알고리즘에 비하여 제안된 방법의 성능이 초기 연결가중치에 덜 의존함을 보여준다.

표 1. 20 개의 7 비트 패턴에 대한 패리티검사의 100 번 시도에 따른 실험결과  
Table 1. Results for the 20 of the 7-bit patterns parity check of 100 trials.

$\eta$	$\alpha$	BP algorithm				Proposed method			
		$N_{\bar{x}}$	$t_{\bar{x}}$	$E$	$P_c$	$N_{ch}(N_{bp}, N_{pm})$	$t_{pm}$	$E$	$P_c$
1.0	x	19721.5	194.3	0.0001	10	2406.4(1822.1,564.3)	29.5	0.000100	100
	o	351.2	1.4	0.0000		1758.2(925.9,885.9)	28.6	0.000001	
0.8	x	6404.0	69.5	0.0001	99	95.7(761.3,204.4)	11.1	0.000099	100
	o	1893.6	18.0	0.0000		544.1(332.3,245.8)	7.6	0.000002	
0.5	x	5094.1	48.2	0.0001	99	802.7(634.1,168.6)	9.2	0.000099	100
	o	1496.3	14.0	0.0000		643.6(347.7,313.5)	9.2	0.000002	
0.3	x	19301.1	190.4	0.0001	9	2203.4(1770.8,432.6)	24.9	0.000100	100
	o	680.3	13.4	0.0000		689.0(571.4,202.0)	8.2	0.000000	

$\bar{x}$  : Mean,  $\sigma$  : Standard deviation,  $t_c$  : CPU time in [sec],  $P_c$  : Convergence ratio [%]

2) 패턴 분류

패턴분류에서는 각 학습패턴에 대해 해당되는 출력 뉴런만 "1"이 되고, 나머지 출력 뉴런들은 "0"을 출력

하도록 학습시킨다. 실험에서는 첫 번째 학습패턴에 대해서 첫 번째 출력 뉴런만 "1"을 출력하고 나머지 출력 뉴런은 "0"을 출력하는 즉, 각 패턴의 입력순서와 같은 번호에 해당되는 출력 뉴런만 반응하도록 학습시켰다.

실험은 학습패턴의 크기가 8 x 8 (64) 비트인 10 개의 패턴에 대하여 실시 하였으며, 학습의 종료는 허용 오차함수 PEV가 0.001 이하일 때로 하였다. 또한 신경망의 구조로 입력층과 은닉층의 뉴런 수는 각각 64 개로 하고 출력층의 뉴런 수는 10 개로 하였다.

그림 3은 학습률  $\eta=0.5$ 와 모멘트  $\alpha=0.9$ 일 때에 랜덤시드를 각각 10, 50, 그리고 90의 세가지에 대하여 실험한 전체 오차함수의 변화를 나타낸 것이다.

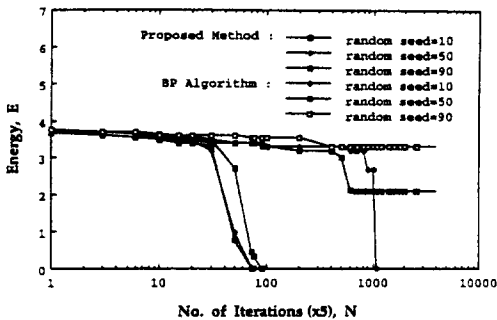


그림 3. 랜덤시드의 변화에 따른 10 개의 64 비트 패턴에 대한 패턴분류 문제의 오차함수 E (학습률  $\eta=0.5$ , 모멘트  $\alpha=0.9$ )

Fig. 3. Error function E of pattern classification to random seed variation at the learning rate  $\eta=0.5$  and the momentum  $\alpha=0.9$ .

그림에서 보면 랜덤시드가 10인 경우에는 제안된 방법과 후향전과 알고리즘 둘다 최적해로 수렴되었으나, 50과 90인 경우에는 제안된 방법만 최적해로 수렴되었다. 이는 후향전과 알고리즘의 경우에 초기 연결가중치의 설정에 따라 그 학습능력이 달라지고, 국소최적해를 만났을 때에 이를 벗어나지 못함을 보여준다. 그림 3에서도 그림 2의 결과와 유사함을 알 수 있다. 이 결과는 랜덤시드의 단 세가지에 대해서만 얻은 것으로 평균적인 특징차이는 보기 어렵다.

표 2는 10 개의 64 비트 패턴에 대한 패턴분류에 있어서 후향전과 알고리즘과 제안된 방법을 각각 적용하여 랜덤시드를 다르게 하여 100 번씩 시도한 결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 각각 평균  $\bar{x}$ 와 표준편차  $\sigma$ 를 나타낸 것이다. 표에

서도 후향전과 알고리즘과 제안된 방법은 학습률과 모멘트의 값에 따라 최적해로의 수렴확률 및 그 수렴속도가 매우 의존됨을 알 수 있다. 표에서 보는 바와같이 제안된 방법은 후향전과 알고리즘에 비하여 반복횟수의 경우에는 약 1.7 배에서 약 6.6 배 정도 개선되었으며, 학습시간의 경우에는 약 1.2 배에서 약 4.2 배 정도 개선되었다. 또한 제안된 방법은 기존의 후향전과 알고리즘에 비하여 최적해로의 수렴확률은  $\eta=1.0$ 과  $\alpha=0.0$  및  $\eta=0.3$ 과  $\alpha=0.7$ 의 경우에는 같으나  $\eta=0.8$ 과  $\alpha=0.8$ 의 경우에는 1.1 배 정도,  $\eta=0.5$ 와  $\alpha=0.9$ 의 경우에는 약 7.1 배 정도 개선되었다. 제안된 방법에서 반복횟수에 대한 표준편차는  $\eta=1.0$ 과  $\alpha=0.0$ 의 경우에는 후향전과 알고리즘의 표준편차보다 약 2.8 배 정도 큰 값이나,  $\eta=0.8$ 과  $\alpha=0.8$ ,  $\eta=0.5$ 와  $\alpha=0.9$ , 그리고  $\eta=0.3$ 과  $\alpha=0.7$ 의 경우에는 각각 후향전과 알고리즘의 표준편차보다 약 10.1 배, 약 22.5 배, 그리고 약 6.6 배 정도 적은 값이다. 또한 학습시간에 대한 표준편차도 3 가지 학습파라미터의 경우에 각각 약 6.3 배, 약 11.6 배, 그리고 약 3.6 배 정도 적은 값이다. 여기서도 후향전과 알고리즘에 비하여 제안된 방법의 성능이 초기 연결가중치에 덜 의존함을 보여준다.

표 2. 10 개의 64 비트 패턴에 대한 패턴분류의 100 번 시도에 따른 실험결과

Table 2. Results for the 10 of the 64-bit patterns pattern classification of 100 trials.

$\eta$	$\alpha$	$\bar{x}$	BP algorithm				Proposed method			
			$N_{tr}$	$t_{tr}$	E	P <sub>c</sub>	$N_{opt}(N_{tr}, N_{in})$	tpm	E	Pr
1.0	0.0	x	2219.1	700.9	0.001000	100	1288.6(782.7,465.9)	571.0	0.000998	100
		$\sigma$	59.0	28.9	0.000000		167.8(84.3,83.6)	84.0	0.000006	
0.8	0.8	x	1249.3	369.6	0.000998	95	360.0(291.2,68.8)	133.9	0.000964	100
		$\sigma$	1020.6	302.1	0.000002		101.4(80.6,59.2)	47.9	0.000045	
0.5	0.9	x	2826.0	826.8	0.000997	14	425.1(235.0,190.1)	197.6	0.000983	100
		$\sigma$	2131.0	624.1	0.000004		94.8(34.3,68.3)	53.6	0.000005	
0.3	0.7	x	2222.6	683.2	0.001000	100	443.8(346.7,97.1)	174.7	0.000982	100
		$\sigma$	60.4	20.0	0.000000		9.2(7.1,2.7)	5.5	0.000006	

x : Mean,  $\sigma$  : Standard deviation,  $t_r$  : CPU time in [sec], P<sub>c</sub> : Convergence ratio [%]

#### IV. 결론

본 논문에서는 후향전과 알고리즘과 동적터널링 시

시스템을 조합 사용함으로써 다층신경망의 학습성능을 개선할 수 있는 효율적인 방법을 제안하였다. 제안된 방법에서는 빠른 수렴속도로 전역최적화가 가능하도록 하기 위하여 기울기하강법의 후향전과 알고리즘을 사용하고, 국소최적해를 만났을 때 이를 벗어난 다른 연결가중치의 설정을 위하여 동적터널링 시스템을 적용함으로써 빠른 수렴속도로 전역최적화가 가능하도록 하였다.

제안된 방법을 20 개의 7 비트 패턴에 대한 패리티 검사와 10 개의 64 비트 패턴에 대한 패턴분류에 각각 적용하여 그 타당성과 성능을 확인하였다. 20 개의 7 비트 패턴에 대한 패리티검사에 있어서 제안된 방법은 기존의 후향전과 알고리즘에 비해 최적해로의 수렴확률은 최대 약 11 배 정도 개선되었으며, 수렴속도에서도 최대 약 8.8 배 정도 개선되었다. 또한 10 개의 64 비트 패턴에 대한 패턴분류에서도 제안된 방법은 기존의 후향전과 알고리즘에 비해 최적해로의 수렴확률은 최대 약 7 배 정도 개선되었고, 수렴속도에서는 최대 약 6.6 배 정도 개선된 우수한 학습성능을 가짐을 확인할 수 있다.

향후 제안된 방법을 좀 더 규모가 크고 다양한 학습 문제 및 조합형 최적화 문제에 확장 적용할 수 있는 수단 마련과 개발된 다른 학습 알고리즘들과도 그 성능을 비교하는 일이 남아 있다.

※ 이 논문은 1995학년도 영남전문대학 연구조성비 지원에 의한 것임.

### 참 고 문 헌

- [1] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, John Wiley & Sons, Inc., London, pp. 10-43, 1973.
- [2] J. T. Tou and R. C. Gonzalez, *Pattern recognition principles*, Addison Wesley Pub. Co., London, pp. 158-242, 1974.
- [3] S. Akaho and S. Amari, "On the capacity of three-layer networks," *International Joint Conference on Neural Networks*, Sandiego, vol. 3, pp. 1-6, June 1990.
- [4] A. J. Maren, C. T. Harston, and R. M. Pap, *Handbook of neural computing application*, Academic Press, pp. 107-250, 1990.
- [5] J. A. Freeman and D. M. Skapura, *Neural networks : algorithms, Applications, and programming techniques*, Addison Wesley, London, pp. 89-168, 1991.
- [6] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, MIT press, Cambridge, MA., pp. 282-362, 1986.
- [7] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Journal of Neural Networks*, vol. 1, pp. 295-307, Nov. 1988.
- [8] C. H. Chen and R. S. Nutter, "Improving the training speed of three-layer feed-forward neural nets by optimal estimation of the initial weights," *International Joint Conference on Neural Networks*, Singapore, vol. 3, pp. 2063-2068, Nov. 1991.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Y. Yao, "Dynamic tunneling algorithm for global optimization," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 19, no. 5, pp. 1222-1230, Sept/Oct. 1989.

## — 저 자 소 개 —



趙鏞鉉(正會員)

1957年 3月 18—生. 1979年 2月 경북대학교 전자공학과 졸업(공학사). 1981年 2月 경북대학교 대학원 전자공학과 졸업(공학석사). 1993年 2月 경북대학교 대학원 전자공학과 졸업(공학박사). 1983年 9月 ~ 1984年 2月 삼성전자(주) 근무. 1984年 3月 ~ 1987年 2月 한국전자통신연구소 근무. 1987年 3月 ~ 현재 영남전문대학 전자과 부교수. 주관심 분야는 신경망, 병렬분산처리, 최적화 및 스위칭네트워크 등임.