

하이브리드 알고리즘을 이용한 신경망의 학습성능 개선

김 원 옥[†] · 조 용 현^{††} · 김 영 일^{†††} · 강 인 구[†]

요 약

본 논문에서는 공액기울기법과 터널링 시스템을 조합사용하여 신경망의 학습성능을 향상시킬 수 있는 효율적인 방법을 제안하였다. 빠른 수렴속도의 학습을 위하여 공액 기울기법에 기초한 후향전파 알고리즘을 사용하였고, 국소최적해를 만났을 때 이를 벗어난 다른 연결가중치의 설정을 위해 동적터널링 시스템에 기초한 후향전파 알고리즘을 조합한 학습 알고리즘을 적용하였다. 제안된 방법을 패리티 검사 및 패턴분류 문제에 각각 적용하여 기존의 기울기 하강법에 기초한 후향전파 알고리즘 및 기울기 하강법과 동적터널링 시스템을 조합한 후향전파 알고리즘방법의 결과와 비교 고찰하여 제안된 방법이 다른 방법들 보다 학습성능에서 우수함을 나타내었다.

Improving the Training Performance of Neural Networks by using Hybrid Algorithm

Weon Ook Kim[†] · Yong Hyun Cho^{††} · Young Il Kim^{†††} · In Ku Kang[†]

ABSTRACT

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

※본 연구는 과학기술처의 지원에 의해서 수행되었음.

† 정 회 원: 연암공업전문대학

†† 정 회 원: 효성카톨릭대학교 전자·정보공학부

††† 정 회 원: 경상대학교 전자공학과

논문접수: 1997년 5월 10일, 심사완료: 1997년 9월 18일

1. 서 론

컴퓨터비전이나 로보틱스 등과 같은 여러 응용분야에 이용되고 있는 순차적 기법의 대안으로 최근 신경망이 이용되고 있다. 이러한 신경망은 단순한 처리 능력을 가지는 뉴런(neuron)들이 상호 연결되어 병렬 분산처리, 일반화 그리고 학습등의 속성을 가지고 있어 NP-copmplete 조합형 최적화 문제나 인식등과 같은 문제들을 해결하는데 많이 이용되고 있다. 그러나 이러한 문제들의 해를 구하기 위하여 신경망을 이용할 때에 아직까지는 그 학습시간이 많이 요구되며, 또한 국소최적해로 수렴되어 정확한 해를 구하지 못하고 있어 그에 대한 연구가 절실한 실정이다.⁽¹⁻⁸⁾ 따라서 이러한 제약점을 해결할 수 있는 효율적인 다층 신경망의 학습방법의 개발이 절실히 요구되며 음성이나 영상신호의 처리에 있어서도 신경망 자체가 가지는 고유한 특성을 그대로 살릴 수 있도록 하는 방법의 연구 또한 매우 필요하리라 생각된다.

신경망을 학습에 이용할 때에는 인식이나 분류의 능력을 높이기 위하여 한 개 이상의 은닉층(hidden layer)을 가지는 다층 구조를 흔히 이용한다.⁽²⁾ 이때 많이 이용되고 있는 후향전파 알고리즘(backpropagation algorithm: BP)은 기울기하강(gradient descent)법에 기초한 결정론적 방법^(9, 10)으로 최적해로의 수렴속도는 빠르나 국소최적해를 만났을 때 이를 벗어나기가 어렵다. 이러한 후향전파 알고리즘에서는 학습률과 모멘트등의 학습 파라미터, 초기 연결가중치 또는 신경망의 구조등에 따라 그 성능이 달라진다.⁽¹⁻⁸⁾

지금까지의 연구에서는 전역최적해로의 수렴확률을 개선하거나 또는 학습속도를 개선하기 위한 연구중에서 어느 하나에 주로 역점을 두었을 뿐, 이 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 구조 변경등을 사용하였다. 그러나 학습중에 국소최적해를 만났을 때 이를 벗어남으로써 전역최적해로의 수렴확률 및 학습속도를 동시에 개선하려는 체계적인 연구는 거의 찾아보기가 힘들다.

본 연구에서는 지도처리 인식과정에서 인식능력을

향상시키기 위한 기존의 후향전파 알고리즘에 비해 더욱 개선된 학습성능을 가진 새로운 학습방법의 다층 신경망을 제안한다. 제안된 방법에서는 빠른 수렴속도의 학습을 위하여 공액 기울기법(conjugate gradient method)⁽¹²⁻¹⁴⁾에 기초한 후향전파 알고리즘을 사용하였고, 국소최적해를 만났을 때 이를 벗어난 다른 연결가중치의 설정을 위해 동적터널링 시스템(dynamic tunneling system)⁽¹¹⁾에 기초한 후향전파 알고리즘(conjugate gradient and dynamic tunneling system: CGDTS)을 조합한 학습 알고리즘을 적용하였다. 제안된 방법을 패리티 검사 및 패턴분류 문제에 각각 적용하여 기존의 기울기 하강법에 기초한 후향전파 알고리즘 및 기울기 하강법과 동적터널링 시스템을 조합한 후향전파 알고리즘(backpropagation and dynamic tunneling system: BPDTS)방법의 결과와 비교 고찰하였다.

2. 공액 기울기법의 후향전파 알고리즘과 동적터널링 시스템의 후향전파 알고리즘을 조합한 다층 신경망의 효율적 학습

2.1 공액 기울기법

기존의 순차적 최적화 방법중에서 수렴속도가 빠른 기법중에 하나인 공액 기울기법은 선형이나 비선형 방정식의 해를 구하는 문제에 적용될 수 있다. 일반적인 선형방정식

$$Tx(t) = b, t = 0, 1, 2, \dots \quad (1)$$

에서 $x(t)$ 와 b 는 $n \times 1$ 벡터이고, T 는 $n \times n$ 행렬로 표현되는 방정식 계수로서 positive definite 이며 invertible 하다고 가정할 때 이를 풀기 위한 비용함수 $f(x(t))$ 는

$$f(x(t)) = -\frac{1}{2} x(t)' Tx(t) + x(t)' b \quad (2)$$

로 정의 될 수 있다. 여기서 '는 전치행렬을 의미한다. 식 (2)에서 T 가 positive definite이면 $f(x(t))$ 는 완전 오목함수(strictly convex function)로 오직 하나의 최소값을 갖게 된다. 또한 $Tx(t) = c (c = 2b)$ 이면 식 (2)는 영의 값을 가져 선형방정식을 푸는 문제는 비용함수 $f(x(t))$ 를 최소화하는 문제로 변환될 수 있다. 그러므

로 $\nabla f(x(t))^* = 0$ 이면 $x(t)^*$ 는 비용함수를 최소로 한다. 즉, 식 (2)를 미분하면

$$\nabla f(x(t)) = \frac{\partial f(x(t))}{\partial x(t)} = -Tx(t) + b = 0 \quad (3)$$

가 되어 $x(t)$ 를 구하면, $Tx(t) = b$ 의 최적해가 된다. $x(t+1)$ 은 반복에 의하여

$$x(t+1) = x(t) + \gamma(t)s(t) \quad (4)$$

로 구해진다. 여기서 t 는 반복수이며 $s(t)$ 는 update의 방향이고 $\gamma(t)$ 는 $f(x(t) + \gamma s(t)) = \min_{\gamma \in R} f(x(t) + \gamma s(t))$

에 의해 정의되는 스칼라 step size이다. 이때 $s(t)$ 와 $\gamma(t)$ 는

$$g(t) = \nabla f(x(t)) = -Tx(t) + b \quad (5)$$

$$g(t-1) = \nabla f(x(t-1)) = -Tx(t-1) + b \quad (6)$$

$$s(t) = -g(t) + q(t)s(t-1) = Tx(t) - b + q(t)s(t-1) \quad (7)$$

$$s(0) = -g(0) \quad (7)$$

$$q(t) = \frac{(g(t)'g(t))}{(g(t-1)'g(t-1))} \quad (8)$$

$$\gamma(t) = -\frac{(s(t)'g(t))}{(s(t-1)' \Gamma s(t-1))} \quad (9)$$

에 의해 구해지며 식 (4)에서 (9)까지의 과정을 공액 기울기법이라 한다. 여기서 Γ 는 Hessian matrix이다. 공액 기울기법에서는 비용함수값이 단조 감소하는 결정론적 동작특성으로 기울기 하강법에 비해 수렴 속도는 현저히 증가되나 기울기 하강법이 가지는 국소최소해로의 수렴은 그대로 나타난다.⁽¹²⁻¹⁴⁾

2.2 동적터널링 시스템

동적 흐름(dynamic flow)과 터널링 기능을 갖는 동적터널링 알고리즘은 다극점함수(multixtremal function)의 전역최소점을 찾는 전역최적화 알고리즘이다.⁽¹¹⁾ 최적화는 다극점함수 $f(x)$ 의 국소최소점 x^* 를 구하는 동적최적화 시스템과 국소최소점에서 x^* 에 교란 ϵ 을 추가하여 $f(x^*) \geq f(x^0)$, $x^* \neq x^0$ 인 새로운 시작점 x_0 를 구하는 동적터널링 시스템으로 이루어져 있으며 이 두 시스템을 반복 적용함으로써 전역최소

점을 구할 수 있다.

여기서 다극점함수 $f(x)$ 의 국소최소점을 구하는 동적최적화 과정은

$$\frac{dx_i}{dt} = -\frac{\partial f}{\partial x_i} \quad (10)$$

와 같다. 이 과정은 임의의 초기값으로부터 국소최소점 x^* 를 찾는 기울기 하강법이다. 그러나 기울기 하강법 대신 앞 절에서 설명한 더욱 빠른 수렴속도를 가지는 공액 기울기법을 적용할 수 있다.

또한, 구해진 국소최소점 x^* 에 교란 ϵ 을 추가하여 시작하는 동적터널링 과정은

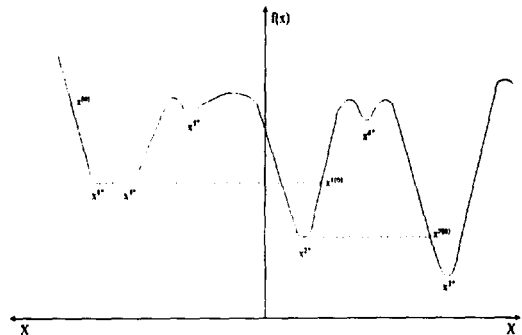
$$\frac{dx_i}{dt} = -\frac{\partial f / \partial x_i}{[(x-x^*)'(x-x^*)]^\lambda} - Kf^*(\hat{f}(x)) \frac{\partial f}{\partial x_i} \quad (11)$$

이다. 여기서 '는 전치를 의미하고 $\hat{f}(x)$ 는 $\hat{f}(x) = f(x) - f(x^*)$ 로 정의된다. K 는 터널링 패널티(tunneling penalty)이고, λ 는 $\partial f / \partial x$ 의 영(zero)점 x^* 의 차수보다 작거나 같은 값이다. 또한 $f^*(z)$ 는 piecewise-linear 함수인 heaviside 함수로 정의한다. 즉,

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

이다.

(그림 1)은 연속적이고 미분 가능한 6 개의 최소점을 가지는 다극점함수 $f(x)$ 에 동적최적화와 동적터널링 과정을 반복 적용함으로써 전역최적화의 과정을 나



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

타낸 것이다. 그림에서 보는 바와 같이 임의의 값 $x^{(0)}$ 를 초기값으로 하여 함수 $f(x)$ 에 기울기 하강법의 동적최적화 과정을 적용하면 국소최소점 x^* 에 도달하고, x^* 에 교란 ϵ 을 추가한 초기값으로 동적터널링 과정을 적용하면 $x^{(0)}$ 에 수렴하게 된다.

또한 $x^{(0)}$ 를 초기값으로 동적최적화 과정을 적용하면 x^* 에 도달하고, 동적터널링 과정에 의하여 $x^{2(0)}$ 를 구하며, 다시 $x^{2(0)}$ 를 초기값으로 하여 동적최적화 과정에 의해서 전역최소점인 x^* 를 구할수 있다. 이 과정에서 x^4 와 x^5 그리고 x^6 는 터널링 현상에 의하여 제거된다. 따라서 동적최적화 과정으로 공액 기울기법을 사용하고 전역 최소점을 찾기 위한 과정으로 동적터널링 시스템을 이용하면 빠른 수렴속도의 전역 최적화를 위한 방법의 신경망을 이용한 학습에도 그 대로 이용될 수 있을 것이다.

2.3 공액 기울기법의 후향전파 알고리즘과 동적터널링 시스템의 후향전파 알고리즘을 조합한 학습방법

공액 기울기법은 그 동작 특성상 국소 최적해에 빠질 확률이 있지만 수렴속도가 기울기 하강법에 비해 매우 빠른 동작특성을 가지고 있으며, 동적터널링 시스템은 전역최적화를 위한 알고리즘이다. 그러므로 공액 기울기법과 동적터널링 시스템의 장점을 조합할 수 있으면 더욱 우수한 수렴특성을 가지는 최적화가 가능할 것이다. 즉, 전역최적해로 수렴속도를 빠르게 하는 데는 공액 기울기법을 이용하고 전역최적해로의 수렴을 확실히 하는 데는 동적터널링 시스템을 이용하면 고속으로 학습문제의 전역최적해를 구할 수 있다. 학습을 위한 다층 신경망에서 연결가중치 벡터 w 와 학습패턴 p 에 대한 평균자승 오차함수(mean square error function) $E_p(w)$ 의 최소값을 찾는 과정은 공액 기울기법이나 동적터널링 시스템에서 상태벡터 x 에 대한 비용함수 $f(x)$ 의 최소값을 찾는 과정으로 대응시킬 수 있다. 그러므로 다층 신경망의 학습에 동적최적화 과정으로 공액 기울기법의 후향전파 알고리즘을 적용하여 평균자승 오차함수의 값이 최소인 연결가중치를 구하고, 구해진 연결가중치에 교란을 추가한 새로운 연결가중치를 초기값으로 하여 터널링 기능을 가지는 동적터널링의 후향전파 알고리즘을 적용하면 학습시간의 단축 및 전역최소값을 찾

을 수 있는 신경망의 학습이 가능함을 알 수 있다.

먼저, 다층 신경망에서 전체 오차함수 $E(w)$ 및 학습패턴에 대한 평균자승 오차함수 $E_p(w)$ 는 각각

$$E(w) = \sum_p E_p(w) \tag{12}$$

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

로 정의된다. 여기서 d_{pk} 와 o_{pk} 는 각각 p 번째 패턴에 대한 출력층 뉴런 k 의 원하는 출력과 실제 출력이다. 따라서 평균자승 오차함수 $E_p(w)$ 에 공액 기울기법을 적용할 때 p 번째 패턴에 대한 연결가중치 벡터 w 의 변화는

$$w(t+1) = w(t) + \gamma(t)s(t)$$

$$g(t) = \nabla_w E_p(w)$$

로 구해진다. 여기서 $\gamma(t)$ 와 $s(t)$ 는 각각 반복횟수 t 에 따른 최적의 step size와 공액 기울기이며, $g(t)$ 는 $E_p(w)$ 의 기울기이다. 이때 $\gamma(t)$ 및 $s(t)$ 는 앞 절에서의 식 (7)에서 식 (9)까지에 의하여 계산된다.

한편, 평균자승 오차함수 $E_p(w)$ 에 동적터널링 시스템을 적용할 때 p 번째 패턴에 대한 연결가중치 벡터 w 의 변화는

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

$$\Delta_p w = -\frac{\partial E_p(w)}{\partial w} \left(\frac{1}{[(w-w^*)(w-w^*)]^k} \right) + kE_p^*(\hat{E}_p(w)) \tag{13}$$

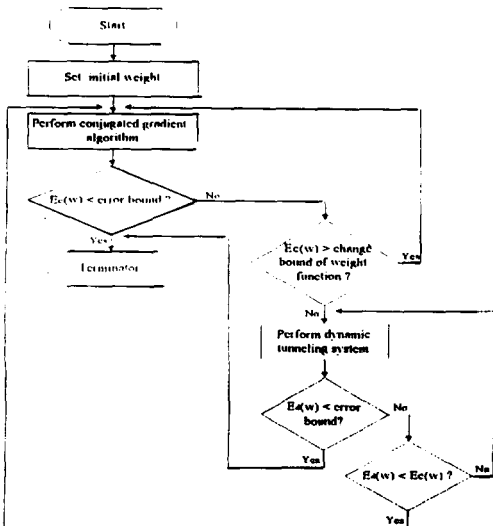
로 구할 수 있다. 여기서 η 와 α 는 각각 학습률(learning rate)과 모멘트(momentum)이고, $E_p(w) = E_p(w) - E_p(w^*)$ 이다. 또한 $E_p^*(z)$ 는 $z > 0$ 이면 z 이고, $z < 0$ 이면 0의 값을 갖는다.

따라서 다층 신경망에서 공액 기울기법과 동적터널링 시스템에 기초한 후향전파 알고리즘을 이용하여 학습성능을 개선하기 위해 제안한 알고리즘은 다음과 같다.

단계 1: 다층 신경망의 전체오차함수 $E(w)$ 및 학습패턴 p 에 대한평균 자승 오차 함수 $E_p(w)$ 를 각각 정의한다.

- 단계 2: 출력층과 은닉층이 연결가중치를 각각 무작위로 발생시킨다.
- 단계 3: 공액 기울기법의 후향전파 알고리즘을 수행하여 계산되는 전체 오차함수의 값이 허용치 이하이면 학습을 종료하고, 그렇지 않고 공액 기울기법의 후향전파 알고리즘을 반복함에 따라 전체 오차함수의 값이 허용치 이상으로 변하지 않으면 후향전파 알고리즘을 종료한다.
- 단계 4: 단계 3에서 계산된 연결가중치에 교란 ϵ 을 추가하여 동적터널링 시스템의 후향전파 알고리즘을 수행한다.
- 단계 5: 동적터널링 시스템의 후향전파 알고리즘 수행으로 계산된 전체 오차함수의 값이 허용치 이하이면 학습을 종료하고, 그렇지 않고 계산된 전체 오차함수의 값이 단계 3의 공액 기울기법의 후향전파 알고리즘에서 계산된 오차함수값 이하이면 단계 3으로 간다.

여기서 학습 알고리즘의 종료는 전체 오차함수의 값이 설정된 허용치 이하일 때로 한다.



(그림 2) 공액 기울기법과 동적터널링 시스템을 조합한 후향전파 알고리즘의 학습 흐름도

(Fig. 2) The flow chart of backpropagation learning algorithm using a hybrid of conjugate gradient and dynamic tunneling system

이와같이 공액 기울기법의 후향전파 알고리즘을 수행하여 국소최적해로 수렴될 경우, 교란을 가한 연결가중치를 새로운 초기값으로 하여 터널링 기능을 가진 동적터널링 시스템의 후향전파 알고리즘을 실행하게 된다.

따라서 동적터널링 과정으로 기존의 공액 기울기법의 후향전파 알고리즘이 학습 파라미터나 신경망의 구조등과 같은 요인에 따라 국소최적해에 수렴되어 버리는 문제를 해결할 수 있으며, 공액 기울기법이 가지는 빠른 수렴속도의 특성도 그대로 살릴 수 있다. 이상에서 제안된 학습 알고리즘의 흐름도는 (그림 2)와 같다.

(그림 2)에서의 $E_c(w)$ 와 $E_d(w)$ 는 각각 공액 기울기법과 동적터널링 시스템의 후향전파 알고리즘을 이용하여 구해진 전체 오차함수이다.

3. 시뮬레이션 결과 및 고찰

제안된 방법의 타당성과 학습성능을 확인하기 위한 응용실험 대상으로 패리티 검사와 패턴 분류를 택하였다. 패리티 검사는 가장 유사한 패턴들이 서로 다른 출력층을 요구하는 특징을 가진 여러 개의 국소 최적해가 존재하는 문제 중의 하나이다. 이 문제는 출력층 뉴런의 수가 1개인 구조의 신경망을 이용하며, XOR 문제는 2비트 패리티 검사의 대표적인 문제에 속한다. 또한 패턴 분류는 학습을 통하여 불완전한 입력 패턴을 분류해 내는 특징을 가진 문제로서 영상이나 음성의 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 있다¹⁾. 따라서 기존의 기울기 하강법에 근거한 후향전파 알고리즘, 기울기 하강법과 동적터널링 시스템을 조합한 후향전파 알고리즘(BPDTS) 그리고 공액 기울기법과 동적터널링 시스템을 조합한 후향전파 알고리즘(CGDTTS)을 이들 문제에 각각 적용하여 그 타당성을 확인하였으며, 무작위로 설정한 초기의 연결 가중치를 이용하는 기존의 후향전파 알고리즘의 결과와 그 성능을 비교 고찰하였다.

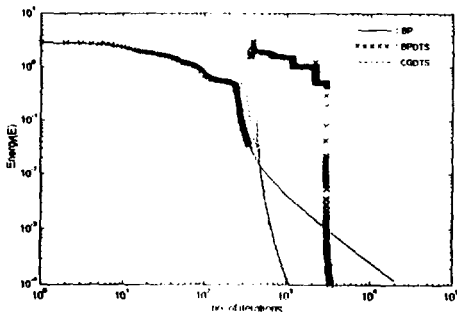
학습에 이용된 다층 신경망의 구조로는 층 사이의

뉴런간에 완전한 연결을 갖는 입력층, 은닉층 및 출력층으로 구성된 삼층 구조를 택하였다. 초기 연결가중치들의 범위는 -0.5 와 $+0.5$ 사이의 값으로 하였으며 무작위 씨앗(random seed)의 변경으로 연결가중치들을 변화시킬 수 있도록 하였다. 알고리즘의 종료는 계산 반복횟수(number of iterations)가 20000번 이상이거나, 전체 오차함수의 값이 주어진 값 PEV(permissible error value) 이하일 때로 하였다. 여기서 계산 반복 횟수는 모든 입력패턴이 연결가중치를 경신하기 위해서 한번씩 다 이용된 때를 1 회로 하였다. 특히 BP와 BPDTS, 그리고 CGDTS에서 학습률 η 와 모멘트 α 는 각각 그 조합으로 실험한 결과 중에서 가장 우수한 조합의 값으로 설정하였다.

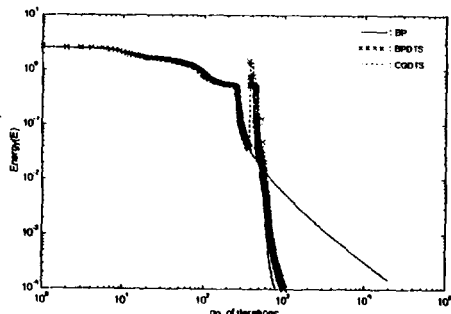
3.1 패리티검사

패리티 검사는 각 학습 패턴 내에 포함된 "1"의 개수에 따라 출력을 결정하는 문제로서, 실험에서는 일부분(partial) 및 우수(even) 패리티로 "1"의 개수가 짝수 개이면 "1", 홀수 개이면 "0"이 출력되도록 학습

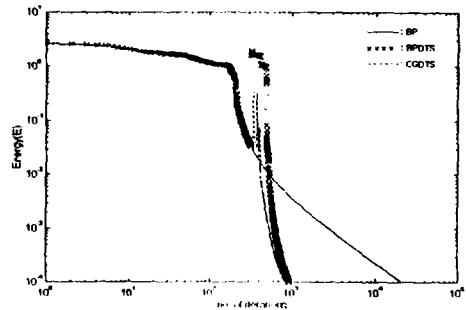
시켰다. 실험은 학습 패턴의 수가 20, 크기가 7인 문제를 대상으로 하였으며, 허용오차함수 PEV는 0.0001로 하였다. 또한 신경망의 구조로 입력층과 은닉층의 뉴런 수는 각각 7개로 하고 출력층의 뉴런 수는 1개로 하였다.



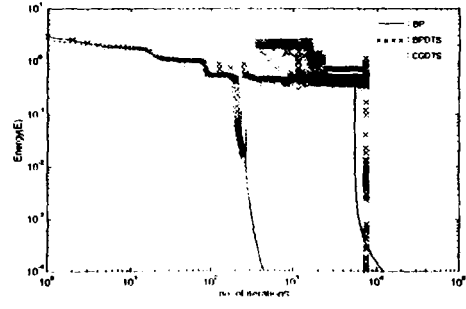
(a) seed = 40, $\eta = 1.0$, $\alpha = 0.0$



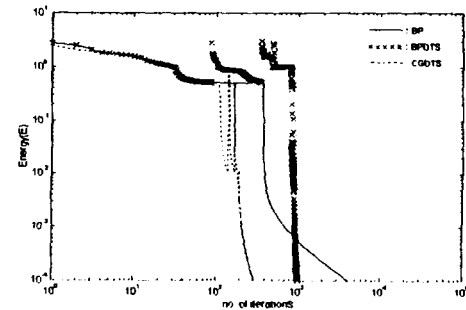
(b) seed = 36, $\eta=0.1$, $\alpha=0.9$



(c) seed=75, $\eta=0.3$, $\alpha=0.7$



(e) seed= 42, $\eta=0.8$, $\alpha=0.8$



(d) seed = 70, $\eta = 0.5$, $\alpha = 0.9$

(그림 3) 학습률 η 와 모멘트 α 에 따른 20 개의 7 비트 패턴에 대한 패리티 검사의 오차 함수 E

(Fig. 3) Error function E of parity check for the 20 of the 7-bit patterns in each learning rate η and momentum α

(그림 3)은 여러가지 무작위 씨앗에 대해서 학습률 η 와 모멘트 α 를 각각 1.0과 0.0, 0.1과 0.9, 0.3과 0.7, 0.5과 0.9 그리고 0.8과 0.8로 하여 BP, BPDTS, 그리고 CGDTS를 각각 적용할 때의 반복횟수에 따른 전체 오차함수 E의 변화를 나타낸 것이다.

(그림 3)에서 나타난 것처럼 기존의 BP보다 BPDTS는 최적해로의 수렴속도가 빠르고 더구나 CGDTS는 앞의 두 경우보다 더욱 더 개선되었음을 알 수 있다. 특히 BPDTS와 CGDTS의 적용시는 최적해로의 수렴 확률이 높음을 알 수 있다. 또 각각의 알고리즘에서 수렴속도는 학습률과 모멘트에 의존함을 알 수 있다. 또한 BP에서는 오차함수의 변화분이 미소할 때의 수렴속도는 현저하게 느려지나 BPDTS나 CGDTS는 오차함수의 변화율이 적은 부분에서 동적터널링 동작을 수행함으로써 최적해로의 수렴속도가 빠름을 알 수 있다.

(표 1)은 20 개의 7 비트 패턴에 대한 패리티 검사에 있어서 BP, BPDTS, 그리고 CGDTS를 각각 적용하여 무작위 씨앗을 다르게 100 번씩 시도한 결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 각각 평균 \bar{x} 와 표준편차 σ 를 나타낸 것이다. (표 1)

에서 N_{bp} , N_{bd} , N_{cd} , N_{cg} 그리고 N_{dts} 는 각각 BP, BPDTS, CGDTS, 공액 기울기법, 그리고 동적터널링 시스템의 반복횟수이며, t_{bp} , t_{bd} , t_{cd} 은 각각 BP, BPDTS, 그리고 CGDTS에서 소요된 CPU 시간이다. 특히 t_{bd} 와 t_{cd} 은 전역최적해를 구하기 위해 기울기 하강법과 동적터널링 시스템 및 공액기울기법과 동적터널링 시스템에서 각각 소요된 CPU시간의 합이다.

(표 1)에서 보는 바와 같이 각각의 알고리즘에서 학습률과 모멘트의 값에 따라 최적해로의 수렴확률 및 그 수렴속도가 매우 의존됨을 알 수 있다. CGDTS와 BPDTS는 BP에 비하여 반복횟수의 경우에 CGDTS는 약 10.7배에서 약 22.2배, BPDTS는 약 8.3배에서 22.1배 정도 개선되었으며, 학습시간의 경우에 CGDTS는 약 5.3배에서 약 10.5배, 그리고 BPDTS는 약 2.6배에서 약 7.5배정도 개선되었다. 또한 CGDTS와 BPDTS는 기존의 BP에 비하여 학습률 η 와 모멘트 α 의 값이 각각 $\eta=0.8$ 과 $\alpha=0.8$ 및 $\eta=0.5$ 과 $\alpha=0.9$ 인 경우에는 비슷한 최적해로의 수렴확률을 보여주나 $\eta=1.0$ 과 $\alpha=0.0$, $\eta=0.1$ 과 $\alpha=0.9$ 그리고 $\eta=0.3$ 과 $\alpha=0.7$ 일 경우에 CGDTS와 BPDTS는 11.1배에서 16.7배 정도 그 수렴확률이 개선되었다. 여기에서 CGDTS 및 BPDTS

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

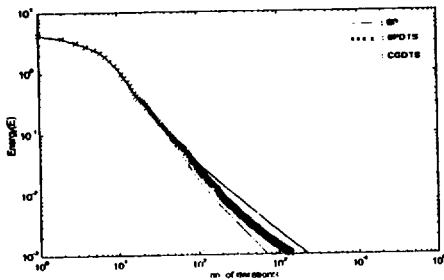
η	α	BP Algorithm				BPDTS				CGDTS			
		N_{bp}	E	t_{bp} (sec)	P_r (%)	$N_{bd}(N_{bp}, N_{dts})$	E	t_{bd} (sec)	P_r (%)	$N_{cd}(N_{cg}, N_{dts})$	E	t_{cd} (sec)	P_r (%)
1.0	\bar{x}	19535.8	0.000100	49.3	6	1510.7(911.1, 600)	0.000100	11.6	100	1295.2(817.8, 477.3)	0.000100	6.9	100
	σ	345.6	0.000000	0.94		962.3(478.1, 502)	0.000001	7.8		361.0(193.1, 211.5)	0.000001	2.1	
0.1	\bar{x}	19288.4	0.000100	48.5	8	872.7(598.5, 274.2)	0.000098	6.47	100	868.9(597.6, 271.3)	0.000098	4.6	100
	σ	526.4	0.000000	1.66		184.0(125.5, 70.6)	0.000002	1.5		175.9(124.8, 61.9)	0.000002	1.0	
0.3	\bar{x}	19301.1	0.000100	48.7	9	1116.2(726.4, 389.8)	0.000099	8.4	100	1060.9(686.4, 374.4)	0.000098	5.7	100
	σ	660.3	0.000000	1.9		387.5(214.3, 183.2)	0.000001	3.1		247.4(146.8, 120.0)	0.000002	1.4	
0.5	\bar{x}	5084.1	0.000100	12.7	99	613.1(350.0, 263.1)	0.000099	4.8	100	445.3(263.8, 181.5)	0.000099	2.4	100
	σ	1466.3	0.000000	3.8		1439(744.4, 695.8)	0.000001	11.9		527.7(278.3, 252.7)	0.000001	3.0	
0.8	\bar{x}	5798.0	0.000100	16.0	99	628.1(374.9, 253.2)	0.000099	4.86	100	539.9(318.9, 221.0)	0.000099	2.8	100
	σ	1609.8	0.000000	4.8		828.4(515.1, 326.1)	0.000000	6.5		519.0(250.9, 273.3)	0.000001	2.9	

N_{bp} , N_{bd} , N_{cd} , N_{dts} , N_{cg} : no. of iterations η : learning rate α : momentum \bar{x} : mean σ : standard deviation t : CPU time P_r : convergence ratio

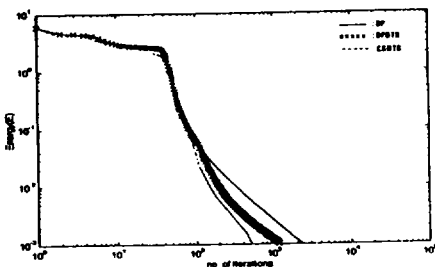
는 5 가지 학습파라미터의 조합 각각에 대하여 100 번씩 시도한 결과 모두 최적해로 수렴되어 국소최적 해로의 수렴문제를 해결할 수 있음을 알 수 있다. 그리고 CGDTS와 BPDTS의 반복횟수에 대한 표준편차는 BP의 표준편차에 비하여 $\eta=1.0$ 과 $\alpha=0.0$ 인 경우를 제외한 나머지 4가지의 학습파라미터의 경우 CGDTS는 약 1.9배에서 3.0배 정도 적은 값이고, BPDTS는 약 1.0배에서 2.85배 정도 적은 값이다. 또한 학습시간에 대한 표준편차는 BPDTS가 BP에 비하여 $\eta=0.1$ 과 $\alpha=0.9$ 인 경우에 약 1.1배 정도 적은 값이며 CGDTS는 BP에 비하여 $\eta=1.0$ 과 $\alpha=0.0$ 의 경우를 제외하고는 약 1.7배에서 약 1.3배 정도 적은 값이다. 이는 BP에 비하여 BPDTS와 CGDTS방법의 성능이 초기연결 가중치에 덜 의존함을 보여준다.

3.2 패턴분류

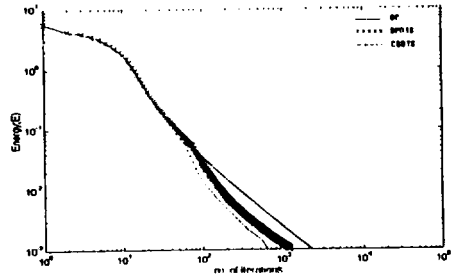
실험은 학습패턴의 크기가 64 비트인 10 개의 패턴에 대하여 실시하였으며, 학습의 종료는 허용오차함수 PEV가 0.001 이하일 때로 하였다. 또한 신경망의 구조로 입력층과 은닉층의 뉴런 수는 각각 64개로 하고 출력층의 뉴런 수는 10개로 하였다.



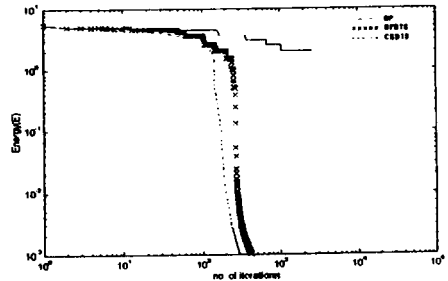
(a) seed=40, $\eta=1.0$, $\alpha=0.0$



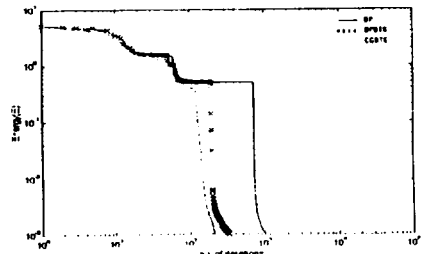
(b) seed=36, $\eta=0.1$, $\alpha=0.9$



(c) seed=75, $\eta=0.3$, $\alpha=0.7$



(d) seed=70, $\eta=0.5$, $\alpha=0.9$



(e) seed=42, $\eta=0.8$, $\alpha=0.8$

(그림 4) 무작위 씨앗의 변화에 따른 10개의 64비트 패턴에 대한 패턴분류의 오차함수 E

(Fig. 4) Error function E of pattern classification to random seed variation at each learning rate η and momentum α

(그림 4)는 여러가지 무작위 씨앗에 대해서 학습률 η 와 모멘트 α 가 각각 1.0과 0.0, 0.1과 0.9, 0.3과 0.7, 0.5와 0.9, 그리고 0.8과 0.8로 하여 BP, BPDTS, 그리고 CGDTS를 각각 적용할 때의 반복횟수에 따른 전체 오차함수 E의 변화를 나타낸 것이다. 그림에서는 여러가지 무작위 씨앗에 따라 각각은 수렴하였으나

그림 d의 경우에 BP는 수렴하지 못하였고 제시한 두 알고리즘에서는 완전히 수렴함을 알 수 있다. 여기서의 수렴확률, 수렴속도 등은 (그림 3)에서의 결과와 유사함을 알 수 있다. <표 2>는 10개의 64 비트 패턴에 대한 패턴분류에 있어서 BP, BPDTS, 그리고 CGDTS를 각각 적용하여 무작위 씨앗을 다르게 100 번씩 시도한 결과, 최적해로 수렴된 경우들에 대한 반복횟수 및 CPU 시간의 각각 평균 \bar{x} 와 표준편차 σ 를 나타낸 것이다.

<표 2>에서도 BP, BPDTS, 그리고 CGDTS는 학습률과 모멘트의 값에 따라 최적해로의 수렴확률 및 그 수렴속도가 매우 의존됨을 알 수 있다. <표 2>에서 보는 바와 같이 CGDTS와 BPDTS는 BP에 비하여 반복횟수의 경우에 CGDTS는 약 1.9배에서 약 15.3배, BPDTS는 약 1.6배에서 11.1배 정도 개선되었으며, 학습시간의 경우에 CGDTS는 약 1.2배에서 약 11.6 배, 그리고 BPDTS는 약 1.1배에서 약 8.3배 정도 개선되었다. 또한 CGDTS와 BPDTS는 기존의 BP에 비하여 학습률 η 와 모멘트 α 의 값이 $\eta=0.5$ 과 $\alpha=0.9$ 인 경우를 제외하고는 비슷한 최적해로의 수렴확률을 보여주나 $\eta=0.5$ 과 $\alpha=0.9$ 일 경우에 약 4.2배 정도 그

수렴확률이 개선되었다. 그리고 CGDTS와 BPDTS의 반복횟수에 대한 표준편차는 BP의 표준편차에 비하여 CGDTS는 약 1.7배에서 42.9배 정도 적은 값이고, BPDTS는 약 1.4배에서 33.4배 정도 적은 값이다. 또한 학습시간에 대한 표준편차는 CGDTS는 BP에 비하여 약 1.1배에서 약 29.8배 정도 적은 값이며 BPDTS는 BP에 비하여 $\eta=1.0$ 과 $\alpha=0.0$, $\eta=0.3$ 과 $\alpha=0.7$ 인 경우를 제외하고 약 1.2배에서 약 21.2배 정도 적은 값이다. 여기서도 BP에 비하여 BPDTS와 CGDTS 방법의 성능이 초기 연결가중치에 덜 의존함을 보여준다.

4. 결 론

본 논문에서는 공액 기울기법과 동적터널링 시스템을 조합사용함으로써 다층 신경망의 학습성능을 개선할 수 있는 새로운 방법을 제안하였다. 제안된 방법에서는 공액 기울기법의 후향전파 알고리즘을 사용하고, 국소최적해를 만났을 때 이를 벗어난 다른 연결가중치의 설정을 위하여 동적터널링 시스템의 후향전파 알고리즘을 적용함으로써 빠른 수렴속도로 전역최적화가 가능하도록 하였다.

<표 2> 10개의 64 비트 패턴에 대한 패리티검사의 100번 시도에 따른 실험결과
 <Table 2> Results for the 10 of the 64-bit patterns parity check of 100 trials

η	α	BP				BPDTS				CGDTS			
		N_{bp}	E	t_{bp} (sec)	P_r (%)	$N_{bd}(N_{bpb}, N_{dts})$	E	t_{bp} (sec)	P_r (%)	$N_{cd}(N_{cgb}, N_{dts})$	E	t_{cd} (sec)	P_r (%)
1.0	\bar{x}	2198.0	0.000990	276.3	100	1341.4(317, 1024.1)	0.000998	257.2	100	1149.2(271.7, 877.5)	0.001000	220.4	100
	σ	62.0	0.000001	7.7		42.5(8.9, 34.4)	0.000002	8.4		36.4(7.6, 29.5)	0.000003	7.2	
0.1	\bar{x}	2312.4	0.001000	288.8	100	1191.3(315.8, 875.5)	0.000997	224.4	100	1131.7(300.1, 831.7)	0.000997	212.8	100
	σ	96.9	0.000000	11.3		56.6(37.4, 31.9)	0.000002	9.2		53.8(35.5, 18.3)	0.000002	8.7	
0.3	\bar{x}	2223.1	0.001000	280.0	100	1157.1(261.1, 896.0)	0.000997	222.3	100	925.6(208.8, 716.8)	0.000998	177.8	100
	σ	60.9	0.000000	7.2		37.1(7.8, 30.2)	0.000002	7.2		31.5(10.9, 20.4)	0.000001	6.8	
0.5	\bar{x}	4738.1	0.000997	570.8	24	424.4(234.2, 190.1)	0.000993	68.7	100	308.1(170.0, 138.1)	0.000995	48.9	100
	σ	3169.0	0.000003	378.5		94.8(34.2, 68.3)	0.000005	17.8		73.8(24.8, 49.1)	0.000001	12.7	
0.8	\bar{x}	1250.0	0.000997	152.3	95	337.7(177.0, 160.6)	0.000993	55.4	100	303.9(159.3, 144.6)	0.000998	47.2	100
	σ	1021.5	0.000002	123.8		40.0(26.7, 20.8)	0.000004	6.6		36.5(24.0, 12.5)	0.000001	5.3	

N_{bp} , N_{bd} , N_{cd} , N_{dts} , N_{cgb} : no. of iterations η : learning rate α : momentum \bar{x} : mean σ : standard deviation t: CPU time P_r : convergence ratio

제안된 방법을 20 개의 7비트 패턴에 대한 패리티 검사와 10 개의 64 비트 패턴에 대한 패턴분류에 각각 적용하여 그 타당성과 그 성능을 확인하였다. 20 개의 7 비트 패턴에 대한 패리티검사에 있어서 기존의 BP에 비해 최적해로의 수렴확률은 CGDTS와 BPDTS는 최대 약 16.7배 정도 개선되었고, 수렴속도에서는 CGDTS는 최대 약 22.2배, BPDTS는 최대 약 22.1배 정도 개선되었다. 또한 10 개의 64 비트 패턴에 대한 패턴분류에서도 CGDTS와 BPDTS는 BP에 비해 최적해로의 수렴확률은 최대 약 4.2배 정도 개선되었고, 수렴속도에서는 CGDTS는 최대 약 15.3배, BPDTS는 최대 약 11.1배 정도 개선된 우수한 학습성능을 가짐을 확인할 수 있다.

따라서 제안된 학습방법의 다층 신경망을 지도처리 등에 적용하면 신경망이 가지는 속성을 더욱 더 잘 활용할 수 있을 것이다. 향후 제안된 방법을 좀 더 큰 학습 문제 및 조합형 최적화 문제에 확장 적용할 수 있는 연구수단을 마련하는 일이 남아 있다.

참 고 문 헌

- [1] D.C.Plaut, S.J.Nowlan, and G.E.Hinton, "Experiments on learning by backpropagation," Carnegie-Mellon Univ., CMU-CS-86-126, 1986.
- [2] 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.
- [3] D. E. Rumelhart and J. L. McClelland, 'Parallel Distributed Processing', Mit Press, Cambridge, MA., pp. 282-362, 1986.
- [4] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," Journal of Neural Networks, vol. 1, pp. 295-307, Nov. 1988.
- [5] C. H. Chen and R. S. Nutter, "Improve 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, 2063-2068, Nov. 1991.
- [6] Y. Hirose, K. Yamashita, and S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," Neural Networks, vol. 1, pp. 61-66, 1991.
- [7] 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.
- [8] 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.
- [9] J.J.Hopfield and D.Tank, "Neural computation of decision in optimization problems," Biol. cybern., vol. 52, pp. 141-152,1985.
- [10] J.Ramanujam and P.Sadayappan, "Optimization by neural networks," IEEE Conference on Neural Networks, San Diego, vol. 2, pp. 325-332,July 1988.
- [11] 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.
- [12] C. Charalmbous, "Conjugate gradient algorithm for efficient training of artificial neural networks," IEE. Proceedings-G, vol. 139, No. 3, June 1992.
- [13] C. H. CHEN and Hongtao Lai, "A comparison study of the Gradient Descent and the Conjugate Gradient Backpropagation Neural Networks," WCDN '93, vol. 3, Oregon. July 1993.
- [14] A.Zaghw and W.M.Dong, "An automated approach for selecting the learning rate and momentum in backpropagation networks," ICNN '94, vol. 1, June/July 1994.



김 원 욱

1979년 경북대학교 전자공학과 졸업(학사)
1981년 경북대학교 대학원 전자공학과(공학석사)
1994년 경상대학교 대학원 전기·전자공학과 박사과정 수료

1984년~현재 연암공업전문대학 전자과 교수
관심분야: 신경망, 인공지능, 병렬분산처리

조 용 현

1979년 경북대학교 전자공학과 졸업(학사)
1981년 경북대학교 대학원 전자공학과(공학석사)
1993년 경북대학교 대학원 전자공학과(공학박사)
1983년~1987년 한국전자통신연구소
1987년~1997년 영남전문대학 전자과 부교수
1997년~현재 효성카톨릭 대학교 전자·정보공학부 조교수

관심분야: 신경망, 인공지능, 병렬분산처리, 통화코덱 위치 등



김 영 일

1979년 경북대학교 전자공학과 졸업(학사)
1981년 연세대학교 대리원 전자공학과(공학석사)
1985년 연세대학교 대학원 전자공학과(공학박사)
1987년~현재 경상대학교 전자공학과 교수

관심분야: 음성언어처리, 문자인식, 영상처리



강 인 구

1955년 해군사관학교(학사)
1961년 미국 해군대학원 전기공학과(공학석사)
1967년 미국 New Mexico University 전자공학과(Ph.D.)
1993년~1995년 (주)금성사 부사장

1995년~현재 연암공업전문대학 학장
관심분야: 정보통신, Intranet, 공업정책