

적응 오류 제약 Backpropagation 알고리즘

최수용*, 고균병**, 홍대식**

Adaptive Error Constrained Backpropagation Algorithm

Soo-Yong Choi*, Kyun-Byoung Ko**, Dae-Sik Hong** *Regular Members*

요약

Multilayer perceptrons (MLPs)를 위한 일반적인 BP 알고리즘의 학습 속도를 개선하기 위하여 제약을 갖는 최적화 기술을 제안하고 이를 backpropagation (BP) 알고리즘에 적용한다. 먼저 잡음 제약을 갖는 LMS (noise constrained least mean square : NCLMS) 알고리즘과 영잡음 제약 LMS (ZNCLMS) 알고리즘을 BP 알고리즘에 적용한다. 이러한 알고리즘들은 다음과 같은 가정을 반드시 필요로 하여 알고리즘의 이용에 많은 제약을 갖는다. NCLMS 알고리즘을 이용한 NCBP 알고리즘은 정확한 잡음 전력을 알고 있다고 가정한다. 또한 ZNCLMS 알고리즘을 이용한 ZNCBP 알고리즘은 잡음의 전력을 0으로 가정, 즉 잡음을 무시하고 학습을 진행한다. 본 논문에서는 확장된(augmented) Lagrangian multiplier를 이용하여, 비용함수(cost function)를 변형한다. 이를 통하여 잡음에 대한 가정을 제거하고 ZNCBP와 NCBP 알고리즘을 확장, 일반화하여 적응 오류 제약 BP(adaptive error constrained BP : AECBP) 알고리즘을 유도, 제안한다. 제안한 알고리즘들의 수렴 속도는 일반적인 BP 알고리즘보다 약 30 배정도 빠른 학습 속도를 나타내었으며, 일반적인 선형 필터와 거의 같은 수렴속도를 나타내었다.

ABSTRACT

In order to accelerate the convergence speed of the conventional BP algorithm, constrained optimization techniques are applied to the BP algorithm. First, the noise-constrained least mean square algorithm and the zero noise-constrained LMS algorithm are applied (designated the NCBP and ZNCBP algorithms, respectively). These methods involve an important assumption: the filter or the receiver in the NCBP algorithm must know the noise variance. By means of extension and generalization of these algorithms, the authors derive an adaptive error-constrained BP algorithm, in which the error variance is estimated. This is achieved by modifying the error function of the conventional BP algorithm using Lagrangian multipliers. The convergence speeds of the proposed algorithms are 20 to 30 times faster than those of the conventional BP algorithm, and are faster than or almost the same as that achieved with a conventional linear adaptive filter using an LMS algorithm.

Key Words: BP, LMS, NCBP, ZNCBP, Lagrangian multipliers

I. 서론

신경망(또는 신경회로망)은 통신과 신호처리 연구 분야에서 많은 주목을 받고 있으며, 또한 주목할 만한 연구 성과들이 계속 발표되고 있다^{1, 2}. 하지만

구조적인 비선형 특성으로 인하여 신경망은 아직까지 해결해야 할 문제들이 존재한다. 다층 구조의 퍼셉트론 (multilayer perceptron : MLP)의 경우, 느린 학습 속도와 많은 지역 극소 문제가 대표적이라 할 수 있다. 본 논문에서는 신경망 학습 속도의 개선에 대한 연구를 수행하도록 한다.

* Institute for Neural Computation, University of California, San Diego (sychoi@ucsd.edu), ** Center for Information Technology of Yonsei University (CITY), 연세대학교 전기전자공학과 (pumi@itl.yonsei.ac.kr, daesikh@itl.yonsei.ac.kr)
 논문번호 : 020509-1127, 접수일자 : 2002년 11월 27일

**이 논문은 한국과학재단의 목적기초연구 사업(R01-2002-000-00153-0(2002))의 지원으로 수행되었습니다.

신경망의 학습 속도를 개선하기 위하여 많은 기술들이 개발되고 있다. 한 가지 방법은 현재 gradient의 값에 따라 학습률 (learning rate)을 동적으로 변화시키는 것이다^[3]. 또 다른 접근 방법은 line search를 이용한 conjugate gradient (CG) direction 기법을 들 수 있다^[4, 5]. 이외에도 비용함수의 최적화 방법 등 다양한 기술들이 제안되고 있다^[6, 7]. 하지만 line search, Newton, 그리고 CG 기법 등은 각 학습 단계마다 역행렬 또는 이차 미분 등의 복잡한 연산 과정이 필요하다.

본 논문에서는 일반적인 backpropagation (BP) 알고리즘의 수렴 속도를 개선하고자 제약을 갖는 최적화 (constrained optimization) 기술을 제안한다. 먼저 참고문헌 [8]에서 제안한 잡음 제약 LMS (noise constrained least mean square : NCLMS) 알고리즘과 ZNCLMS (zero noise constrained LMS) 알고리즘의 최적화 기법을 BP 알고리즘에 적용, NCBP 알고리즘과 ZNCBP 알고리즘을 제안한다. 그러나 NCBP 알고리즘과 ZNCBP 알고리즘은 NCLMS 알고리즘과 ZNCBP 알고리즘과 마찬가지로 잡음에 대한 가정이 필요하다. 본 논문에서는 확장된(augmented) Lagrangian multiplier를 이용하여 이러한 가정을 제거하는 기법을 제안하고, 이를 이용하여 적응 오류 제약을 갖는 BP(adaptive error constrained BP : AECBP) 알고리즘을 제안한다.

II. 잡음제약 BP (NCBP) 알고리즘

은닉층이 하나이며, K 개의 출력노드, J 개 은닉 노드 그리고 I 개의 입력노드를 갖는 일반적인 MLPs를 위한 BP 학습 알고리즘에 대한 비용함수 (cost function)를 고려하자.

$$E_p^{BP} = \frac{1}{2} \sum_{k=1}^K \delta_{pk}^2, \quad \text{for } 1 \leq p \leq P \quad (1)$$

여기서 $\delta_{pk}^2 = (d_{pk} - o_{pk})^2$, d_{pk} 와 o_{pk} 는 k 번째 출력 노드에 대한 p 번째 목적(target) 패턴 값과 실제 출력 값을 나타낸다. 식 (1)의 비용함수를 [8]에서와 같이 변수 λ 와 γ 를 이용한 Lagrangian 방법을 사용하여 다음과 같이 변형할 수 있다.

$$E_p^{NCBP} = \frac{1}{2} \sum_{k=1}^K \delta_{pk}^2 + \gamma \lambda (\delta_{pk}^2 - \sigma_n^2) - \gamma \lambda^2 \quad (2)$$

여기서 부가 백색 가우시안 잡음 (additive white Gaussian noise : AWGN)의 평균인 η_k 는 0으로 분산은 σ_n^2 으로 가정한다. 따라서 BP 알고리즘 유도 방법과 같이 연쇄 미분식 (calculus chain rule)을 이용하여 식 (2)로부터 각각의 학습 식을 얻을 수 있다. 은닉층의 j 번째 노드와 출력층의 k 번째 노드를 연결하는 연결 가중치, w_{kj}^o 에 대한 학습식은

$$w_{kj}^o(t+1) = w_{kj}^o(t) + \alpha(t) \delta_{pk}^o i_{pj}, \quad \text{for } 1 \leq j \text{ and } 1 \leq k \leq K \quad (3)$$

$$\alpha(t+1) = \alpha(t) + \mu_\alpha [1 + \gamma \lambda(t)] \quad (4)$$

이고, $\delta_{pk}^o = \delta_{pk} f_k'(net_{pk}^o)$, f_k^o 는 k 번째 출력 노드 값을, net_{pk}^o 은 k 번째 출력 노드의 입력 값을 나타낸다. 또한 $i_{pj} = f_j^h(net_{pj}^h)$ 이며, f_j^h 는 j 번째 은닉 노드 출력 값을, net_{pj}^h 은 j 번째 은닉 노드의 입력 값을 나타낸다. μ_α 는 학습률을 나타낸다. 입력층 또는 하위 은닉층의 i 번째 노드와 j 번째 상위 은닉층 노드를 연결하는 연결 가중치, w_{ij}^h 에 대한 학습식은

$$w_{ij}^h(t+1) = w_{ij}^h(t) + \alpha(t) \delta_{pj}^h x_{pi}, \quad \text{for } 1 \leq i \leq I \text{ and } 1 \leq j \leq J \quad (5)$$

이고, $\delta_{pj}^h = \delta_{pk} f_j^h'(net_{pj}^h)$, $f_j^h = \frac{\partial f_j^h}{\partial net_{pj}^h}$ 그리고

$\alpha(t)$ 는 식 (4)와 같으며, x_{pi} 는 i 번째 입력 노드의 p 번째 입력 패턴이다. 마지막으로 $\lambda(t)$ 에 대한 학습식은 다음과 같다.

$$\lambda(t+1) = \lambda(t) + \mu_\lambda \gamma \left[\frac{1}{2} (\delta_{pk}^2 - \sigma_n^2) - \lambda(t) \right] \quad (6)$$

여기서 μ_λ 는 λ 에 대한 학습률이다. 따라서 식 (3), (4), (5) 그리고 (6)은 NCBP 알고리즘을 구성한다.

ZNCBP 알고리즘은 NCBP 알고리즘의 특별한 경우로서 NCBP 알고리즘을 유도한 비용함수 식 (2)에서 잡음에 대한 분산인 σ_n^2 을 0으로 대체 함으로써 얻을 수 있다. 즉 ZNCBP 알고리즘은 NCBP 알고리즘의 식 (6)에서 σ_n^2 을 0으로 대체 함으로써 얻을 수 있다. 따라서 NCBP와 ZNCBP 알고리즘은 기존의 BP 알고리즘에 대하여 연결 가중치마다

아닌 매 학습마다 식 (4)와 (6)만이 추가적으로 필요할 뿐이다. 그리고 식 (4)와 (6)은 계산량이 많은 벡터 또는 행렬의 형태가 아닌 스칼라 양이다. 따라서 BP 알고리즘에 비해 제한한 NCBP와 ZNCBP 알고리즘의 계산량 증가는 아주 미미하다고 할 수 있다.

III. 적용오류제약 BP(AECBP) 알고리즘

앞에서 제안한 NCBP와 ZNCBP 알고리즘은 수신기 또는 필터에서 반드시 잡음의 분산을 알아야 하거나 잡음을 무시한다는 가정이 있다. 본 절에서는 확장된 Lagrangian multiplier를 이용하여 이러한 가정이 필요 없는 AECBP 알고리즘을 제안한다.

잡음에 대한 가정을 제거하기 위하여, 식 (2)에서 잡음의 분산인 σ_n^2 을 새로운 변수 ζ 로 치환하고, 확장된 Lagrangian 기법을 이용하자. 따라서 식 (2)는 다음 식과 같이 또 다른 제약 최소화 문제 (constrained minimization problem)로 변환된다.

$$E_b^{AECBP} = \frac{1}{2} \sum_{k=1}^K \delta_{pk}^2 + \gamma \lambda (\delta_{pk}^2 - \zeta) - \gamma \lambda^2 + \rho \zeta^2, \quad \text{for } 1 \leq p \leq P \quad (7)$$

여기서 ρ 는 양의 상수이다. 식 (7)에서 제안한 제약 조건은 $\zeta^2 = 0$ 이고, 식 (7)은 연결 가중치와 ζ 에 대하여 최소화되어야 하고, λ 에 대하여 최대화되어야 한다. 확장된(augmented) Lagrangian에 대한 자세한 설명은 참고문헌 [5]에 있다.

NCBP 알고리즘의 유도와 같은 방법으로 AECBP 알고리즘에 대한 학습 식을 유도할 수 있다. 출력층과 은닉층의 연결 가중치에 대한 학습 식과 $\alpha(t)$ 에 대한 학습식은 각각 식 (3), (4) 그리고 (5)와 같다. 또한 $\lambda(t)$ 와 $\zeta(t)$ 에 대한 학습식은 각각 다음과 같다.

$$\lambda(t+1) = \lambda(t) + \mu_\lambda \gamma \left[\frac{1}{2} (\delta_{pk}^2 - \zeta(t)) - \lambda(t) \right] \quad (8)$$

$$\zeta(t+1) = \zeta(t) + \mu_\zeta [\gamma \lambda(t) - 2\rho \zeta(t)] \quad (9)$$

여기서 μ_λ 와 μ_ζ 는 각각 λ 와 ζ 에 대한 학습률이다. 따라서 AECBP 알고리즘은 식 (3), (4), (5), (7), 그리고 (8)로 구성된다. $4\rho \ll \gamma$ 라는 가정 하에, 식 (7), (8), 그리고 (3)을 이용하여 각각 다음을 얻

을 수 있다.

$$E[\lambda(t)] \approx \frac{2\rho}{\lambda} E[\delta_{pk}^2(t)] \quad (10)$$

$$E[\zeta(t)] = \frac{\lambda}{2\rho} E[\lambda(t)] = E[\delta_{pk}^2(t)] \quad (11)$$

$$E[\alpha(t)] = \mu_\alpha (1 + 2\rho E[\delta_{pk}^2(t)]) \quad (12)$$

따라서, 제한한 알고리즘들은 최종 mean square error(MSE) 값에 의하여 조절된다.

IV. 실험 결과

그림 1과 같은 정보 전송 시스템에서의 등화 문제와 은닉층이 하나인 MLPs를 고려하자. 등화 문제는 일반적으로 inverse modeling 문제로 간주된다. 신호가 전송되는 채널을 통하여 원래의 신호는 왜곡된다. 이러한 왜곡 신호를 복원하는 문제가 등화 문제이다. 정보 신호는 이진 변조(binary phase shift keying : BPSK) 신호를, 채널은 다음을 사용하였다.

$$H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2} \quad (13)$$

각 학습 알고리즘에 대한 인자는 다음과 같다. NCBP, ZNCBP 그리고 AECBP 알고리즘에 대하여 $\mu_\alpha = 0.01$, $\mu_\lambda = 0.01$, $\gamma = 100$ 을, 또한 AECBP 알고리즘에 대하여 부가적으로 $\mu_\zeta = 0.00001$, $\rho = 0.0001$ 을 사용하였다. 일반적인 BP 알고리즘의 경우 0.01의 학습률과 0.1의 모멘텀(momentum)을 사용하였다. 사용된 MLPs의 구조는 입력 노드 수가 4개, 은닉노드 수가 16개 그리고 하나의 출력 노드를 사용하였다. 본 논문에서 도시한 MSE와 α 에 대한 그림은 1000번의 반복적인 실험 결과의 평균이다. 각 실험에서 신경망의 학습을 위해 10000개의 샘플을 이용하였다. 또한 연결 가중치는 -1.0에서 1.0 사이의 값을 임의적으로 선택하여 매번 다

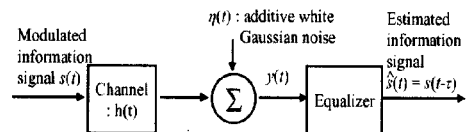


그림 1 단순화된 정보 전송 시스템

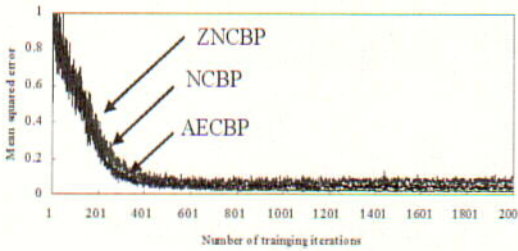


그림 3 제안한 NCBP, ZNCBP, AECBP 알고리즘들에 대한 MSE 수렴 곡선

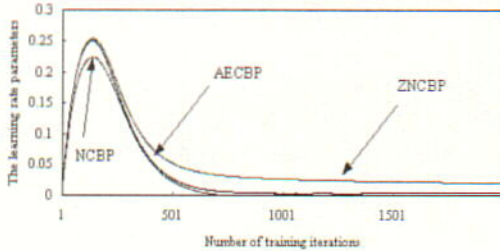


그림 2 제안한 NCBP, ZNCBP, AECBP 알고리즘들에서의 학습률 변화 곡선

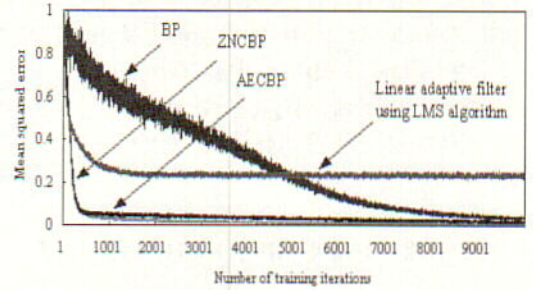


그림 4 ZNCBP, AECBP 알고리즘과 BP 알고리즘, 선형 필터와 MSE 비교

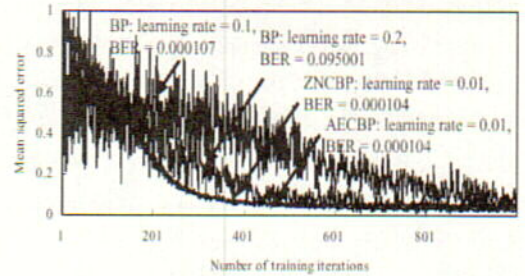


그림 5 ZNCBP, AECBP, 큰 학습률의 BP 알고리즘들과 MSE 및 BER 비교

큰 값으로 초기화하였고, 제안한 알고리즘들에서 $\alpha(t)$, $\lambda(t)$, 그리고 $\zeta(t)$ 는 각각 0으로 초기화하였다.

그림 2는 신호 대 잡음 비 (signal-to-noise ratio : SNR)가 20dB인 경우, 제안한 알고리즘들의 MSE 수렴 곡선을 나타내었다. 그림 2로부터 제안한 알고리즘들의 MSE 수렴 속도가 거의 비슷함을 알 수 있다.

우리는 식 (4)로부터 제안한 알고리즘들이 일종의 적응 학습률을 갖는 BP (variable step-size BP : VSBP) 알고리즘임을 알 수 있다. 이러한 VSBP 알고리즘은 일반적인 BP 알고리즘과 거의 같은 복잡도를 가지며 동적인 학습률 변화를 통하여 빠른 수렴 속도를 가진다. 그림 3은 20dB인 경우, 제안한 알고리즘들의 동적인 학습률 변화 곡선을 나타낸다. 그림으로부터 각 학습 단계의 진행에 따라 제안한 알고리즘들의 학습률들이 동적으로 활발하게 변화함을 알 수 있다.

이번에는 제안한 알고리즘과 일반적인 BP 알고리즘 그리고 LMS 알고리즘을 이용하는 4개의 탭을 갖는 선형 등화기와 MSE 수렴 곡선을 비교하였다. 선형 등화기의 경우 0.01을 학습률로 사용하였다.

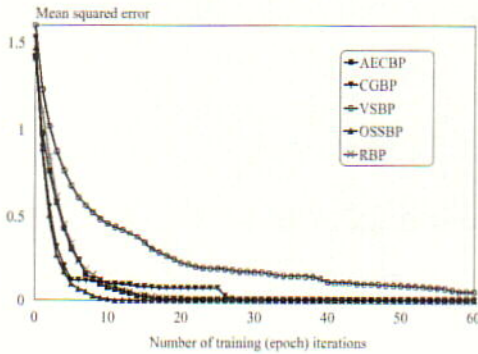
그림 4의 경우 SNR이 20dB인 경우의 MSE를 나타내었다. 20dB일 때 0.1이하의 MSE를 얻기 위하여 필요한 학습 횟수를 보면, 제안한 ZNCBP와 AECBP의 경우 300회 이내의 학습으로 수렴한 반면, BP의 경우 6000번 이상의 학습이 필요하였다. 또한 제안한 알고리즘들은 단지 4개의 탭을 갖는 선형 필터와 거의 유사한 수렴 속도를 가지며, 훨씬 우수한 MSE 수렴 특성을 가짐을 알 수 있다.

지금까지는 일반적인 BP 알고리즘과 제안한 알고리즘들과 같은 학습률로서 실험을 하였다. 이번에는 제안한 알고리즘보다 큰 학습률을 갖는 BP 알고리즘과 비교하여 보았다. 그림 5에 BP 알고리즘의 학습률을 0.1과 0.2로 하여 학습을 수행한 후에 등화기의 비트 오류 확률(bit error probability or bit error rate : BER)값과 MSE 값을 나타내었다. 학습률이 0.2이상으로 증가하게 되면 BP의 MSE 수렴 속도는 제안한 알고리즘들과 거의 비슷하나 BER 값은 원하는 결과를 얻을 수 없었다. 즉 0.2 이상의 값을 BP 알고리즘의 학습률로 이용할 경우, 신경망이 제대로 학습되지 않았다. 따라서 일반적인 BP 알고리즘은 아무리 학습률을 크게 하더라도 BER과 MSE 측면에서 제안한 알고리즘과 같은 성능을 얻

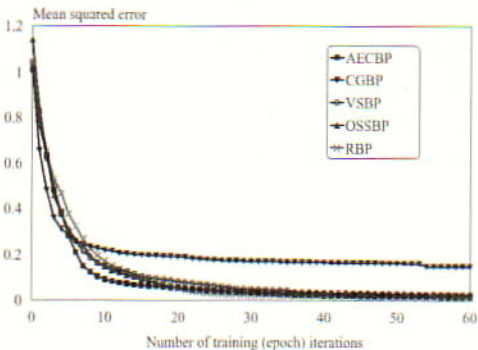
을 수 없었다.

이번에는 제안한 알고리즘과 MLPs 학습을 위한 기존의 알고리즘들 중 빠른 수렴 특성을 갖는 기법과의 비교를 수행한다. 비교하고자 하는 알고리즘들은 다음과 같다. VSBP^[10], RBP (resilient BP)^[11], CGBP (conjugate BP)^[12], 그리고 OSSBP (one-step secant BP)^[7] 알고리즘이다. 이러한 알고리즘들은 학습 속도는 빠르나 역행렬의 계산이나 이차 미분값등의 계산과 같은 복잡한 연산 과정이 필요하다. 제안한 AECBP 알고리즘은 VSBP를 제외하고 다른 알고리즘에 비해 훨씬 간단한 구조를 갖는다. XOR 문제와 sine 함수 근사화 문제를 이용하여 각 학습 방법의 학습 속도를 비교한다. VSBP를 위한 초기 학습률과 모멘텀(momentum)은 각각 0.1과 0.5를 사용하였다. AECBP 알고리즘에서는 $\mu_a = 0.1$, $\mu_\lambda = 0.01$, $\gamma = 100$, $\mu_\xi = 0.0001$ 그리고 $\rho = 0.0001$ 을 사용하였다.

먼저 XOR 문제를 고려하자. 사용된 MLPs의 구



(a)



(b)

그림 6. 제안한 알고리즘과 MLPs를 위한 빠른 알고리즘과의 MSE 수렴 비교 (a) XOR 문제 (b) 함수 근사화 문제

조는 입력 노드 수 2개, 은닉 노드 수 6개 그리고 1개의 출력 노드를 사용하였다. 그림 6 (a)에서 제안한 AECBP 알고리즘이 거의 비슷한 복잡도를 갖는 VSBP보다는 훨씬 빠르며, 매우 복잡한 계산을 요구하는 다른 알고리즘과는 거의 같은 수렴 속도를 나타냄을 알 수 있다.

그림 6 (b)는 $d_k = \sin(2\pi k) + 0.1\eta_k$ 의 잡음이 섞인 sine 함수 근사화 문제를 입력 노드 수가 1개, 은닉 노드 수가 20개 그리고 출력 노드 수가 1개인 신경망을 이용하여 얻은 MSE 곡선이다. k 값의 범위는 $-1 < k < 1$ 이다. 여기서 $k+1 = k+0.05$ 이며 AWGN의 분산은 1이다. XOR 문제에서와 마찬가지로 제안한 AECBP 알고리즘이 가장 간단한 구조로 가장 빠른 수렴 속도를 가짐을 알 수 있다.

V. 결론

본 논문에서는 일반적인 BP 알고리즘의 학습 속도를 개선하기 위하여 제약을 갖는 최적화 기법을 제안하였다. 먼저 잡음에 대한 가정을 이용하는 NCBP 그리고 잡음에 대한 가정을 무시하는 ZNCBP 알고리즘을 제안하였다. 또한 확장된 개념의 Lagrangian multiplier를 이용하여 NCBP 알고리즘을 확장하고, 일반화하여, AECBP 알고리즘을 유도, 제안하였다.

제안한 NCBP, ZNCBP 그리고 AECBP 알고리즘들은 일반적인 BP 알고리즘보다 약 20-30배 정도의 빠른 학습 속도를 나타내었다. 또한 제안한 알고리즘들을 입력 노드 수가 4개, 은닉 노드 수가 16개 그리고 출력 노드 수가 1개인 MLPs에 적용하였을 경우, 은닉노드를 갖지 않고 입력 탭이 4개뿐인 선형 필터와 거의 같은 학습 속도를 나타내었다.

마지막으로 제안한 알고리즘들을 MLPs를 위한 빠른 학습 방법들과 비교하였다. 제안한 AECBP 알고리즘은 훨씬 간단한 구조로 거의 같은 학습 속도를 가짐을 확인 할 수 있었다. 따라서 약간의 계산량 증가를 가지고 놀라운만한 수렴 속도의 개선 효과를 제안한 알고리즘을 이용하여 얻을 수 있음을 확인하였다.

또한 향후 연구 과제로서 제안한 기법의 포괄적인(global) 최적화 기법에 적용을 들 수 있다. 즉 제안한 기법을 포괄적인 최적화 기법에 적용함으로써 수렴 속도의 개선뿐만 아니라 신경망의 최종 성능 또한 개선할 수 있을 것으로 기대된다.

