

---

# 다층퍼셉트론의 강하 학습을 위한 최적 학습률

## Optimal Learning Rates in Gradient Descent Training of Multilayer Perceptrons

---

오상훈

목원대학교 정보통신전공공학부

Sang-Hoon Oh(shoh@mokwon.ac.kr)

---

### 요약

이 논문은 다층퍼셉트론의 학습을 빠르게 하기 위한 최적 학습률을 제안한다. 이 학습률은 한 뉴런에 연결된 가중치들에 대한 학습률과, 중간층에 가상의 목표값을 설정하기 위한 학습률로 나타난다. 그 결과, 중간층 가중치의 최적 학습률은 가상의 중간층 목표값 할당 성분과 중간층 오차함수를 최소화 시키고자 하는 성분의 곱으로 나타난다. 제안한 방법은 고립단어인식과 필기체 숫자 인식 문제의 시뮬레이션으로 효용성을 확인하였다.

□ 중심어 : | 다층퍼셉트론 | 강하학습 | 최적학습률 |

### Abstract

This paper proposes optimal learning rates in the gradient descent training of multilayer perceptrons, which are a separate learning rate for weights associated with each neuron and a separate one for assigning virtual hidden targets associated with each training pattern. Effectiveness of the proposed error function was demonstrated for a handwritten digit recognition and an isolated-word recognition tasks and very fast learning convergence was obtained.

□ Keyword : | Multilayer Perceptrons | Gradient Descent | Optimal Learning Rates |

---

## 1. 서론

다층퍼셉트론(MLP: Multilayer Perceptron)은 층분한 수의 중간층 뉴런이 있으면 임의의 함수를 허용 오차 내에서 근사화 할 수 있다는 특성 때문에 패턴인식, 시계열 예측, 비선형 제어, 통신 등에 응용되고 있다. MLP의 학습으로는 EBP(Error Back Propagation)가 가장 널리 사용되는 데, 이는 고정된 학습률을 사용한

강하 학습법(gradient descent)이다[1]. 이 방법은 학습 속도가 느리기 때문에, 학습속도 가속을 위한 방법으로 여러 가지 오차함수가 제안되었다[2]. 그렇지만, 이 방법은 여전히 최적이지 아닌 학습률을 사용한 강하학습에 기반을 둔 것이다.

또 다른 방법이 conjugate gradient나 Newton 과 같은 2차 비선형 최적화 방법이다. 이 방법들의 가장 큰 약점은 많은 응용 문제에서 나타나는 Hessian 행렬의 나

쁜 특성과 Hessian에 관련된 많은 계산량 문제이다[3]. 이에 대한 해결책으로 LBL(layer-by-layer) 최적화 방법이 제안되었는데, 이는 MLP의 각 계층을 비선형 부분과 선형 부분으로 나누어 학습을 진행시킨다. 그렇지만, 이 방법은 중간층에 가상의 목표값을 설정할 때, 학습이 더 이상 진전되지 않는 정체 현상이 나타난다[4].

위와 같은 접근방법 이외에, EBP의 학습률을 가변시키는 방법도 제안되었다. EBP가 최적의 학습률을 지나게 되면 학습속도가 아주 빨라질 것이므로, 이 논문은 이 학습률 가변 방법에 초점을 맞춘다. 학습률 가변의 한 형태로 BD(Bold Driver) 방법이 있다[5]. 이 방법은 오차가 연속 감소하는 경우에는 학습률을 증가시키며, 반대의 경우에는 학습률을 감소시킨다. Jacobs는 각 가중치에 대한 학습률을 학습 진행에 따라 변동시키는 DD(delta-delta) 방법을 제안하였다[6]. 이 방법은 오차함수의 특정 가중치에 대한 미분이 학습의 진행 시 연속적으로 같은 부호를 지니면 학습률을 증가시키며, 그렇지 않으면 감소시킨다. 가중치들이 서로 독립이라는 가정 하에 이의 수정된 형태인 DBD(delta-bar-delta) 방법도 제안되었다. 그렇지만, 실제 가중치들은 독립이 아니다. 또한, Fahlman은 오차 표면을 가중치들의 다항식으로 근사화 시킨 quick-prop을 제안하였다[7]. 비록 이상과 같은 휴리스틱 방법들이 어떤 경우에는 학습속도의 향상을 가져오기도 하였지만, 이들은 여전히 최적 학습률을 찾아주지 못한다.

이 논문은 EBP 알고리즘에서 각 뉴런과 학습패턴에 대한 최적의 학습률을 찾는 방법을 제안한다. II장에서 MLP에 대하여 간략히 설명하고, III장에서 출력층 가중치에 대한 최적 학습률을 제안한다. IV장에서 중간층 가중치에 대한 최적 학습률을 제안하는 데, 이를 위하여 가상의 중간층 목표값이 할당되며 새로운 중간층 오차함수를 도입한다. V장에서 시뮬레이션을 통하여 제안한 방법의 효용성을 검증하고 VI장에서 결론을 맺는다.

## II. 다층퍼셉트론

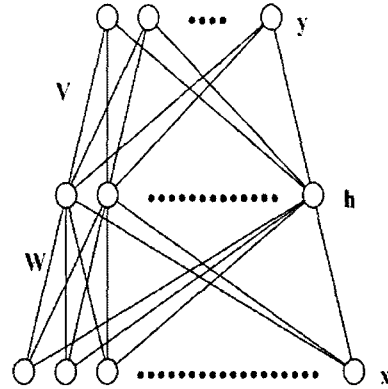


그림 1. 다층퍼셉트론 구조

다층퍼셉트론(MLP)이  $N$ 개의 입력 뉴런과  $H$ 개의 중간층 뉴런 및  $M$ 개의 출력 뉴런들로 구성되어 있다고 하자. 어떤 입력패턴  $\mathbf{x} = [x_1, x_2, \dots, x_N]$ 이 MLP에 입력되면,  $j$ 번째 중간층 뉴런의 값은

$$h_j = f(\tilde{h}_j) = \tanh(\tilde{h}_j/2), j = 1, 2, \dots, H \quad (1)$$

와 같이 주어진다. 여기서  $f(\cdot)$ 는 중간층 뉴런의 시그모이드 비선형 함수이며

$$\tilde{h}_j = \sum_{i=0}^N w_{ji}x_i \quad (2)$$

는 중간층 뉴런에 입력되는 가중치 합이다.  $w_{ji}$ 는  $x_i$ 와  $h_j$ 를 연결하는 중간층 가중치이며  $x_0 = 1$ 로 주어지기에  $w_{j0}$ 는 바이어스라고 불린다. 즉, 중간층 뉴런은 입력층 공간에 대하여 가중치 벡터를 이용하여 사상을 한 후 그 결과에 대하여 비선형 변환을 한 결과 값을 지닌다. 같은 형태로  $k$ 번째 출력 뉴런에 입력되는 가중치 합은

$$\hat{y}_k = \sum_{j=0}^H v_{kj}h_j, k = 1, 2, \dots, M \quad (3)$$

이고,  $v_{kj}$ 는  $h_j$ 와  $y_k$ 를 연결하는 출력층 가중치이고,  $h_0 = 1$ 이며  $v_{k0}$ 는 바이어스이다. 이 논문은 출력층 뉴런이 선형이라고 가정하여  $y_k = \hat{y}_k$ 로 둔다.

P개의 학습패턴  $\mathbf{x}^{(p)} (p=1, 2, \dots, P)$ 와 이들의 출력층 목표벡터  $\mathbf{t}^{(p)} = [t_1^{(p)}, t_2^{(p)}, \dots, t_M^{(p)}]$ 가 주어지면, 가중치들은 출력층에서

$$E^{out} = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^M (t_k^{(p)} - y_k^{(p)})^2 \quad (4)$$

로 주어지는 MSE(mean squared error)를 최소화시키도록 변경된다[8].

### III. 최적학습률

#### 1. 출력층 가중치 최적 학습률

EBP 알고리즘에서 출력층의 각 뉴런에 연결된 가중치에 별도의 학습률이 할당된 경우, 출력층 가중치는 오차함수의 미분에 의해 감소하는 방향으로 가중치가 변하도록

$$\Delta v_{kj} = -\eta_k^{out} \frac{\partial E^{out}}{\partial v_{kj}} = \eta_k^{out} \sum_{p=1}^P (t_k^{(p)} - y_k^{(p)}) h_j^{(p)} \quad (5)$$

와 같이 학습된다[2]. 여기서,  $\eta_k^{out}$  은  $y_k$  에 연관된 출력층 가중치들의 학습률이다. 수식 (5)와 같이 출력층 가중치들이 변경되고 난 후 MSE는

$$E^{out}(\eta_k^{out}) = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^M (t_k^{(p)} - y_k^{(p)})^2 + \eta_k^{out} \sum_{j=0}^H \frac{\partial E^{out}}{\partial v_{kj}} h_j^{(p)} \quad (6)$$

이 될 것이다. 따라서,  $h_j^{(p)}$ 가 고정되어 있다는 가정하

에  $\eta_k^{out}$ 의 최적값은  $\frac{\partial E^{out}(\eta_k^{out})}{\partial \eta_k^{out}} = 0$  이 되는 값이므로

$$\eta_k^{out} = \frac{\sum_{j=0}^H \left( \frac{\partial E^{out}}{\partial v_{kj}} \right)^2}{\sum_{p=1}^P \left( \sum_{j=0}^H \frac{\partial E^{out}}{\partial v_{kj}} h_j^{(p)} \right)^2}, k=1, 2, \dots, M \quad (7)$$

로 구해진다. 출력층 가중치는 이 최적 학습률을 이용하여 학습된다.

#### 2. 중간층 가중치 최적 학습률

중간층 가중치의 최적 학습률을 구하기 위하여 먼저 가상의 중간층 목표값을 할당한다. 그 다음, 이 가상의 목표값에 대한 중간층 오차함수를 제안한다. 그리고, 중간층 가중치의 변경량을 계산한 후, 가상의 목표값을 제거하여 EBP 학습에서 중간층 가중치 변경식에 대한 최적 학습률을 유도한다.

먼저,  $h_j^{(p)}$ 의 가상 목표값을 강하 학습법에 따라

$$z_j^{(p)} = h_j^{(p)} + \zeta_p \beta_j^{(p)} \quad (8)$$

와 같이 정한다. 여기서,

$$\beta_j^{(p)} \equiv -\frac{\partial E^{out}}{\partial h_j^{(p)}} = \sum_{k=1}^M (t_k^{(p)} - y_k^{(p)}) v_{kj} \quad (9)$$

이고,  $\zeta_p$ 는 가상의 중간층 목표값을 할당하기 위한 학습률이다.  $v_{kj}$ 가 고정되었다고 가정하면, 출력층의 MSE는  $\zeta_p$ 의 2차함수이므로  $\partial E^{out} / \partial \zeta_p = 0$  인 조건에서  $\zeta_p$ 의 최적값이

$$\zeta_p = \frac{\sum_{j=1}^H (\beta_j^{(p)})^2}{\sum_{k=1}^M \left( \sum_{j=1}^H v_{kj} \beta_j^{(p)} \right)^2}, p = 1, 2, \dots, P \quad (10)$$

와 같이 구해진다. (10)을 이용하여 (8)과 같이 가상의 중간층 목표값이 정해진 후,  $-1 < z_j^{(p)} < 1$ 을 만족하도록 절삭의 과정을 거친다. 그 다음  $h_j^{(p)}$ 에 대해 입력되는 가중치 합의 목표값이

$$\widehat{z}_j^{(p)} = f^{-1}(z_j^{(p)}) = 2 \tanh^{-1}(z_j^{(p)}) \quad (11)$$

와 같이 구해진다.

(11)과 같이 주어진 가상의 중간층 목표값에 대하여 새로운 오차함수를

$$E_n^{hid} = \frac{1}{2} \sum_{p=1}^P \sum_{j=1}^H (\widehat{z}_j^{(p)} - \widehat{n}_j^{(p)})^2 [f(\widehat{z}_j^{(p)})]^2 \quad (12)$$

와 같이 제안한다. 이 오차함수에서 마지막 항은 중간층 뉴런이 시그모이드 함수의 선형영역에 있는 지 혹은 포화 영역에 있는 지에 따라 중간층 가중치의 변경량을 조절하는 역할을 한다. 이는 MLP의 출력층에서 MSE를 최소화 시키는 것이 EBP 학습의 목적이므로,  $\widehat{z}_j^{(p)}$ 와  $\widehat{h}_j^{(p)}$ 의 거리보다는  $z_j^{(p)}$ 와  $h_j^{(p)}$ 의 거리를 고려하여 중간층 가중치가 변경되도록 하는 것이다. 만약  $\widehat{z}_j^{(p)}$ 와  $\widehat{h}_j^{(p)}$ 가 같은 포화영역에 위치하면, 비록  $\widehat{z}_j^{(p)}$ 와  $\widehat{h}_j^{(p)}$ 의 거리가 크더라도  $z_j^{(p)}$ 와  $h_j^{(p)}$ 의 거리는 작다. 이 경우  $\widehat{z}_j^{(p)}$ 와  $\widehat{h}_j^{(p)}$ 의 큰 거리에 따라 중간층 가중치를 변경하는 것은 불합리하다. 한편,  $\widehat{z}_j^{(p)}$ 가 선형영역에 있는 경우에 중간층 가중치는  $\widehat{z}_j^{(p)}$ 와  $\widehat{h}_j^{(p)}$ 의 거리에 따라 변경되어야 한다.

$E_n^{hid}$ 를 최소화시키기 위하여 중간층 가중치는

$$\Delta w_{ji} = -\eta_j^{hid} \frac{\partial E_n^{hid}}{\partial w_{ji}} \quad (13)$$

에 따라 변경될 것이다. 여기서,  $\eta_j^{hid}$ 는 학습률이고

$$\frac{\partial E_n^{hid}}{\partial w_{ji}} = -\sum_{p=1}^P (\widehat{z}_j^{(p)} - \widehat{n}_j^{(p)}) [f(\widehat{z}_j^{(p)})]^2 x_i^{(p)} \quad (14)$$

이다. (13)에 따라 변경된 중간층 가중치를 (12)에 대입하면

$$E_n^{hid}(\eta_j^{hid}) = \frac{1}{2} \sum_{p=1}^P \sum_{j=1}^H (\widehat{z}_j^{(p)} - \widehat{n}_j^{(p)})^2 + \eta_j^{hid} \sum_{i=0}^N \frac{\partial E_n^{hid}}{\partial w_{ji}} x_i^{(p)2} [f(\widehat{z}_j^{(p)})]^2 \quad (15)$$

이 된다. 따라서, 최적의  $\eta_j^{hid}$  ( $j = 1, 2, \dots, H$ )는 조건  $\partial E_n^{hid}(\eta_j^{hid}) / \partial \eta_j^{hid} = 0$ 에 의해

$$\eta_j^{hid} = \frac{\sum_{i=0}^N \left( \frac{\partial E_n^{hid}}{\partial w_{ji}} \right)^2}{\sum_{j=1}^H \left( \sum_{i=0}^N \frac{\partial E_n^{hid}}{\partial w_{ji}} x_i^{(p)} \right)^2 [f(\widehat{z}_j^{(p)})]^2} \quad (16)$$

와 같이 구해진다.

이제 가상의 중간층 목표값을 제거하여 EBP 알고리즘 형태를 지닌 중간층 가중치 변경식을 유도하겠다. (8)에서  $\zeta_p \beta_j^{(p)}$ 가 작은 값을 지닐 경우, 테일러 급수 전개에 의해 (11)은

$$\widehat{z}_j^{(p)} = f^{-1}(h_j^{(p)} + \zeta_p \beta_j^{(p)}) \approx \widehat{n}_j^{(p)} + \zeta_p \beta_j^{(p)} \frac{2}{1 - h_j^{(p)2}} \quad (17)$$

와 같이 근사화 된다. 또한,  $z_j^{(p)}$ 가 급격하게 변하지 않는다면

$$f(\hat{z}_j^{(p)}) \approx f(\hat{n}_j^{(p)}) = \frac{1 - \hat{h}_j^{(p)^2}}{2} \quad (18)$$

라고 두어도 된다. 이제 (17)과 (18)을 (14)에 대입하면

$$\frac{\partial E_n^{hid}}{\partial w_{ji}} \approx \sum_{p=1}^P \zeta_p \delta_j^{hid}(\mathbf{x}^{(p)}) x_i^{(p)} \quad (19)$$

와 같이 된다. 여기서,

$$\delta_j^{hid}(\mathbf{x}^{(p)}) = -\frac{\partial E^{out}}{\partial \hat{n}_j^{(p)}} = \beta_j^{(p)} f'(\hat{n}_j^{(p)}) \quad (20)$$

이다. 따라서, (13)은

$$\Delta w_{ji} \approx \eta_j^{hid} \sum_{p=1}^P \zeta_p \delta_j^{hid}(\mathbf{x}^{(p)}) x_i^{(p)} \quad (21)$$

와 같이 근사화 된다. 한편, 일반적인 EBP 학습에서 중간층 가중치의 변경식은

$$\Delta w_{ji} = \eta \sum_{p=1}^P \delta_j^{hid}(\mathbf{x}^{(p)}) x_i^{(p)} \quad (22)$$

와 같이 주어진다.

(22)를 (21)과 비교해보면, 이 논문에서 제안한 중간층 가중치 벡터의 변경식 (21)에는 두 종류의 최적 학습률이 있음을 알 수 있다. 하나는 각각의 학습패턴에 대한 가상의 중간층 목표값을 설정하는  $\zeta_p$  이고, 또 다른 하나는 새로운 중간층 오차함수를 최소화 시키는  $\eta_j^{hid}$  이다. 이 논문에서 제안한 EBP 알고리즘에서, 중간층 가중치는 목표값을 할당하지 않고 (21)에 따라 변경된다. 즉, 중간층 목표값은 새로운 중간층 오차함수를 이용하여 최적의 학습률을 유도하기 위한 가상의 목표값이며 실제로 존재하지는 않는다.

만약 이 논문에서 제안한 중간층 오차함수 (12) 대신에 일반적으로 사용되는 MSE를 사용하면, 중간층 가중치의 변경량은 같은 절차에 의해

$$\Delta w_{ji} \approx \eta_j^{hid} \sum_{p=1}^P \frac{\zeta_p}{f'(\hat{n}_j^{(p)})^2} \delta_j^{hid}(\mathbf{x}^{(p)}) x_i^{(p)} \quad (23)$$

와 같이 유도된다. (23)을 보면 시그모이드 함수의 기울기가 분모에 위치하는데, 이 값이 영에 가까우면 중간층 가중치의 변경량이 급격히 증가하게 된다. 결국, 중간층 뉴런들이 아주 심하게 포화되어 학습에 나쁜 영향을 미친다.

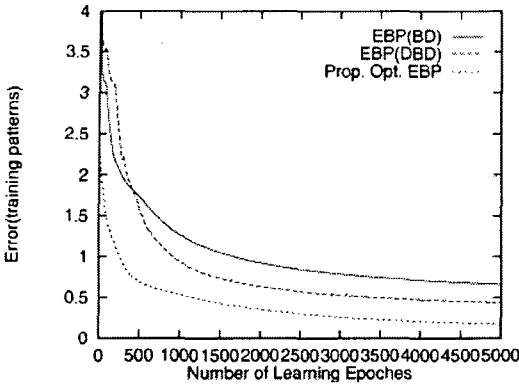
이 논문에서 제안한 최적 학습률을 이용한 EBP 알고리즘을 정리하면 다음과 같다.

- 단계 1: 중간층 가중치를 고정시키고서, (7)로 주어진 최적 학습률  $\eta_k^{out}$  을 이용하여 (5)에 따라 출력층 가중치를 변경시킨다.
- 단계 2:  $\beta_j^{(p)}$ ,  $\zeta_p$ ,  $\partial E_n^{hid} / \partial w_{ji}$  를 (9), (10), (19)와 같이 계산한다.
- 단계 3: (16)에 따라  $\eta_j^{hid}$  를 계산한 후, 중간층 가중치를 (21)에 따라 변경시킨다.

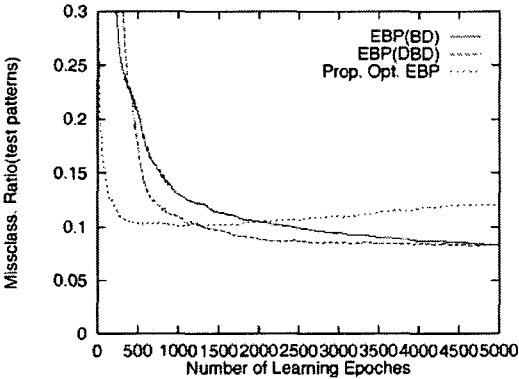
여기서 주의하여야 할 점은, 출력층 가중치의 변경  $\Delta v_{kj}$  와 중간층 가중치의 변경  $\Delta w_{ji}$  가 교차적으로 수행되어야 하는 것이다. 이는, 이 둘이 서로 다른 것이 고정되었다는 가정 하에 구해진 최적 값이기 때문이다.

#### IV. 시뮬레이션

제안한 방법의 효율성을 검증하기 위하여 필기체 숫자 인식 문제를 시뮬레이션하여 보았다. CEDAR 데이터베이스[9]에서 5000개의 숫자 영상을 추출하여 크기 정규화 과정을 거친 후 학습에 사용하였다. 숫자 영상의 크기는 12×12 픽셀이며 각 픽셀은 16가지 레벨의 값을 지닌다. MLP는 입력 144, 중간층 30, 출력층 10개의 뉴런들로 구성되었다.



(a) 학습패턴에 대한 MSE



(b) 시험패턴에 대한 오인식률

그림 2. 필기체 숫자인식 시뮬레이션 결과

MLP의 학습방법으로는 DB[5]를 이용한 EBP, DBD를 이용한 EBP, 그리고, 제안한 최적 학습률을 이용한 EBP를 사용하였다. BD를 이용한 EBP는 초기 학습률을 0.05로 하였고,  $\rho = 1.1$ 과  $\sigma = 0.5$ 를 사용하였다. DBD의 파라미터는  $\theta = 0.9$ ,  $\phi = 0.01$ ,  $\kappa = 1 \times 10^{-7}$ 로 정하였다. 그리고, 이 논문에서 제안한 학습 방법에서는 중간층 가중치들이 학습되지 않은 상황에서 출력층 가중치들이 매우 큰 값을 지니는 것을 방지하기 위하여, 학습의 두 번째 epoch까지는  $\eta_k^{out}$ 의 크기가 1을 넘지 않도록 제한하였다. 각각의 학습방법에 대하여  $[-1 \times 10^{-4}, 1 \times 10^4]$ 에서 균일분포를 지니도록 임의로 선택된 가중치들을 초기치로 하여 학습시키는 것을 4번 실시하여 평균치를 그림으로 그렸다.

[그림 2] (a)는 학습패턴에 대한 출력층의 MSE를 보

여준다. BD와 DBD 방법은 학습률이 최적 값이 아니므로 학습속도가 느리다. 이에 반하여, 제안한 방법은 최적 학습률을 사용하였기에 MSE가 학습초기부터 급격히 줄어든다. [그림 2] (b)는 2213개의 학습시키지 않은 시험패턴에 대한 오인식률을 보여준다. 학습이 진행될수록 시험패턴에 대한 오인식률이 증가하는 것은 학습패턴에 대한 학습이 필요 이상으로 많이 되어 시험패턴에 대한 특성은 오히려 나빠지기 때문이다.

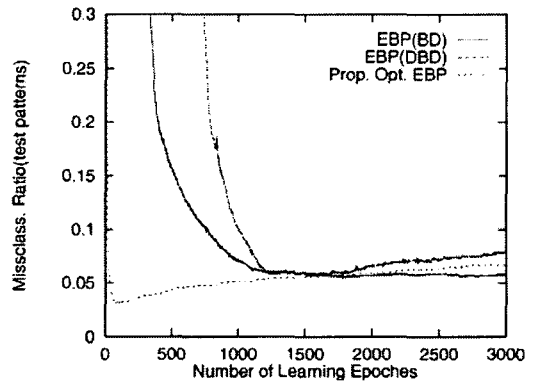


그림 3. 고립단어 인식문제에서 시험패턴에 대한 오인식률

더욱 더 철저한 효용성 검증을 위하여 고립단어 인식문제도 시뮬레이션하였다. 50단어가 인식대상이며, 900개의 학습패턴에서 1024차원의 ZCPA 특징[10]을 추출 후 50개의 중간층 뉴런을 지닌 MLP에 학습시켰다. [그림 3]은 1050개의 시험패턴에 대한 오인식률을 나타내는 데, 제안한 방법이 BD와 DBD보다 훨씬 더 학습속도가 빠름을 알 수 있다. 이 그림에서도 [그림 2] (a)에서와 마찬가지로 학습이 진행될수록 시험패턴에 대한 오인식률이 증가하는 현상이 나타난다. 이는 학습패턴에 대한 과도한 학습이 시험패턴에 대한 특성을 저하시키기 때문이다.

## V. 결론

이 논문에서는 MLP의 EBP 알고리즘에 대한 학습속도 향상의 방안으로 최적 학습률을 제안하였다. 출력층 가중치의 최적 학습률은 중간층 가중치가 고정되었다는

가정 하에 유도하였다. 중간층 가중치의 최적 학습률은 두 가지 성분으로 이루어짐을 보였다. 한 가지는 가상의 중간층 목표값을 할당하기 위하여 모든 학습패턴마다 다른 값을 지니는 형태로 유도되었으며, 또 다른 한 가지 성분은 제안한 중간층 오차함수를 최소화시키기 위한 성분이었다. 여기서, 새로운 중간층 오차함수는 중간층 가중치의 변경량이 중간층 뉴런 값의 시그모이드 함수에 위치한 영역에 따라 적절히 변화하도록 하는 성질을 지녔다.

필기체 숫자 인식과 고립 단어 인식 문제의 시물레이션으로 제안한 방법을 BD 및 DBD와 비교하였는데, 주장한 바와 같이 제안한 방법이 학습초기부터 MSE가 급격히 줄어듦을 확인할 수 있었다.

**참고 문헌**

[1] D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing. MIT Press, Cambridge, MA, 1986.

[2] S.-H. Oh, "Improving the error backpropagation algorithm with a modified error function," IEEE Trans. Neural Networks, Vol. 8, No. 3, pp.799-803, 1997.

[3] R. Parisi, E. D. Di Claudio, G. Orlan야, and B. D. rao, "A generalized learning paradigm exploiting the structure of feedforward neural networks," IEEE Trans. Neural Networks, Vol. 7, pp.1450-1459, 1996.

[4] G.-J. Wang and C.-C. Chen, "A fast multilayer neural networks training algorithm based on the layer-by-layer optimizing procedures," IEEE Trans. Neural Networks, Vol. 7, pp.768-775, 1996.

[5] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Accelerating the convergence of the back-propagation method," Biol. Cybern., Vol. 59, pp.257-263, 1988.

[6] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," Neural Networks, Vol. 1, pp.295-307, 1988.

[7] S. E. Fahlman, "Faster-learning variations on back-propagation: an empirical study," Proc. of the 1988 Connectionist Models Summer School, pp.38-51, 1988.

[8] 오상훈, "비선형 변환에 의한 중간층 뉴런의 상관 계수 감소", 한국콘텐츠학회 논문지 제3권 제3호, pp.98-102, 2003.

[9] J. J. Hull, "A database for handwritten text recognition research," IEEE Trans. Pat. Ana. Mach. Int., Vol. 16, No. 5, pp.550-554, May 1994.

[10] D.-S. Kim, S.-Y. Lee, and R. M. Kil, "Auditory processing of speech signals for robust speech recognition in real-world noisy environments," IEEE Trans. Speech and Audio Processing, Vol. 7, pp.55-69, 1999.

**저자 소개**

오 상 훈(Sang-Hoon Oh)

종신회원



- 1986년 2월 : 부산대학교 전자공학과(공학사)
- 1988년 2월 : 부산대학교 대학원 전자공학과(공학석사)
- 1999년 8월 : 한국과학기술원 전기 및 전자공학과(공학박사)

- 1988년 1월~1989년 12월 : LG 반도체(주) 사원
- 1990년 1월~1998년 6월 : 한국전자통신연구원 기초 기술연구부 및 이동통신기술연구소 선임연구원
- 1999년 8월~2000년 3월 : 한국과학기술원 뇌과학연구센터 연구원
- 2000년 4월~2000년 10월 : 일본 RIKEN, Brain Science Institute, Research Scientist
- 2000년 10월~2001년 10월 : (주)엑스텔테크놀로지 연구소장
- 2001년 11월~2002년 2월 : 한국과학기술원 초빙교수
- 2002년 3월~현재 : 목원대학교 정보통신전과학부 조교수

<관심분야> : 지능정보처리 알고리즘 개발 및 IT에의 응용, 독립성분분석, 음성신호 처리