

論文94-31B-8-22

Robustness를 향상시키기 위한 Hybrid 학습법칙을 갖는 다층구조 신경회로망

(Multi-layer Neural Network with Hybrid Learning Rules for Improved Robust Capability)

鄭 東 奎*, 李 壽 永*

(Dong-Gyu Jeong and Soo-Young Lee)

要 約

대부분의 신경회로망에서 네트워크의 동작은 weighted sum 으로부터 뉴런의 출력으로 비선형적으로 변환이 이루어진다. 이러한 비선형특성을 갖는 은닉층의 뉴런을 고찰하여 적절한 제한(restriction)을 가함으로써 기존의 다층퍼셉트론보다 잡음에 강한 특성을 갖도록 할 수 있으며, 더우기 이러한 제한사항을 잘 부여함으로써 기존의 오차역전파 학습법칙과 Hebb 학습법칙이 동시에 적용되는 복합학습법칙으로 네트워크를 학습시킬 수 있음을 보이하고자 한다. 즉, 시냅스 weight 에 제한을 가하여 네트워크 complexity 를 줄이는 기존의 방법과는 달리 은닉층의 뉴런에 부가적 제한성을 잘 부여함으로써 이 제한적 요소가 Hebb 의 학습법칙으로 연결될 수 있는데 이 복합학습법칙은 기존의 오차역전파 학습법칙보다 훨씬더 빠른 학습속도를 얻을 수 있으며, 아울러 은닉층의 오차도입에 따라 robustness 가 향상될 수 있음을 컴퓨터 시뮬레이션 예제에서 보여진다.

Abstract

In this paper we develop a hybrid learning rule to improve the robustness of multi-layer Perceptions. In most neural networks, the activation of a neuron is determined by a nonlinear transformation of the weighted sum of inputs to the neurons. Investigating the behaviour of activations of hidden layer neurons, a new learning algorithm is developed for improved robustness for multi-layer Perceptrons. Unlike other methods which reduce the network complexity by putting restrictions on synaptic weights, our method based on error-backpropagation increases the complexity of the underlying problem by imposing its saturation requirement on hidden layer neurons. We also found that the additional gradient-descent term for the requirement corresponds to the Hebbian rule, and our algorithm incorporates the Hebbian learning rule into the error back-propagation rule. Computer simulation demonstrates fast learning convergence as well as improved robustness for classification and hetero-association of patterns.

*正會員, 韓國科學技術院 電氣 및 電子工學科
(Dept. of Elec. Eng., KAIST)

接受日字 : 1993年 12月 20日

I. 서론

기존의 계산방식에 의해 문제를 해결하고자 할 때는 그 문제가 잘 정리되어 있어야 하고, 문제해결의 방법이 사전에 잘 정리되어 있어야 한다. 그러나, 신경망은 복잡하거나 잘 정리되어 있지 않는 문제, 즉 수학적 모델이나 해결방법을 찾기 어려운 문제도 효율적으로 적용될 수 있다. 신경망 모델 중 다층퍼셉트론^[1]은 복잡한 패턴분류 및 함수근사화(function approximation)을 하는데 우수한 성능을 보여준다.^[2,3] 그런데, 우수한 generalization 특성은 학습시키고자 하는 데이터의 크기, 문제의 복잡성(complexity), 그리고 네트워크의 복잡성등의 조합에 의해 얻어질 수 있다.^[4] 주어진 데이터수와 문제의 복잡도가 주어졌을때 네트워크의 크기(size)가 너무 작으면 그 문제를 정확하게 나타내기가 어렵다. 반면에 너무 큰 크기의 네트워크는 더 나쁜 generalization 특성을 보여준다. 이 경우에 generalization 능력을 향상시키기 위하여 시냅스 weight에 제한성(restriction)을 주거나 네트워크 complexity를 줄이기도 한다. 시냅스 weight pruning^[5]이나 국부적 연결(local connection)^[5-8], weight 소거(elimination)^[9] 등이 이러한 방식에 속한다. 최근에는 generalization 특성과 함께 잡음에 대한 robustness 연구가 또한 관심을 불러 일으키고 있다.^[10-11] Robustness 문제는 주어진 학습패턴에 잡음이 끼었을 때 본래 학습패턴을 얼마나 잘 찾아 내는가에 관한 것으로, 이 문제를 다루기 위해서는 비선형 특성의 네트워크 다이내믹스(network dynamics)에 대한 고찰이 필요하다.

본 논문에서는 위와는 다른 한가지 방법으로 잡음에 robust한 특징을 갖게 하는 다층구조신경회로망 학습방법을 제시하고자 한다. 즉, noise robustness를 향상시키기 위해 네트워크 크기나 weight에 제한을 가하는 기존의 방법들과는 달리, 은닉층의 뉴런에 제한을 가함으로써 잡음에 강한 특성을 갖게 하고 아울러 은닉층뉴런의 제한성을 적절히 가함으로써 오차역전파법칙과 Hebb의 학습법칙이 동시에 적용되는 복합(hybrid)방식이 이루어 질 수 있는 한가지 방법을 제시하고자 한다. 즉, 다층구조에서 학습이 이루어진 후 네트워크에 학습된 메모리패턴을 인가하였을 때 은닉층의 뉴런값이 시그모이드 함수의 zero crossing 점 부근의 선형성이 강한 영역에서 얻어진다면 잡음이 섞인 패턴이 인가될 때 자신의 메모리패턴을 찾기 어렵다. 반면에 학습후 메모리패턴이 인가되었을때 은닉층의 뉴런값이 시그모이드 함수의 포화영역

(saturation region)에서 얻어진다면 이 네트워크는 잡음에 강한 특성을 보여준다. 이러한 현상을 바탕으로 기존의 다층구조퍼셉트론에서 정의되는 출력층의 오차를 은닉층을 포함하여 얻어지는 오차로 새로이 정의함으로써 잡음에 강한 네트워크를 얻었다. 뿐만 아니라, 은닉층의 오차 정의에 사용되는 특수한 형태의 함수를 도입함으로써 학습시에 오차역전파기법과 Hebb법칙이 동시에 적용될 수 있는 복합방식을 얻을 수 있음을 보이고자 한다. 이러한, 복합방식이 기존의 오차역전파기법보다 훨씬 더 빠른 학습속도를 얻을 수 있고 동시에 잡음에 강한 특성을 가짐을 시뮬레이션 예제에서 보인다.

II. 은닉층 뉴런의 제한성

여기서 제안된 방법은 일반적인 다층구조 신경회로망에 적용이 가능하나 설명의 편의성을 위해 아래의 그림1 과 같이 2층구조를 갖는 퍼셉트론(1개 은닉층 포함)에 적용해 보기로 한다. 일반적으로 학습이 끝난 후에 메모리 패턴(=학습패턴)을 인가했을 때 은

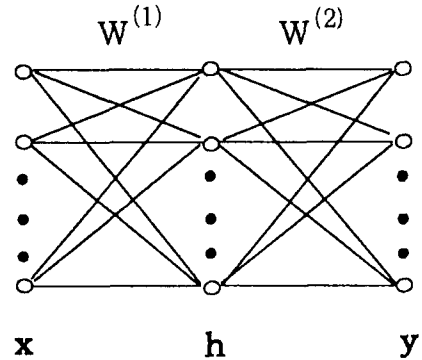


그림 1. 2층구조 신경회로망 구조
Fig. 1. 2-layer feedforward neural network.

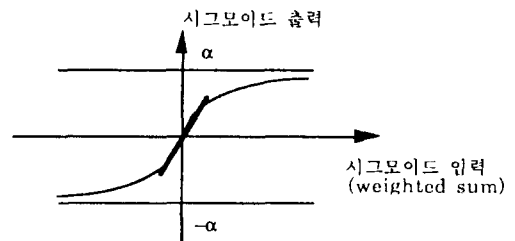


그림 2. 비선형 특성을 갖는 시그모이드함수
Fig. 2. Nonlinear Sigmoid function.

닉층의 뉴런값은 그림2 에서의 특성에 의해 비선형 시그모이드함수의 양극값 α 와 $-\alpha$ 사이의 임의의 아날로그값으로 분포되어진다. 학습이 끝난 후 이들 중 임의의 메모리패턴이 인가되었을때의 은닉층 뉴런값이, weighted sum으로부터 시그모이드함수의 좌우 대칭부분(그림2의 굵은 실선 근처부분)인, 선형 특성이 강한 영역에 있는 경우에는 그 메모리패턴에 잡음이 끼인 상태가 들어오면 다음 층으로 넘어갈때 많은 오차를 발생시킬 수도 있다. 특히, 은닉층 뉴런값이 weighted sum으로부터 부호에 민감한 부분(zero crossing point)으로부터 얻어지는 경우 약간의 잡음이 은닉층 뉴런의 값을 반대부호(+값 \rightarrow -값, -값 \rightarrow +값)을 가지게 되어 뒷 층으로 신호가 전달될때 많은 오차를 유발시킨다. 따라서, 학습후 메모리 패턴이 인가되었을 때 은닉층의 뉴런값이 가능한 한 weighted sum으로부터 시그모이드함수의 포화영역에서 얻어지는 경우 메모리 패턴들은 잡음에 robust하게 될 것이다. 이러한 문제와 결부시켜 은닉층의 뉴런에 적절한 제한(restriction)을 가함으로 잡음에 강한 특성을 주는 것과 동시에 복합방식의 학습법칙(hybrid learning rule)을 유도할 수 있다. 기존의 출력층에 오차사승치를 오차로 정의하는 것과는 달리, 다음과 같이 출력층과 은닉층을 동시에 고려한 새로운 형태의 오차를 정의하고자 한다.

$$\begin{aligned}
 E &= E_o + \gamma E_h \\
 &= \frac{1}{M} \sum_s E_o^s + \frac{\gamma}{M} \sum_s E_h^s \\
 E_o^s &= \frac{1}{2N_o} \sum_i (t_i^s - y_i^s)^2, \\
 E_h^s &= \frac{1}{2N_h} \sum_j [1 - (h_j^s)^2].
 \end{aligned} \tag{1}$$

여기서 E_o^s 와 E_h^s 는 정규화 에러로, E_o^s 는 기존의 퍼셉트론에 쓰이는 오차와 같고 E_h^s 는 새로 도입된 부가적 오차항으로 은닉층의 뉴런값으로부터 구해진다. 한편, M, N_o , 그리고 N_h 는 각각 학습패턴의 수, 출력층 뉴런의 수, 그리고, 은닉층의 뉴런수를 나타낸다. 은닉층의 뉴런 출력값(activation)에 의한 오차 E_h 와 출력층의 오차 E_o 는 각각 이들 수에 의해 정규화되어진다. t_i^s 와 y_i^s 는 각각 s번째 학습패턴에 대응하는 i번째 출력층 뉴런의 목표치(target value)와 실제값(activations)이며, h_j^s 는 대응하는 s번째 학습패턴에 대한 j번째 은닉층 뉴런의 값이다. 즉, 은닉층과 출력층의 뉴런값(activation)은 학습패턴이 입력에 인가되었을때 다음과 같이 표현되며,

$$\begin{aligned}
 h_j^s &= f(\bar{h}_j^s), j=1,2,\dots,N_h, s=1,2,\dots,M, \\
 y_i^s &= f(\bar{y}_i^s), i=1,2,\dots,N_o,
 \end{aligned}$$

$$\bar{h}_j^s = \sum_k w_{jk}^{(1)} x_k^s + \theta_j,$$

$$\text{그리고, } \bar{y}_i^s = \sum_j w_{ij}^{(2)} h_j^s + \theta_i. \tag{2}$$

$f(\cdot)$ 는 비선형 시그모이드함수이며, θ_i 와 θ_j 는 각각 출력층과 은닉층의 바이어스를 나타낸다. 만약, 은닉층의 뉴런값이 시그모이드 함수의 zero crossing 점 부근의 선형성이 강한 영역에 머무르면 입력의 잡음에 매우 민감하게 될 것이고 높은 은닉층의 오차 E_h^s 를 갖게 된다. 따라서, E_h 를 최소화함으로 네트워크를 포화영역 가까이에 놓이게 함으로 잡음에 대한 robustness를 향상시킬 수 있다. 여기서, γ 는 출력층의 오차 E_o 에 대한 은닉층의 오차 E_h 의 상대적 중요도를 나타내는 파라메타이다.

네트워크는 오차 E에 대해 아래와 같이 최대경사법(steepest descent method)를 도입하여 학습된다. 즉,

$$\begin{aligned}
 w_{jk}^{(1)}[n+1] &= w_{jk}^{(1)}[n] + \Delta w_{jk}^{(1)} \\
 &= w_{jk}^{(1)}[n] - \eta \frac{\partial E}{\partial w_{jk}^{(1)}} \\
 w_{ij}^{(2)}[n+1] &= w_{ij}^{(2)}[n] + \Delta w_{ij}^{(2)} \\
 &= w_{ij}^{(2)}[n] - \eta \frac{\partial E}{\partial w_{ij}^{(2)}}
 \end{aligned} \tag{3}$$

와 같이 표현된다. 그리고, η 는 학습률을 나타내는 것으로 은닉층과 출력층을 달리 줄 수도 있다. 마지막층의 weight updating식은 새로 도입된 부가적인 은닉층의 오차항에 의해 영향을 받지 않지만, 첫번째 층에서 각 weight에 대한 오차 E의 편미분항은 부가적인 새로운 항을 가지며 그 updating식은 다음과 같이 되어진다.

$$\begin{aligned}
 \Delta w_{jk}^{(1)} &= \frac{\eta}{MN_o} \sum_s x_k^s \delta_j^s f'(\bar{h}_j^s) + \frac{\eta\gamma}{MN_h} \sum_s x_k^s h_j^s f'(\bar{h}_j^s) \\
 &= \frac{\eta}{M} \sum_s (\delta_j^s / N_o + \gamma h_j^s / N_h) x_k^s f'(\bar{h}_j^s).
 \end{aligned} \tag{4}$$

단, δ_j^s 는 s번째 학습패턴을 j번째 은닉층뉴런에 대하여 오차를 역전파시키는 양이며, $f'(\cdot)$ 는 시그모이드 함수의 미분치로 항상 양의 값을 갖게 된다. 식(4)에서 두번째항이 은닉층 오차감소에 관계되는 부가적인 새로운 항으로 입력신호 x_k^s 와 은닉층의 뉴런값 h_j^s 가 작용하는 Hebb학습률을 유도시키는 것이다. 그리

고, Hebbian term이 $f'(h_j^*)$ 에 의해 곱해지는 것에 유의할 필요가 있으며, 이 양은 시냅스 weight가 무한대값으로 증가하는 것을 막아주는 중요한 역할을 한다.

한가지 중요한 사실은 네트워크의 학습능력과 γ 는 중요한 관계를 갖는다는 것이다. 일반적으로 네트워크의 complexity가 풀고자 하는 문제의 complexity보다 높을 경우에는 γ 값은 학습에 별 영향을 미치지 않는다. 그러나, 문제의 complexity가 네트워크의 complexity보다 더 높을 경우에는 학습 후 성능이 좋지 않을 수도 있으나 특별한 문제의 해결이 요구될 때는 이 경우를 배제할 수 없다. 이러한 경우에는 γ 가 너무 큰 값을 가지게 되면 학습이 이루어지지 않을 수도 있으며, 반면에 너무 작은 값 ($\gamma \approx 0.0$)을 가지게 되면 오차역전파기법과 차이를 보이지 않을 것이다. 하지만 문제에 따라서 최적의 γ 를 선택하는 것은 쉽지 않다. 따라서, 풀고자 하는 문제와 네트워크의 complexity가 주어질 때 learning epoch에 따라 나타나는 은닉층의 오차와 출력층의 오차에 따라 γ 를 변화시키는 adaptive γ 의 도입이 필요하다고 생각하며, 이는 현재 연구가 진행중이며 추후에 발표를 계획중이다.

이 새로운 알고리즘은 기존의 우수한 두가지 학습법칙인 오차역전파학습법과 Hebb학습법칙을 동시에 이용될 수 있는 복합학습법칙을 가능케 해 준다. Hebb의 법칙과 오차역전파기법을 도입한 기존의 연구는 콜라레이션 매트릭스 시냅스를 기초로 되었지만¹²⁾, 본 연구에서 도입된 새로운 알고리즘은 더욱 일반성을 갖는 adaptive Hebbian learning을 나타낸다.

III. 시뮬레이션 결과

새로이 정의된 은닉층오차의 개념을 도입하여 복합방식의 학습법칙을 유도하였으며, 이 학습법칙을 다층구조를 갖는 네트워크의 분류기 또는 연상기억메모리에 적용하였다. 10개의 이진패턴을 네트워크가 분류기와 연상기억메모리로 작용했을 때의 결과가 그림 3~그림 7에 보여진다. [부록] 그림 3 ~ 그림 6은 랜덤패턴에 의한 시뮬레이션 결과이고, 그림 7은 그림 8에 나타내어진 10개의 영문대문자 패턴을 분류기에 적용시켰을 때의 결과이다. 분류기의 경우 입력, 은닉층, 출력층의 뉴런수는 각각 35, 30, 10개이고, 연상기억메모리의 경우는 뉴런수가 각각 35, 30, 35개이다.

그림 3은 서로 다른 γ 에 대해 오차 E 를 도시한

것으로, γ 는 출력층 오차 E 에 대한 은닉층의 오차 E_h 의 상대적 중요도를 나타낸다. 여기서 제안된 학습법칙은 기존의 오차역전파($\gamma = 0$)보다 훨씬 빠른 수렴속도를 보인다. 그림 4는 $\gamma = 0.5$ 일 때 출력층오차, 은닉층오차, 그리고 전체오차를 학습회수에 따라 도시한 것으로, 보여진 바와 같이 학습초기에는 출력층의 오차 E_o 는 은닉층의 오차 γE_h 보다 더 크나 학습이 점점 많이 이루어져 감에 따라 더 적어진다. 식(4)에서 알 수 있듯이 학습의 후기에는 출력층오차에 비해 은닉층의 오차의 지배적 의존성(dominance)은 이 알고리즘이 γ 에 큰 영향을 받지 않음을 나타낸다. 이것은 학습초기에는 작은 값의 초기 weight에 기인하여 시냅스 weight는 Hebb법칙에 해당되는 updating양이 매우 작은 값을 가지는 반면에 오차역전파에 대응되는 항의 updating양 δ_j^* 가 큰 값을 가지므로 출력층의 오차감소에 많은 영향을 미치도록 변화되다가, 학습후기에는 입력 학습패턴에 기인한 출력층 뉴런의 값(activation)이 목표치(target value)에 가까운 쪽으로 감에 따라 δ_j^* 의 양에 비해 상대적으로 γh_j^* 의 값이 커지게 된다. 따라서, 학습후기에는 결국 γ 값이 작으나 크나 학습속도에 어느 정도 영향을 주나 weight의 상대적 값에는 별 영향을 주지 않는다. 좀더 자세히 살펴보면 γ 가 큰 값일 수록 학습초기에는 오차가 크나 이것은 초기에는 전체오차에 차지하는 비중이 크기 때문이며, 점점 더 학습이 되어감에 따라 γ 가 큰 쪽이 적은 쪽보다 오차가 약간 작아는 현상을 알수 있는데, 이 현상은 식(4)에서 Δw 에 γ 가 영향을 미치기 때문이다. 아울러, 제안된 복합학습법칙이 기존의 오차역전파학습법칙보다 빠른 학습속도를 보여 주는 것은 Hebb법칙에 따르는 항이 패턴의 특징추출 기능을 가지고 있기 때문으로 해석된다.

그림 5는 학습이 이루어진 후 인식률을 Hamming 거리의 함수로 도시되었는데, Hamming 거리는 학습패턴과 테스트패턴 사이에 몇 bit가 다른가를 나타내는 것이다. 사용된 테스트패턴은 10개의 학습패턴 각각에 대해 100개씩 랜덤하게 발생되었으며, 모든 랜덤패턴은 Hamming 거리를 만족하도록 구성되었다. 즉, 10개의 이진 랜덤패턴이 학습패턴으로 사용되었으며, 각 Hamming 거리에 대해 1000개의 학습패턴이 랜덤하게 발생되어 테스트용으로 사용되었다. 그림 5(a)는 학습이 오차 $E = 0.03$ 에서 멈추게 하였으며, 그림 5(b)는 $E = 0.001$ 에서 멈추게 하였다. 이들 결과는 기존의 오차역전파기법 ($\gamma = 0$)를 갖는 다층구조퍼셉트론에 비해 훨씬 더 잡음에 강한 분류능력을 가짐을 설명해 주고 있다. 학습이 완전하지

않았을 때는 그림5(a)에서 보여진 바와 같이 γ 에 의해 큰 영향을 받으나, 학습이 충분히 이루어진 경우에는 그림 5(b)에 보여진 바와 같이 인식률은 γ 에 덜 민감하다.

유사한 테스트로 그림 6은 연상기억메모리에 적용된 경우의 결과를 보여주며, 그림6(a)는 입력층에 10쌍의 이진랜덤패턴과 출력층에 다른 10쌍의 이진랜덤패턴을 오차를 $E = 0.03$ 까지 학습이 이루어지게한 후 연상을 시킨 것이고, 그림6(b)는 $E = 0.001$ 까지 학습이 이루어지게한 후 연상을 시켰을 때의 인식률을 보인 것인데, 이 결과 역시 분류기와 같은 현상을 보여준다. 한편, 그림7은 $\gamma = 0.0, 0.2$ 인 경우에 그림8에 나타난 10개의 영어 대문자 패턴을 분류기에 적용했을 때의 Hamming거리에 따른 인식률을 보여 주는데, 기존의 다층퍼셉트론($\gamma = 0.0$) 학습법에 의한 것보다 우수한 인식성능을 보여 준다.

IV. 결론

본 연구에서는 잡음에 강한 특성을 얻기 위해 2층의 feedforward 신경회로망에 이용될 수 있는 새로운 학습알고리즘을 제안하였다. 새로이 정의된 오차함수를 바탕으로 은닉층의 뉴런값을 시그모이드함수의 포화영역 가까운 곳에서 얻어질 수 있도록 유도함으로써 네트워크의 robustness 를 증가시킬 수 있었다. 더우기, 은닉층의 오차함수를 적절히 도입함으로써 최대경사법 (gradient-descent algorithm)을 도입시킬때 Hebb학습법칙과 오차역전파법이 동시에 사용될 수 있음을 보였으며, 이는 기존의 오차역전파법칙만 사용했을 때보다 훨씬 더 학습 속도가 빠름을 알 수 있었다. 이 복합학습법칙 (hybrid learning rule)은 일반적인 다층구조에 동일한 방법으로 적용이 가능하다.

부 록

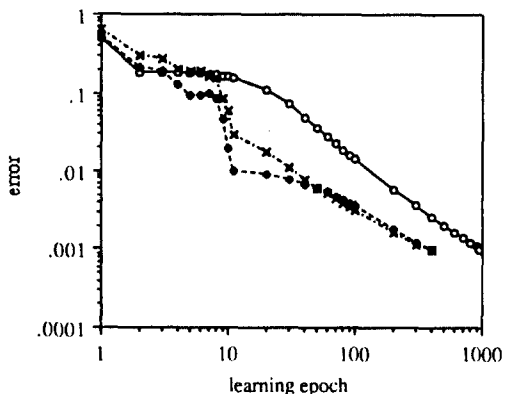


그림 3. 10개의 랜덤패턴을 분류기로 학습시켰을 때 학습반복회수에 따른 전체오차 E. "o", ".", 그리고 "x" 는 각각 $\gamma = 0.0, 0.1, 0.5$ 일때의 오차를 나타냄. $\gamma = 0.0$ 는 기존의 역전파기법에 해당하는 것임.

Fig. 3. Total error versus learning epoch for classification of 10 random patterns. Here "o", ".", and "x" denote cases for $\gamma = 0.0, 0.1$, and 0.5 , respectively. The case $\gamma = 0.0$ corresponds to standard error back-propagation learning rule.

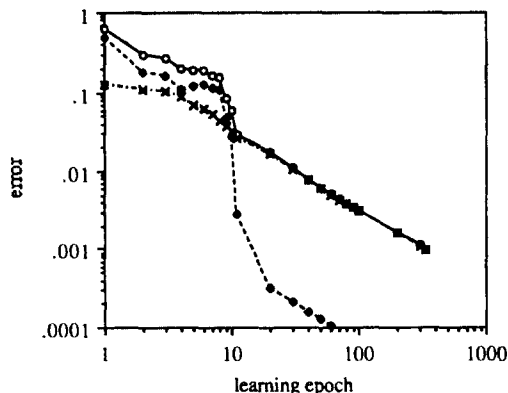
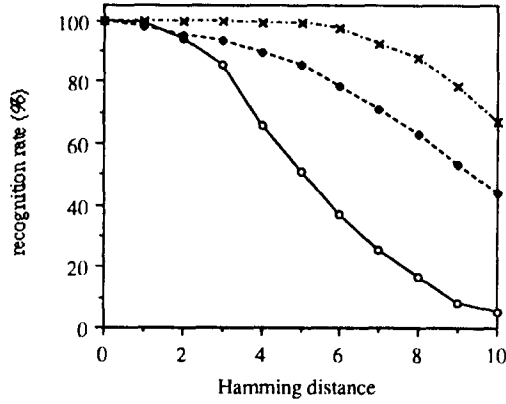
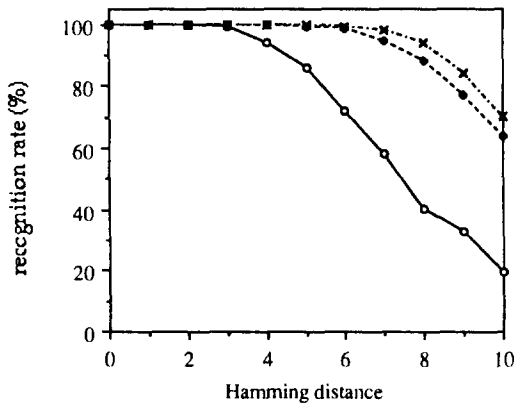


그림 4. 10개의 랜덤패턴을 $g = 0.5$ 상태에서 분류기로 학습시켰을 때 학습반복회수에 따른 오차. 여기서 "o", ".", 그리고 "x" 는 각각 전체오차 E, 출력층오차 E_o , 그리고 은닉층오차 γE_h ($\gamma = 0.5$) 를 나타냄.

Fig. 4. Error versus learning epoch for classification of 10 random patterns. Here "o", ".", and "x" denote the total error E, output error E_o , and hidden-layer error γE_h ($\gamma = 0.5$), respectively.



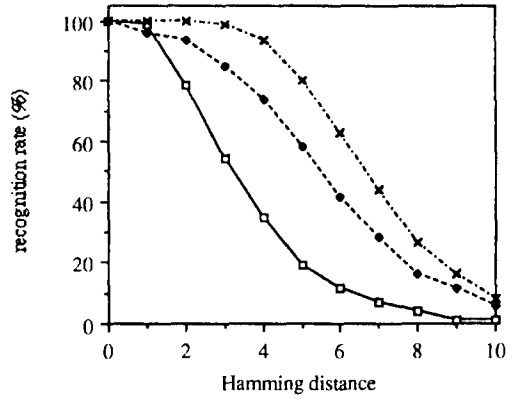
(a)



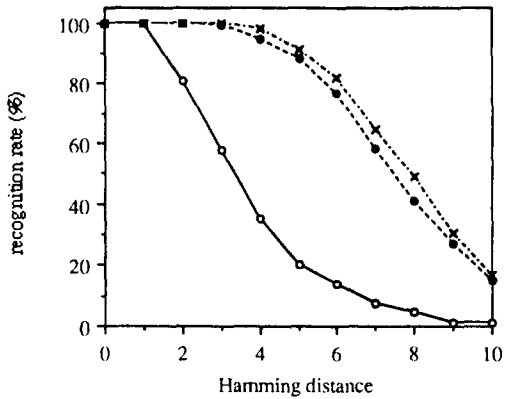
(b)

그림 5. 분류기로서 Hamming거리에 따른 인식률. (a) 전체오차 $E = 0.03$ 까지 학습시켰을때의 인식률, (b) $E = 0.001$ 까지 학습시켰을때의 인식률. 여기서 "o", "•", 그리고 "x"는 각각 $\gamma = 0.0, 0.1, 0.5$ 일때의 인식률임. $\gamma = 0.0$ 는 기존의 오차역전파 기법에 의한 결과임.

Fig. 5. Correct classification probabilities as functions of Hamming distance, (a) at total error $E = 0.03$ and (b) at $E = 0.001$. Here "o", "•", and "x" denote cases with $\gamma = 0.0, 0.1, 0.5$, respectively. The $\gamma = 0.0$ case corresponds to standard error back-propagation learning rule.



(a)



(b)

그림 6. 연상기억메모리로서 10쌍의 랜덤패턴을 다른 10쌍의 랜덤패턴으로 학습시킨후 Hamming 거리에 따른 인식률. (a) 전체오차 $E = 0.03$ 까지 학습시켰을 때의 인식률, (b) $E = 0.001$ 까지 학습시켰을때의 인식률임. 여기서 "o", "•", 그리고 "x"는 각각 $\gamma = 0.0, 0.1, 0.5$ 일때의 인식률임. $\gamma = 0.0$ 는 기존의 오차역전파 기법에 의한 결과임.

Fig. 6. Correct association probabilities for 10 random pattern pairs at input and output layer as functions of Hamming distance, (a) at total error $E = 0.03$ and (b) at $E = 0.001$. Here "o", "•", and "x" denote cases with $\gamma = 0.0, 0.1, 0.5$, respectively. The $\gamma = 0.0$ case corresponds to standard error back-propagation learning rule.

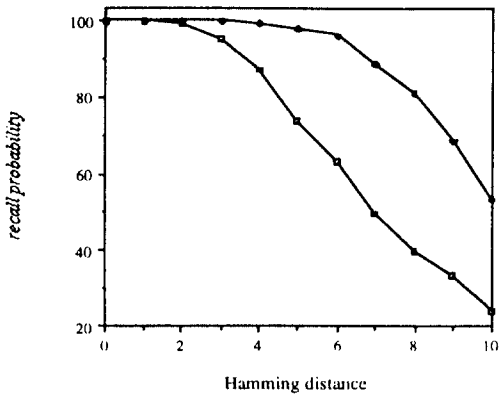


그림 7. (그림8)의 10개의 영문대문자 패턴을 분류기로 E=0.001까지 학습시켰을 때 Hamming거리에 따른 인식률. 여기서, □와 ◆는 각각 $\gamma = 0.0, 0.2$ 일 때의 인식률임.

Fig. 7. Correct classification probabilities as functions of Hamming distance for 10 alphabet patterns in Fig. 8. Here the □ and ◆ correspond to the $\gamma = 0.0, 0.2$, respectively.

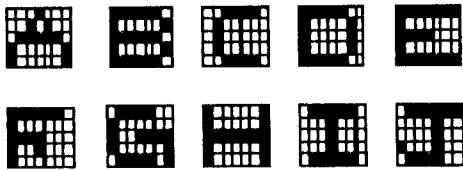


그림 8. 분류기에 사용된 10개의 이진(1, -1) 영문대문자 패턴 (A ~ J)

Fig. 8. The bipolar (1, -1) upper case alphabet patterns from "A" to "J" used in the classification problem.

参考文献

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Integral Representation by Error Propagation," D. E. Rumelhart & J. L. Maclell and (Eds.) *Parallel Distributed Processing : Explorations in the structure of cognitions*, Vol. 1 : Foundations, MIT Press, Chap.8, 1986.

[2] J. M. Zurda, *Introduction to Artificial Neural Systems*, West Publishing Comp., New York, pp.4 - 10, 1992.

[3] V. Kurkova, "Kolmogorov's Theorem and Multi-layer Neural Networks," *Neural Networks*, Vol.5, pp. 501 - 506, 1992.

[4] D. R. Hush, and B. G. Horne, *Progress in supervised neural networks*, IEEE Signal Processing Magazine (1), pp. 8~39, 1993.

[5] Y. Le Cun, Y. S. Denker, and S. A. Solla, "Optimal brain damage", *Advances in Neural Information Processing Systems.2*, Touretzky, ed. Morgan Kaufmann, pp. 598~605, 1990.

[6] K. Fukushima, "Necognitron : A hierachical neural network capable of visual pattern recognition," *Nural Networks* 1(2), pp. 119 - 130, 1988.

[7] K. Fukushima, "Improved generalization ability using constrained neural network architecture," in Proc. IJCNN'93 - Nagoya, pp. 2049 - 2054, 1993.

[8] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme recognition using time-deley neural networks," *IEEE Trans. Acoustics, Signal Process.* 37(3), pp. 328-339, 1989.

[9] A. S. Weigend, D. E. Rumelhart, and B. A. Huberman, "Back-propagation, weight elemation and time-series prediction," in Proc. of the 1990 Connectionist Models Summer School, Morgan Kaufmann, pp. 65 - 80, 1990.

[10] S. Piche, "Robustness of Feedforward Neural Networks," Proc. of the IJCNN, Balti-more, Maryland, USA, pp.II-346 - II-351, 1992.

[11] H. Y. Chow, "The robustness problem in automatic pattern recognition systems," Proc. of the IJCNN, Beijing, China, pp. I-705 - I-710, 1992.

[12] S. H. Koh, S. Y. Lee, J. S. Jang, and S. Y. Shin, "Merging Hebbian learning rule and least-mean-square error

algorithm for two layer neural networks," in Proc. of IJCNN'90 - WASH I, pp. 648 - 650, 1990.

著 者 紹 介

鄭東奎(正會員) 第30卷 第6號 參照
현재 한국과학기술원 박사과정

李壽永(正會員) 第30卷 第6號 參照
현재 한국과학기술원 교수