論文92-29B-1-12

점진적 학습영역 확장에 의한 다층인식자의 학습능력 향상

(Improvement of Learning Capabilities in Multilayer Perceptron by Progressively Enlarging the Learning Domain)

崔 悰 鎬,* 申 盛 植,* 崔 鎭 榮*

(Chong-Ho Choi, Seong Sik Shin, and Jin Young Choi)

要 約

다층인식자(multilayer perceptron)는 임의의 입,출력 사상(mapping)을 표현할 수 있는 능력을 가지고 있음이 알려져 있다. 그러나 오차역전파(error back-propagation) 방법에 의해 학습될 경우 목표함수의 복잡도 및 다층인식자 가중치(weight)의 초기치에 따라 국부 최소점에 빠지는 등 학습을 성공하지 못하거나 학습 속도가 매우 느린 경우가 발생한다. 본 논문에서는 기존의 오차역전파 방법에 의해 학습이 잘 안되는 함수를 학습시킬 때, 이를 극복하는 방법으로서 학습영역을 작은 영역에서 시작하여서 점진적으로 확장하면서 다층인식자를 학습시키는 학습방법을 제안한다. 이 방법은 단변수함수(function of one variable)를 다층인식자로 표현할 때 은닉층(hidden layer) 및 가중치의 역할분석을 통하여 착안되었다. 극점(extremal point)을 많이 가지고 있어 기존 오차역전파 방법으로 학습이 잘 안되는 단변수 및 이변수 함수에 대하여 시뮬레이션을 수행하여 제안된 학습방법의 타당성을 확인하였다.

Abstract

The multilayer perceptron, trained by the error back-propagation learning rule, has been known as a mapping network which can represent arbitrary functions. However depending on the complexity of a function and the initial weights of the multilayer perceptron, the error back-propagation learning may fall into a local minimum or a flat area which may require a long learning time or lead to unsuccessful learning. To solve such difficulties in training the multilayer perceptron by standard error back-propagation learning rule, the paper proposes a learning method which progressively enlarges the learning domain from a small area to the entire region. The proposed method is devised from the investigation on the roles of hidden nodes and connection weights in the multilayer perceptron which approximates a function of one variable. The validity of the proposed method was illustrated through simulations for a function of one variable and a function of two variable with many extremal points.

I. 서 론

다층인식자(multilayer perceptron)는 Rumelhart[1] 등에 의해 오차역전파(error back-propagation) 학습방법이 소개됨으로써 여러 분야에 응용되어 좋은 결과를 보여주고 있다. 이러한 다층인식자는 임의의 입,출력 사상(mapping)을 잘 표현할 수 있는 것으로 알려져 있다. 또한 은닉층(hidden layer)이 하나인 다층인식자로 임의의 연속함수를 충분히 정확하게

*正會員, 서울大學校 制御計測工學科
(Dept. of Cont. Inst. Eng., Seoul Nat'l Univ.)
接受日字 : 1991年 3月 15日

근사화시킬 수 있다는 것이 증명되었다.[2] 그러나 다층 인식자가 오차역전파 학습 방법에 의해 항상 임의의 함수를 학습할 수 있는 것은 아니다. 즉 오차역전파 학습법칙은 오차를 줄이는 방향으로 가중치(weight) 및 임계치(bias)를 조정해 나가는 것인데, 국부최소점(local minimum)의 영향으로 학습에 의해서 최적의 가중치 조합을 항상 찾아낸다는 것은 보장할 수 없는 것이다. 또한 학습이 성공적으로 이루어 지는 경우에도 함수에 따라서는 수렴 속도가 너무 느려지는 문제점이 나타난다. 이러한 문제점은 학습하고자 하는 대상 함수의 차원 및 복잡도(극대점, 극소점의 개수등 함수의 변화 정도)와 신경망 가중치의 초기치에 따라 발생한다. 일반적으로 단사함수(injective function)를 학습하고자 할 때는 고차원인 경우에도 학습이 비교적 잘 된다. 그런데 극대점, 극소점이 많은 비단사 함수(non-injective function)를 기존의 오차역전파 방법으로 학습할 경우, 다차원인 경우는 물로 일차원에서 조차도 국부최소점에 빠지는 등 학습이 용이하지 않다.

이런 문제점 해결을 위해 학습률(learning rate)의 변형,[3] 은닉층의 최적 노드수 추정,[4] 임계치의 초기값 선택,[5] 확률적 탐색법[6]등에 의한 학습성능 향상에 관한 많은 연구가 진행되어 학습속도의 개선 및 국부최소점 문제등에 대하여 많은 방법이 제시되었다. 이러한 방법중 임계치의 초기값 선택방법과 확률적 탐색법이 국부최소점 해결방법으로 가장 가능성 있는 결과이다. 그러나 임계치의 초기값 선정방법은 기존의 방법에 비해 좀더 좋은 수렴 결과를 보여 주지만 국부최소점 극복이 보장된 방법은 아니며 더 연구를 필요로 한다. 또한 확률적 탐색법은 무한히 학습을 반복한다는 가정하에 국부최소점을 극복할 수 있다는 것이 증명되어 있으나 최악의 경우 비효율적인 학습이 될 수 있어 이의 개선에 대한 연구도 필요한 상태이다. 따라서 이러한 문제에 대하여는 신경회로망의 구조, 특성 및 학습법칙 분석 등을 통하여 많은 연구가 진행되어야 할 것이다.

본 논문에서는 이러한 국부최소점 문제를 해결하는 한 방법으로, 학습 영역을 작은 영역에서 시작하여서 점진적으로 확장하면서 다층인식자를 학습시키는 학습방법을 제안한다. Ⅱ장에서 단변수함수(single variable function)를 다층인식자로 표현할 때 은닉층(hidden layer) 및 가중치의 역할을 분석하고, Ⅲ장에서 은닉층 노드 역할에 근거한 점진적 학습방법을 제시하며 점진적 학습시 가중치의 학습과정을 시뮬레이션을 통하여 분석한다. Ⅳ장에서 굴곡이 심하여 기존 오차역전파 방법으로 학습이 잘 안되는 일변수, 이변수 함수에 대하여 시뮬레이션을 수행하여 제안된 학습방법의 타당성을 확인한다.

Ⅱ. 다층인식자의 은닉층 및 가중치 역할

1. 다층인식자의 함수 표현 능력

그림1과 같이 은닉층의 하나인 다층인식자가 $f:\mathbf{R}^1 \Rightarrow \mathbf{R}^1$의 연속함수를 어떻게 사상하는지에 대하여 생각해보자. 폐구간 $[a, b]$에서 연속인 임의의 함수 $f(x)$는 부분별(piecewise) 미분가능하므로 도함수 $f'(x)$는 그림2에서와 같이 계단함수(step function)로 구성된 단순함수(simple function)로 근사화시킬 수 있다. 즉, 구간분할(partition) $a=x_0<x_1<\cdots<x_n=b$에 대해 근사 도함수 $\hat{f}'(x)$는 다음과 같이 주어진다.

$$\hat{f}'(x)=\sum_{i=1}^{n} f'(x_i^-)[u(x-x_{i-1})-u(x-x_i)],$$
$$x\in[a, b]. \tag{1}$$

여기서 $u(\cdot)$는 계단함수이고, $f'(x_i^-)$는 x_i에서의 좌극한 값이다. 그리고 $f(x)$가 연속이면 폐구간에서 유한하므로 근사함수는 다음으로 표현될 수 있다.

$$\begin{aligned}\hat{f}(x)&=\int_a^x \hat{f}'(\tau)d\tau+f(a)=\int_a^x \sum_{i=1}^{n} f'(x_i^-)[u(\tau-x_{i-1})\\&\quad-u(\tau-x_i)]d\tau+f(a)\\&=\sum_{i=1}^{n} f'(x_i^-)\int_a^x (u(\tau-x_{i-1})-u(\tau-x_i))d\tau\\&=\sum_{i=1}^{n}\alpha_i\,\phi_i(x)+f(a),\end{aligned} \tag{2}$$

여기서, $\phi_i(x)$, α_i 및 β_i는

$$\phi_i(x)=\begin{cases}0 & x<x_{i-1}\\ \beta_i x & x_{i-1}\le x\le x_i\\ 1 & x>x_i,\end{cases}$$

$$\alpha_i=\frac{f'(x_i^-)}{\beta_i},\quad \beta_i=\frac{1}{x_i-x_{i-1}}$$

으로 표현된다. 이때 $\phi_i(x)$는 sigmoid 함수를 선형화 시킨것으로 생각하면 다음으로 표현될 수 있다.

$$\phi_i(x)\cong sigmoid(\beta_i x+\gamma_i),$$

$$\gamma_i=\frac{-(x_i+x_{i-1})}{2(x_i-x_{i-1})}$$

여기서 구간을 충분히 세분하면 $f(x)$는 식(2)로 근사화가 가능하며 따라서 출력층 노드 전달함수가 선형이고 은닉층 노드 전달함수가 sigmoid인 다층인식자로 근사표현이 가능하다. 이때 첨자 i는 다층인식자의 은닉층 노드 각각을 나타내며 α_i는 출력층과 은닉층 i번째 노드사이의 연결가중치를 나타내며 γ_i는 은닉층 i번째 노드의 임계치를 나타낸다.

각각의 노드가 주어진 함수를 표현하는데 기여를 하고 있음을 알 수 있다. 은닉층 노드는 표현하고자 하는 함수의 특징을 추출하는 역할을 수행하는데, 은닉층 노드출력의 선형구간이 이에 해당되며 포화구간은 단순한 bias 역할만을 수행한다. 그림3에서 노드 3과 5는 역할이 중복되어 [−1, −0.5] 구간의 감소하는 특징을 나타내며, 노드1이 [0.5, 0.5] 구간의 증가하는 특징을 나타내며, 노드4가 [0.5, 1] 구간의 감소하는 특징을 주로 나타낸다. 그리고 위의 노드들만으로 생기는 약간의 오차를 나머지 노드가 보정하는 것이다. 이 때 각 노드가 맡아서 표현하는 구간의 구분은 함수의 극점, 즉 함수의 증감이 바뀌는 점에서 노드의 역할이 나뉘게 된다. 즉 함수(3)의 경우 증가하는 구간은 $-0.5 \leqq x \leqq 0.5$, 감소하는 구간은 $-1 \leqq x \leqq -0.5$, $0.5 \leqq x \leqq 1$로써, 증가하거나 감소하는 구간이 세 구간으로 나뉘므로 은닉층에 세 개의 노드만이 있으면 함수의 사상이 기본적으로 가능한 것이며 만약에 출력오차를 더 줄이려면 노드수를 증가시키면 되겠으나 그 영향은 앞의 3개의 노드에 비하여 훨씬 적다. 위의 예에서 두 노드의 출력이 비교적 작은 것은 이와 같은 이유때문이다. 그러나 은닉층 노드의 수를 2개로 한 경우에는 큰 출력오차를 주는 것을 확인할 수 있었는데 이는 앞에서 설명한 이유 때문이다.

Ⅲ. 점진적 학습영역 확장에 의한 학습방법

1. 학습방법 제시

앞에서 설명한 대로 함수를 다층인식자로 표현할 때 은닉층의 노드마다 담당하는 구간이 생기는데, 극값이 많은 경우에는 은닉층에서 노드별로 담당하는 구간이 나누어지는 과정에서 시간이 오래 걸리게 되고 또한 학습이 잘 이루어지지 않는 현상이 발생한다.

이런면을 고려하여 학습구간을 작게 시작하여서 점차로 확장시켜 나가는 방법을 생각해 보았다. 예를 들면, 1차원 영역에서 전체 학습구간을 [−1, 1]이라 할 때, 우선 전체의 20%(학습영역확장비율)인 [−0.2, 0.2] 구간을 학습시키고 학습이 잘되면 학습구간을 확장시켜[−0.4, 0.4] 구간을 학습시키는 방법으로 점차 그 구간을 확장시켜서 마지막에는 [−1, 1]의 전구간을 학습시키는 것이다. 그러므로써 초기 학습시 작은 구간의 학습만 시키면 되므로 복잡한 함수라 하더라도 우선은 상대적으로 간단한 함수에 대해서만 학습시키는 효과를 얻을 수 있고, 또한 이러한 이유로 전체 노드중 일부의 노드만이 적은 구간의 사상을 맡으면서 비교적 쉬운 학습을 할 수 있게 된다. 그리고 점차로 학습구간이 확장됨에 따라 이미 학습된 구간에 대해서는 미리 학습된 가중치에 의한 사상이 이루어지고, 새로 추가된 영역의 사상을 전 단계에서 맡은 역할이 없던 노드들이 맡게하는 것이다.

학습영역을 확장시켜 나가는 방법은 작은 영역에 대해 학습을 진행하여 미리 설정된 학습목표가 달성되면 영역을 확장한다. 그러나 학습을 진행중에 국부최소점에 빠진것과 같은 현상이 발생하면 더 이상의 학습은 비효율적이므로 이 때는 학습목표가 달성되지 못하더라도 학습영역을 확장한다. 이 경우 전체 학습영역에 도달하여도 설정된 학습목표를 달성하지 못하고 국부 최소점에 빠진것과 같은 현상이 발생할 수 있다. 이 때는 처음의 작은 영역부터 다시 점진적 영역확장 학습을 반복한다. 점진적 영역확장 학습 방법을 정리하면 다음과 같다.

Step1: 학습목표 g(오차제곱의 평균의 목표치), 학습영역 확장비율 r%, 국부최소점 판정지수 m 및 이의 계산을 위한 최근 sweep 수 s (최근 s sweep에서 오차의 표준편차(standard deviation of error)가 m 이하이면 국부최소점에 빠진것으로 판정한다.) 등의 값을 설정한다.

Step2: 전체학습영역의 r%에 대해 오차역전파 학습을 진행한다.

Step3: 학습결과인 오차제곱 평균이 학습목표 g 보다 작게 되거나, 국부최소점으로 판정되면 학습영역을 전체영역의 r%만큼 확장하여 학습을 진행한다.

Step4: 전체학습영역까지 확장된 후 학습결과가 g 보다 좋아지면 학습을 종결하고 그렇지 않고 국부최소점으로 판정되면 Step2 부터 학습을 반복한다.

입력이 2개 이상인 경우에는 은닉층노드로 들어오는 입력이 여러개이기 때문에 노드의 정확한 역할분담에 대한 분석을 하는 것은 어렵지만, 앞에서 살펴본 바와 같이 1차원 입력일 때 노드가 선형적인 역할을 하는 것이 직선상의 구간으로 구분되는데 비해, 2차원 입력일 때는 평면상의 구역으로 구분되어 지고 두개 노드의 역할이 상호 연관되어 있으며, 일반적으로 다차원 공간에서는 다차원 영역으로 구분된다고 볼 수 있다. 따라서 다차원 입력일 때의 학습에도 영역을 점진적으로 확장시켜 가면서 학습을 하게되면 일차원에서와 같이 좋은 학습결과를 기대할 수 있을 것이다.

2. 시뮬레이션에 의한 가중치 학습과정 분석

이 학습방법을 사용할 때 각 노드와 관련된 가중치가 학습되어 가는 과정을 살펴보기 위하여 다음의

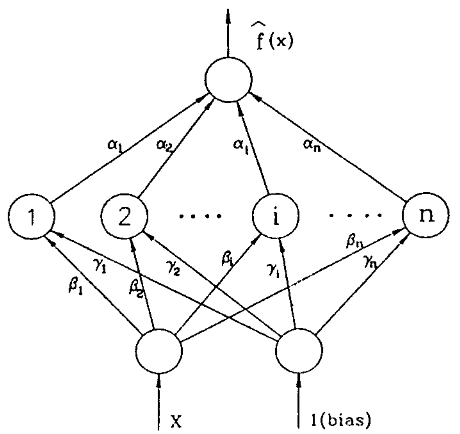


그림 1. 다층인식자의 구조
Fig. 1. Structure of a multilayer perceptron.

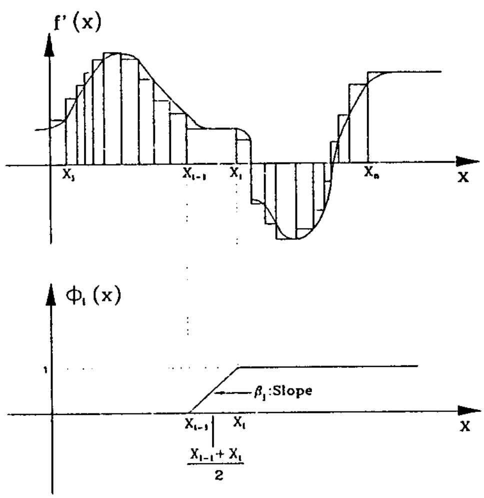


그림 2. 연속함수의 근사화
Fig. 2. Approximation of a continuous function.

2. 은닉층 및 가중치의 역할

은닉층 각 노드의 역할은 특정구간에 대한 사상을 담당한다. 즉 노드1은 $x_0 \leq x \leq x_1$ 구간을, 노드2는 $x_1 \leq x \leq x_2$의 구간을, 그리고 노드i는 $x_{i-1} \leq x \leq x_i$ 구간을 함수 f로 사상하는 역할을 담당한다. 따라서 은닉층 노드가 많을 수록 정확한 근사가 이루어 지는데 함수의 변화정도에 따라 근사화에 필요한 은닉층 노드수가 적절히 선택될 수 있다. 즉 완만한 함수는 적은 노드수를 필요로 한다. β_i는 i번째 구간길이의 역수를 나타내며 이는 sigmoid 함수에서 비례영역의 기울기를 나타낸다. 또한 γ_i/β_i는 비례영역의 중심을 α_i는 함수변화량과 구간길이의 곱을 나타낸다. β_i는 부호가 바뀌어 음수가 될 수 있으며, 이때 α_i의 부호도 같이 바뀌게 된다. 이때 $x_i - x_{i-1} = 1/\beta_i$ 이므로 가중치 β_i의 절대값이 작으면 넓은 범위의 비례영역을 가지며, 크면 좁은 입력범위에 대해서만 비례영역을 갖고, 나머지 입력 범위에서는 0또는 1의 포화치(saturated value)를 갖게 된다. 즉 가중치 β_i가 클수록 함수의 좁은 영역에서의 표현이 쉬워지는 것이다. 함수의 구간을 잘게 나누면 나눌수록 그 구간에서 함수는 직선에 가깝게 된다. 따라서 함수를 구간에 따라 직선들로 근사화시킬 수 있다면 그 구간을 표현하는데 적절한 수의 은닉층노드를 사용하여 표현할 수 있다.

3. 시뮬레이션에 의한 은닉층 노드의 역할분담 고찰

입력층과 출력층의 노드가 각각 한개씩인 다층인식자에서 은닉층의 각 노드마다 구간을 맡아서 사상하는 역할에 대해 시뮬레이션을 통하여 살펴보자. 학습에 사용된 목표함수는

$$f(x) = sin(\pi x), \quad -1 \leq x \leq 1 \tag{3}$$

이고, 은닉층의 노드는 5개를 사용했다. 출력의 최대오차가 0.01 보다 작을 때까지 학습하였으며, 입력x가 $-1.2 \leq x \leq 1.2$일 때 입력에 의해 각 은닉층노드가 출력에 미치는 영향을 알아보기 위해 그 영향을 그림3에 나타냈다.

그림3에서 보면 노드 1, 2, 3, 4, 의 출력은 선형부분의 기울기와 포화치들의 절대값이 비교적 크게 나타난 데 비해, 노드2의 출력은 상대적으로 작게 나타난 것을 볼 수 있다. 결과적으로 노드 2의 출력은 최종출력에 거의 영향을 끼치지 않는다는 것을 알 수 있다. 그 밖의 노드들의 출력의 기울기가 크기 때문에

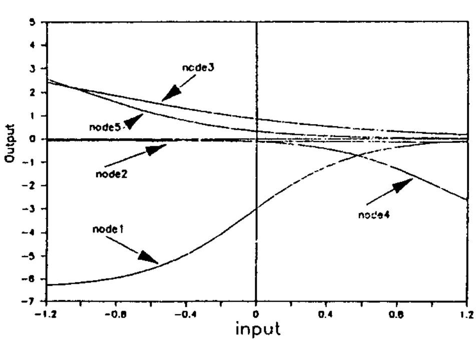


그림 3. 은닉층 노드의 역할
Fig. 3. Roles of hidden nodes.

목표함수를 사용하였다.

$$f(x) = sin(8\pi x), \quad -1 \leq x \leq 1 \qquad (4)$$

여기서 은닉층의 노드는 20개를 사용했다. 위에 설명한 대로 학습구간을 20%, 40%, 60%, 80%, 100%로 점점 확장시켜 나갔다. 그림4는 각 구간의 학습이 끝난 후의 가중치와 임계치의 분포를 나타낸 것이다.

그림4의 점들 옆에 있는 숫자는 은닉층 각 노드의 번호이다. 각 점들의 수평 좌표는 은닉층 노드의 임계치 γ_i를 해당노드와 입력간의 가중치 β_i로 나눈 값으로 은닉층 노드 출력에서 비례영역의 중심, 즉 sigmoid 함수의 출력값이 0.5가 되는 점이다. 그리고 수직좌표는 가중치 β_i로 은닉층 노드 출력에서 비례영역의 기울기가 된다. 즉, 그림4에서 각 점의 수평좌표가 학습영역 밖에 존재하거나, 수직좌표의 절대값이 작다면 해당노드의 역할은 거의 없다고 볼 수 있다.

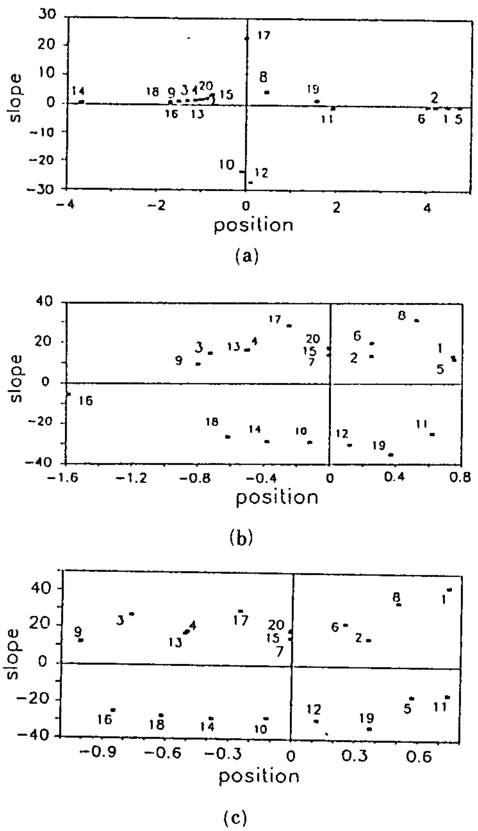


그림 4. 구간별 학습시 가중치의 변화
(a) 전체 구간의 20%를 학습시켰을 때
(b) 전체 구간의 80%를 학습시켰을 때
(c) 전체 구간의 100%를 학습시켰을 때

Fig. 4. Variation of weights in training with proposed method.
(a) after training 20% of input domain,
(b) after training 80% of input domain,
(c) after training 100% of input domain.

그림4(a)에서 보면 1차 학습구간 (전체구간의 20%)에서는 대부분이 작은 가중치를 가지고 있고, 소수의 노드만이 0근처에서 큰 가중치를 가지고 있는 것을 볼 수 있다. 이것은 표현해야 할 함수가 [-0.2, 0.2] 구간에서는 극값을 많이 갖고 있지 않으므로 소수의 노드에게만 역할 분담이 이루어 진 것이다. 그리고 학습영역이 넓어질수록 큰 가중치를 갖는 노드들이 점차로 증가하는 현상을 발견할 수 있다. 즉, 작은 영역의 함수를 식별할 때는 사용되지 않던 노드들이 확장된 구간에서 역할을 맡게 되는 것이다.

이제 노드의 역할 분담과정을 살펴보자. 먼저 가중치와 임계치의 초기값에 따라서 선정된 노드들의 가중치가 증가하면서 각 증감구간을 맡게 되고 학습영역이 넓어지면서 새로운 노드들이 추가로 역할을 맡게 되는데, 이때 새로이 역할을 맡게된 노드들을 살펴보면 역할분담전에 γ_i/β_i가 새로 역할을 맡아야 할 구간의 중심 근처에 있으면서 필요한 부호와 같은 부호의 가중치중 절대값이 큰 노드가 선택되는 것을 볼 수 있다. 즉 그림4의 (b)와 (c)를 비교해 보면 학습영역이 [-0.8, 0.8]에서 [-1, 1]로 확장되면서 (b)에서 활용되지 않던 노드 16이 γ_{16}/β_{16}가 [-1, 1] 근처로 오고 β_{16}도 증가해서, 감소하는 부분의 사상을 맡게 되었으며, 노드9도 (b)에서는 노드3과 같이 증가하는 부분의 사상을 담당하다가 학습영역이 확장되면서 출력의 중심이 좌측으로 이동하고 가중치도 증가하면서 독립적인 역할을 맡게 되었다. 그와 함께 노드 3도 노드9가 분리됨으로써 가중치가 더욱 증가한 현상을 볼 수 있다. 그림4(b)의 오른쪽 영역에서도 역시 학습영역이 확장되면서 노드1과 5가 함께 수행하던 역할을 노드1이 홀로 맡게되면서 가중치가 증가했고 노드5는 가중치가 음수인 노드가 주위에 없으므로 가중치가 양수에서 음수로 바뀌면서 새로운 역할을 맡게 된 현상을 볼 수 있다.

그리고 그림4(c)에서 보면 x=0 부근에 노드7, 15와 20이 같이 중복되어 있고, x=-0.5 부근에 노드 4와 13이 같이 중복되어 있음을 볼 수 있다. 이와 같은 현상은 함수의 증감을 식별하기 위해 필요한 노드수보다 더 많은 노드가 할당되었기 때문에 나타나는 현상이다. 즉, 중복되어 있는 노드들을 제거하고 하나씩만 남긴채로 학습이 진행되어도 거의 같은 학습성능을 얻을 수 있다. 함수(4)의 증감구간은 17개로

나눠는데, 은닉층의 노드는 20개이므로 3개를 제거해도 함수의 표현이 가능한 것이다. 따라서 그림 4와 같은 현상은 중요하지 않은 노드를 제거하는데 좋은 판단기준이 될 것이다.

Ⅳ. 학습성능 비교

1. 단변함함수 학습성능

단변수 함수를 학습할 때, 제시된 학습방법과 기존 학습방법 사이의 성능비교를 위해 극대·극소점이 20개인 다음의 목표함수를 사용했다.

$$f(x)=sin(10\pi x),\quad -1\leq x\leq 1 \qquad (5)$$

학습시 은닉층이 하나이고 은닉층의 노드는 20개인 다층인식자를 사용했으며, 점진적 영역 확장 학습방법에서 학습목표는 $g=0.01$로, 영역확장비율은 $r=20\%$로, 국부최소점 판정지수와 최근 sweep 수는 각각 $m=10^{-5}$, $s=500$sweep으로 선택하였다. 한 sweep 학습시 전체 학습영역에서 200개의 random 입력패턴을 사용하였다. 학습영역확장에 의한 학습시에는 학습영역에 비례하여 한 sweep의 입력패턴 수를 결정하였다. 예를 들면 40%의 학습영역 학습시는 전체 입력패턴수 200개의 40%인 80개의 입력패턴을 사용한다.

그림5는 기존의 오차역전파 방법과 제안된 영역확장 학습방법의 학습 성능 결과를 보여준다. 그림5의 수평축은 학습시간 t를 나타내며 수직축은 출력오차 제곱의 평균(mean square error)을 나타낸다. 그런데 점진적 영역 확장 학습방법에서는 sweep 마다 입력패턴 수가 다르므로 한 sweep 학습에 걸리는 시간이 다르다. 따라서 학습속도를 비교하기위해 전체영역의 입력패턴 수인 200개 패턴을 한번씩 학습하는 시간을 단위로 하여 학습시간 t를 다음과 같이 정의하였다.

$$t=\frac{1}{p}\sum_{i=1}^{c} p_i\ s_i, \qquad (6)$$

여기서, p: 전체학습영역의 입력패턴 수
c: t시간까지의 학습영역 변경횟수
p_i: i번째 학습영역에서의 입력패턴 수
s_i: i번째 학습영역에서의 학습 sweep 수

그림5에서 보면 기존의 방법으로는 국부 최소점에 빠진것과 같이 학습이 잘 안되는 함수를 점진적 학습영역확장 학습방법에 의해 성공적으로 학습이 되었음을 알 수 있다. 여기서 학습영역 확장비율과 국부최소점 판정수에 따라 수렴속도가 다를 수 있으므로 이들 값의 선택도 중요한 변수가 된다. 일반적으로 굴곡이 심한 함수일수록 학습영역 확장비율을 작게 해야 학습이 잘 되는데 너무 작게 했을 경우는 전체학습영역에 도달하는데 많은 시간이 소요될 수 있다. 따라서 적절한 값의 선택이 필요하며, 시뮬레이션을 수행하면서 얻어진 경험으로 보면 함수의 극점이 3~4개 포함되는 영역단위로 확장하는 것이 적절하였다. 함수의 극점을 기준으로 은닉층의 역할이 분담됨을 고려할 때, 은닉층 노드 3~4개의 역할은 동시에 학습이 가능함을 알 수 있다. 국부최소점 판정지수도 너무 크게하면 특정구간에서 충분한 학습이 되기도 전에 학습영역을 변경하여 영역별 학습의 효과를 줄일 수 있으며, 너무 작으면 국부최소점 상태에서 너무 많은 시간을 소요하여 전체적으로 학습완료 시간이 길어지는 단점이 있어, 이 값의 선택도 학습속도에 미치는 중요한 변수가 된다.

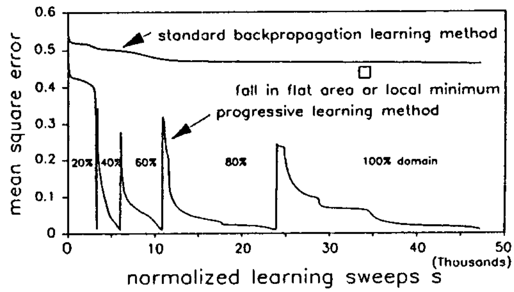


그림 5. 단변수함수의 학습성능
Fig. 5. Learning capabilities for a function of one variable.

2. 다변수 함수 학습성능

다변수 함수 학습시에도 점진적 학습영역 확장에 의한 학습방법이 유용함을 보이기 위해, 극점이 56개인 다음의 이차원 입력을 갖는 함수를 사용했다.

$$f(x,y)=sin(3\pi x)\,cos(3\pi y),\quad -1\leq x\leq 1,\quad -1\leq y\leq 1 \qquad (7)$$

학습에 사용된 다층인식자는 그림1의 구조에서 입력층의 노드는 2개, 은닉층의 노드는 50개를 사용했으며, 점진적 영역 확장 학습방법에서 학습 목표는 $g=0.01$로, 영역확장비율은 $r=20\%$로, 국부최소점 판정지수 및 최근 sweep 수는 각각 $m=10^{-5}$ 및 $s=100$sweep으로 선택하였다. 한 sweep 학습시 전체 학습영역에서 400개의 random 입력패턴을 사용하였다. 영역확장은 원점에서 정방형으로 확장하였다. 그림6에서 보듯이 2변수 함수 학습에서도 제안된 방법이

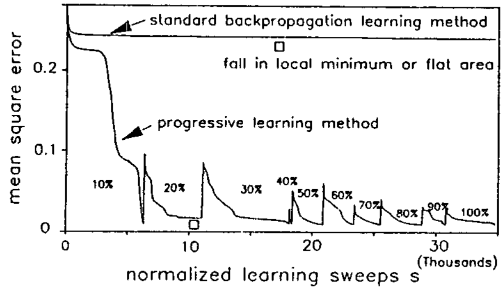


그림 6. 이변수함수의 학습성능
Fig. 6. Learning capabilities for a function of two variable.

타당함을 알 수 있으며 일반적으로 다변수 함수 학습시에도 적용될 수 있을 것으로 기대된다.

V. 결 론

다층인식자가 단변수 함수를 표현하는데 있어서 각 노드의 역할을 가중치 및 임계치를 통하여 설명하였다. 이를 근거로 기존의 오차역전파 학습방법에 의해 국부최소점에 빠지는 등 학습이 잘되지 않는 함수를 더 잘 학습할 수 있는 방법으로서, 학습 영역을 점진적으로 확장시켜 나가는 방법을 제안했다. 극점이 20개인 일변수 함수 및 극점이 56개인 이변수 함수에 대해 시뮬레이션을 수행하여 제안된 방법의 타당성을 확인하였으며, 복잡한 사상 관계를 다층인식자로 나타낼 때 매우 유용하게 사용될 수 있을 것으로 예상된다. 어린이 학습시 또는 동물의 훈련시에 어려운 문제를 한번에 학습하면 학습이 잘 안되지만, 쉬운 문제부터 점진적으로 학습하면 학습이 잘 진행되는 것은 잘 알려져 있다. 이와 연관하여 생각하면 제안된 점진적 학습방법이 다층인식자 뿐아니라 다른 인공신경회로망에도 잘 적용될 수 있을 것으로 생각된다. 또한 이 방법은 기존의 어떤 학습방법과도 병행사용이 용이하며 기존 방법의 학습성능을 높여줄 것이다. 앞으로 다차원 다층인식자의 은닉층 노드 및 가중치의 역할 분석과 특성 연구를 통하여 다층인식자의 사상 능력을 더 잘 이해할 것으로 생각되며 이 분야의 더 많은 연구가 기대된다.

參 考 文 獻

[1] D.E. Rumelhart and J.L. McClelland, "Parallel distributed processing : Explorations in the microstructure of cognition," MIT Press, 1986.

[2] Ken-ichi Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks,* Pergamon Press, vol. 2, pp. 183-192, 1989.

[3] J.R. Chen and P. Mars, "Stepsize variation methods for accelerating the back-propagation algorithm," *Int'l Joint Conf. on Neural Networks,* vol. 1, pp. 601-604, Washington, D.C., Jan. 1990.

[4] M. Gutierrez, J. Wang, and R. Grondin, "Estimating Hidden Unit Number for Two-Layer Perceptrons," *Int'l Joint Conf. on Neural Networks,* vol. 1, pp. 677-681, 1989.

[5] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosting initial values of the adaptive weights," *IEEE Int'l Joint Conf. on Neural Networks,* vol. 3, pp. 21-26, San Diego, California, June 1990.

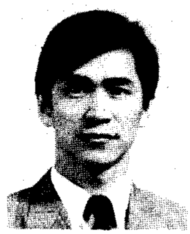
[6] N. Baba, "A Hybrid Algorithm for Finding the Global Minimum of Error Function of Neural Networks," *IEEE Int'l Joint Conf. on Neural Networks,* vol. 1, pp. 585-588, Washington, D.C., Jan. 1990.

[7] Philip D. Wasserman, *Neural computing,* Van Nostrand Reinhold, 1989

著 者 紹 介

崔 悰 鎬 (正會員) 第28卷 B編 第8號 參照
현재 서울대학교 제어계측공학과 교수

申 盛 植(正會員)
1989年 2月 서울대학교 제어계측공학과 학사. 1991年 2月 서울대학교 제어계측공학과 석사. 1991年 1月~현재 금성사 연구원. 주관심분야는 신경회로망, 제어시스템 등임.



崔 鎭 榮(正會員)
1982年 2月 서울대학교 제어계측공학과 학사. 1984年 2月 서울대학교 제어계측공학과 석사. 1984年 3月~현재 한국전자통신연구소 연구원. 1989年 9月~현재 서울대학교 제어계측공학과 박사과정. 주관심분야는 신경회로망 및 응용, 통신시스템, 제어시스템 등임.