

論文2001-38CI-6-3

활성화 함수의 이득 가변화를 이용한 역전파 알고리즘의 성능개선

(The Performance Improvement of Backpropagation Algorithm using the Gain Variable of Activation Function)

鄭盛夫*, 李賢寬**, 嚴基煥***

(Sung Boo Chung, Hyun Kwan Lee, and Ki Hwan Eom)

요 약

일반적인 역전파 알고리즘의 여러 가지 문제점들을 개선하기 위하여 활성화 함수의 이득을 퍼지 로직 시스템을 이용하여 자동 조절하는 방식을 제안하였다. 퍼지 로직 시스템을 구성하기 위하여 먼저 활성화 함수의 이득의 변화가 학습율, 연결강도, 바이어스 등의 변화와 등가인 관계를 조사하였다. 퍼지 로직 시스템의 입력은 마지막층에 대한 오차의 감도와 은닉층에 대한 오차의 평균 감도를 사용하였고, 출력은 활성화 함수의 이득을 사용하였다. 제안한 방식과 일반적인 역전파 알고리즘을 패리티 문제, 함수 근사화 문제 및 패턴 인식 문제등에 대하여 시뮬레이션하여 비교 검토한 결과 수렴비, 평균 학습 반복수, 정밀도 및 새로운 입력에 대한 원하는 오차 범위의 출력을 얻는 등의 성능이 개선됨을 알았다.

Abstract

In order to improve the several problems of the general backpropagation, we propose a method using a fuzzy logic system for automatic tuning of the activation function gain in the backpropagation. First, we researched that the changing of the gain of sigmoid function is equivalent to changing the learning rate, the weights, and the biases. The inputs of the fuzzy logic system were the sensitivity of error respect to the last layer and the mean sensitivity of error respect to the hidden layer, and the output was the gain of the sigmoid function. In order to verify the effectiveness of the proposed method, we performed simulations on the parity problem, function approximation, and pattern recognition. The results show that the proposed method has considerably improved the performance compared to the general backpropagation.

I. 서 론

최근 공학이나 과학뿐만이 아니라 금융, 재정, 경비,

* 正會員, 瑞逸大學

(Seoil College)

** 正會員, 湖南大學校

(Honam University)

*** 正會員, 東國大學校

(Dongguk University)

接受日字:2001年6月4日, 수정완료일:2001年10月9日

의학 등 사회 전 분야에 걸쳐 신경회로망의 응용이 확산되고 있으며, 특히 다층 신경회로망은 비선형 함수근사, 패턴인식, 음성합성, 예측, 최적화, 제어 등 많은 분야에 응용되고 있다^[1-2]. 이러한 다층 신경 회로망은 단순한 기능을 가진 무수히 많은 뉴런 또는 처리 소자들이 병렬 연결된 연산 구조로 즉 입력층, 은닉층, 출력층으로 구성되며, 학습이나 훈련을 통해 연결 강도를 조정함으로써 유연성이나 적응 능력이 뛰어나다. 또한 다층 신경회로망에서 일반적으로 많이 사용하는 학습 알고리즘은 역전파 알고리즘(Backpropagation algorithm)으로, 목표 출력과 네트워크의 실제의 출력의 차의 자

승 합을 목적함수로 하여 그 목적함수의 최소치를 스텝 폭, 고정된 최급강하법(Steepest descent method)으로 구하는 것이다. 최소치를 계산하는 알고리즘은 이외에도 Newton's method, Conjugate gradient method, Levenberg-Marquardt algorithm과 같은 간접탐색법과 nongradient인 Powell 방법과 같은 직접 탐색법으로 대분된다^[2].

그러나 역전파 알고리즘은 다층 신경회로망의 구조에서 은닉층 수의 증가 및 입력층, 은닉층 및 출력층의 뉴런 수가 증가하면 학습과정에 많은 시간이 소요되며, 부분 최소 점에 빠질 수 있고, 최적의 네트워크를 구성하기가 어렵고, 네트워크의 훈련은 주로 주어진 입력과 출력 데이터를 기본으로 하기 때문에 새로운 입력 데이터가 주어지면 원하는 정밀성을 보장할 수 없는 문제점들이 있다.

이러한 문제점을 개선하기 위하여 지금까지 여러 가지 방법들이 제안되고 있으며, 이는 크게 두 개의 부류(category) 즉 경험적(heuristic)기술과 수치 최적화기술로 구분되고 있다. 경험적 기술로는 학습율을 가변화하는 방법, 모멘텀법, 변수들의 재 스케일링, 다층 신경회로망의 구조 즉 은닉층의 층수 및 각 층의 뉴런 수를 조절하는 방법 등이 있다^[3-8]. 모멘텀법은 오차함수가 2차 함수로 충분히 근사할 수 있는 영역에서는 비선형 최적화수법을 이용하는 방법보다 수렴에 시간이 걸리는 단점을 가지고 있지만 알고리즘이 간단하고, 설정하는 파라미터가 적은 특징을 가지고 있어 많은 응용 분야에 이용되고 있다. 그러나 모멘텀법에 있어서도 적용하는 문제에 적절한 파라미터가 서로 다르므로 시행착오적인 파라미터의 설정이 필요하다. 학습율의 가변화를 이용하는 방법에는 2차 보간법 등의 직접 탐색을 넣는 방법, 학습율을 선형적으로 증가시키다가 지수적으로 감소시키는 Jacobs등에 의한 Delta Bar Delta법칙, 학습 반복수와 방향에 의해 학습율을 계산하는 방법, 역학 시스템을 이용하는 방법 등이 있다^{[3-4],[6],[8]}. 신경회로망의 구조를 조정하는 방법으로서 학습과정에서 연결강도가 일정한 값으로 향하는 뉴런들은 고정시키고 끝까지 학습시키지 않는 방법, 필요 없는 뉴런의 가지치기 방법, 학습하는 동안 사용하지 않는 연결강도를 제거하는 방법 등이 있다^{[5],[7]}. 그러나 이러한 방법들의 대부분은 많은 파라미터의 설정과 여분의 메모리가 요구되고, 모멘텀항과 학습율의 값을 결정하기 위한 특별한 계산이 필요하게 된다.

수치 최적화 기술로는 Conjugate gradient algorithm, Levenberg-Marquardt method, Inverse gradient method, 비선형 최적화법을 이용하는 방법, 변형된 오차함수를 이용하는 방법 및 확률적 방식을 이용하는 방법 등이 있다^[9-11]. 비선형 최적화 수법을 이용하는 경우에는 오차함수가 2차 함수로 충분히 근사할 수 있는 영역에서는 고속으로 수렴하고, 정밀도가 좋은 해가 얻어지는 것으로 알려지고 있다. 그러나 파라미터의 수가 많고, 알고리즘이 복잡하다는 문제점들이 있다. 이 기술의 문제점으로서 신경회로망에 직접 적용할 수 없다는 것과 많은 계산량과 기억장치를 요구하는 것이다.

본 논문에서는 역전파 알고리즘의 파라미터인 학습율, 연결강도 및 바이어스의 초기값, 활성화 함수의 이득을 고정된 일반적인 역전파(GBP: General BackPropagation) 알고리즘의 여러 가지 문제점들을 개선하기 위하여 활성화 함수(activation function)의 이득을 퍼지 로직 시스템을 이용하여 자동으로 조정하는 간단한 방식을 제안한다. 퍼지 로직 시스템을 구성하기 위하여 먼저 활성화 함수로서 일반적으로 많이 사용하는 시그모이드 함수의 이득 변화가 학습율, 연결강도 및 바이어스들의 변화와 등가인 관계를 조사한다. 퍼지 로직 시스템의 입력들은 마지막 층에 대한 오차의 감도와 은닉층에 대한 오차의 평균 감도를 사용하고, 출력은 활성화 함수인 시그모이드 함수의 이득을 사용한다. 제안한 방식의 유용성을 확인하기 위하여 N-패리티(parity)문제, 함수 근사화 및 패턴 인식등에 대하여, 활성화 함수로서 시그모이드 함수의 이득을 고정된 일반적인 역전파 알고리즘 방식과 시뮬레이션을 통해 성능을 비교 검토한다.

II. 활성화 함수의 이득과 학습율, 연결강도 및 바이어스 사이의 관계

역전파 알고리즘은 다층 신경회로망의 학습에 일반적으로 많이 사용되고 있는 알고리즘으로 연결강도, 바이어스, 학습율, 활성화 함수, 네트워크 구조 및 활성화 함수의 이득 등의 파라미터들을 가지고 있다. 많은 사람들이 이들의 관계에 대하여 조사하였으며, 특히 Zurada는 활성화 함수의 이득과 연결강도 사이의 관계에 대하여 활성화 함수의 이득이 크면 큰 학습율을 가진 결과와 유사하므로, 1의 표준 값으로 기울기를 유지하고 학습율을 가변하여 학습 속도를 제어하는 것이

바람직하다고 발표하였다^[12-13].

다층신경회로망에 뉴런의 활성화 함수로서는 여러 가지가 사용되고 있으며 일반적으로 다음과 같은 시그모이드 함수를 많이 사용한다.

$$f(x) = \frac{\gamma}{1 + e^{-\beta x}} \quad (2.1)$$

여기서 범위는 $(0, \gamma)$ 이며, hyperbolic tangent 인 $\gamma \tanh(\beta x)$ 는 $(-\gamma, \gamma)$ 의 범위를 갖고, 가우스 함수인 $\gamma e^{-(\beta x)^2}$ 는 $(0, \gamma)$ 의 범위를 갖는다. 이때 파라미터 β 를 활성화 함수의 이득이라고 하며, $\gamma\beta$ 를 기울기 (steepness)라고 한다. γ 가 1인 경우에는 이득과 기울기는 동일하다. 아래의 정리는 두 개의 신경회로망 A와 B의 구조 및 활성화 함수가 동일하다고 할 때, 활성화 함수, 이득, 학습율, 연결강도 및 바이어스 등의 관계를 나타낸다.

정리 1. 두 개의 신경회로망 A와 B가 활성화 함수 $f(x)$, 이득 β , 학습율 η , 연결강도 W 및 바이어스 b 를 갖고 동일한 구조일 때 표2.1과 같은 관계를 가질 때 역전파 알고리즘에서 서로 등가라고 한다.

표 2-1. 활성화 함수, 이득, 학습율, 연결강도 및 바이어스 사이의 관계

Table 2-1. The relationship between activation function, gain, learning rate, weight, and bias.

	Network A	Network B
Activation function	$f(x) = \hat{f}(\beta x)$	$\hat{f}(x)$
Gain	β	$\hat{\beta} = 1$
Learning rate	η	$\hat{\eta} = \beta^2 \eta$
Weights	W	$\hat{W} = \beta W$
Bias	b	$\hat{b} = \beta b$

즉 두 개의 신경회로망이 등가라면 이득이 β 만큼 변화하면 학습율은 $1/\beta^2$, 연결강도와 바이어스는 $1/\beta$ 로 변화된다.

이에 대한 증명을 하기 위하여 먼저 여기서 사용되는 기호나 심볼의 의미를 정의하고, 간단히 역전파 알고리즘에 대하여 설명한다^[2].

p 는 입력, n 는 네트(net) 입력으로 $n = Wp + b$, a 는 뉴런의 출력으로 $a = f(n)$, t 는 목표값, S 는 오차에 대

한 감도이며 영문자 위의 ^표시는 신경회로망 B의 파라미터들이다.

다층 신경회로망의 역전파 알고리즘은 LMS 알고리즘의 일반화이고, 알고리즘은 다음과 같은 평균 자승 오차(mean square error)를 최소화하기 위하여 네트워크 파라미터들을 조절한다.

$$F(\mathbf{x}) = E[e^T e] = E[(t - a)^T (t - a)] \quad (2.2)$$

여기서 \mathbf{x} 는 네트워크의 연결강도와 바이어스의 벡터이며, $E[\]$ 는 기대치를 나타낸다.

LMS 알고리즘을 이용하여 다음 식과 같이 평균 자승 오차를 근사화한다.

$$\begin{aligned} \bar{F}(\mathbf{x}) &= (t(k) - a(k))^T (t(k) - a(k)) \\ &= e^T(k) e(k) \end{aligned} \quad (2.3)$$

여기서 k 는 반복 수이며, 네트워크 파라미터인 연결강도와 바이어스는 최급강하법(steepest descent rule)을 이용하여 식(2.4), (2.5)과 같이 조절 한다.

$$W_i(k+1) = W_i(k) - \eta S_i(a_{i-1})^T \quad (2.4)$$

$$b_i(k+1) = b_i(k) - \eta S_i \quad (2.5)$$

이상과 같이 다층 신경회로망은 학습되며, 모든 뉴런은 자신의 학습율과 활성화 함수를 가지며, 일반적인 경우에 학습율과 활성화 함수는 전체 네트워크에서 동일하다.

정리 1의 증명은 역전파 알고리즘에서 전방향에 대한 전제(lemma)1과 역방향에 대한 전제 2로 나누어 한다.

Lemma 1. 표 2-1의 관계를 만족하는 동일한 뉴런들의 같은 활성화 함수를 가진 두 개의 네트워크 A와 B에서 $a_i = \hat{a}_i$ 이다.

증명 : 입력 층이 아닌 1층의 뉴런 j 에서

$$\begin{aligned} a_{1,j} &= \hat{a}_{1,j} \\ \Leftrightarrow f_{1,j}(n_{1,j}) &= \hat{f}_{1,j}(\hat{n}_{1,j}) \\ \Leftrightarrow \hat{f}_{1,j}(\beta_{1,j} n_{1,j}) &= \hat{f}_{1,j}(\hat{n}_{1,j}) \\ \Leftrightarrow \beta_{1,j} n_{1,j} &= \hat{n}_{1,j} \\ \Leftrightarrow \beta_{1,j} (w_{1,j} \cdot a_{i-1} + b_{1,j}) &= \hat{w}_{1,j} \cdot \hat{a}_{i-1} + \hat{b}_{1,j} \\ \Leftrightarrow \beta_{1,j} (w_{1,j} \cdot a_{i-1} + b_{1,j}) &= (\beta_{1,j} w_{1,j}) \cdot \hat{a}_{i-1} + \beta b_{1,j} \end{aligned}$$

여기서 낮은 층의 활성화 값들이 동일함을 알 수 있

으며, 또한 $\beta_{l,j} n_{l,j} = \hat{n}_{l,j}$ 임을 알 수 있다.

Lemma 1의 증명에서 $\hat{w}_{l,j} = \beta_{l,j} w_{l,j}$, $\hat{b}_l = \beta b_l$ 이라는 특성이 사용되었는데 역방향 전송에서는 연결 강도가 변화하므로, 이 특성이 역방향 전송 단계에서도 불변이라는 것을 보여 주어야 한다.

Lemma 2. $\hat{a}_{l,j} = a_{l,j}$ 와 $\hat{n}_{l,j} = \beta_{l,j} w_{l,j} + \beta b_{l,j}$ 의 특성을 갖는 두 개의 네트워크 A 와 B에서 $\forall j, l: \hat{w}_{l,j} = \beta_{l,j} w_{l,j}$, $\hat{b}_{l,j} = \beta b_{l,j}$ 의 특성이 역방향 전송에서도 불변이다.

증명 : 만약에 $\Delta \hat{w}_{l,j} = \beta_{l,j} \Delta w_{l,j}$, $\Delta \hat{b}_{l,j} = \beta_{l,j} \Delta b_{l,j}$ 이라면 위의 식이 증명된다.

$$\begin{aligned} \Delta \hat{w}_{l,j} &= \beta_{l,j} \Delta w_{l,j} \\ \Leftrightarrow \hat{\eta}_{l,j} \hat{S}_{l,j} \hat{a}_{l-1} &= \beta_{l,j} \eta_{l,j} S_{l,j} a_{l-1} \\ \Leftrightarrow \beta_{l,j}^2 \eta_{l,j} \hat{S}_{l,j} &= \beta_{l,j} \eta_{l,j} S_{l,j} \\ \Delta \hat{b}_{l,j} &= \beta_{l,j} \Delta b_{l,j} \\ \Leftrightarrow \hat{\eta}_{l,j} \hat{S}_{l,j} &= \beta_{l,j} \eta_{l,j} S_{l,j} \\ \Leftrightarrow \beta_{l,j}^2 \eta_{l,j} \hat{S}_{l,j} &= \beta_{l,j} \eta_{l,j} S_{l,j} \end{aligned}$$

이 성립하기 위해서는 $\beta_{l,j} \hat{S}_{l,j} = S_{l,j}$ 이 성립하여야 하므로 이에 대한 증명은 출력층과 출력층이 아닌 층에 대하여 한다.

$$\begin{aligned} \beta_{L,j} \hat{S}_{L,j} &= S_{L,j} \\ \Leftrightarrow \beta_{L,j} \hat{f}'_{L,j}(\hat{net}_{L,j})(t_j - \hat{a}_{L,j}) &= f'_{L,j}(net_{L,j})(t_j - a_{L,j}) \\ \Leftrightarrow \beta_{L,j} \hat{f}'_{L,j}(\hat{net}_{L,j}) &= f'_{L,j}(net_{L,j}) \\ \Leftrightarrow \beta_{L,j} \hat{f}'_{L,j} \beta_{L,j} net_{L,j} &= f'_{L,j}(net_{L,j}) \end{aligned}$$

여기서 L은 출력층이다.

출력층이 아닌 층의 뉴런 j 에 대하여서는

$$\begin{aligned} \beta_{l,j} \hat{S}_{l,j} &= S_{l,j} \\ \Leftrightarrow \beta_{l,j} \hat{f}'_{l,j}(\hat{net}_{l,j}) \sum_k \hat{S}_{l+1,k} \hat{w}_{l+1,j,k} &= f'_{l,j}(net_{l,j}) \sum_k S_{l+1,k} w_{l+1,j,k} \\ \Leftrightarrow \beta_{l,j} f'_{l,j}(\beta_{l,j} net_{l,j}) \sum_k \hat{S}_{l+1,k} \hat{w}_{l+1,j,k} &= f'_{l,j}(net_{l,j}) \sum_k S_{l+1,k} w_{l+1,j,k} \\ \Leftrightarrow \sum_k \hat{S}_{l+1,k} \hat{w}_{l+1,j,k} &= \sum_k S_{l+1,k} w_{l+1,j,k} \\ \Leftrightarrow \sum_k \hat{S}_{l+1,k} \beta_{l+1,k} w_{l+1,j,k} &= \sum_k S_{l+1,k} w_{l+1,j,k} \end{aligned}$$

(361)

이상과 같이 정리 1의 증명에 의해 역전파 알고리즘의 여러 파라미터 중에서 활성화 함수의 이득을 변화시키면 학습율, 연결강도 및 바이어스도 변화함을 알 수 있다. 학습율, 연결강도 및 바이어스의 변화는 역전파 알고리즘의 영향을 미친다는 것은 아는 사실이므로, 활성화 함수의 이득을 가변화하는 방법은 역전파 알고리즘에 큰 영향을 미친다고 본다.

III. 퍼지 논리를 이용한 시그모이드 함수의 이득 자동 조절 방식

3-1. 오차 함수의 감도와 시그모이드 함수의 이득과의 관계

역전파 알고리즘은 평균 자승 오차를 최소화하기 위하여 네트워크 파라미터들을 최급강하법을 이용하여 조절하는 알고리즘이다. 즉 앞장의 식(2.3)의 평균자승 오차를 최소화 하기 위하여 오차역전파에 의하여 식(2.4), (2.5)와 같이 연결강도 및 바이어스를 최급강하법으로 갱신한다. 이때 연결강도 W의 변화량 및 바이어스 b의 변화량들은 오차에 대한 감도 S에 비례하므로, 감도가 큰 영역에서는 연결강도 및 바이어스의 변화량도 크고, 감도가 작은 영역에서는 변화량도 작게 된다. 또한 오차에 대한 감도 S는 활성화 함수의 미분 및 오차에 비례한다.

식(2.1)의 시그모이드 함수를 예로 들면 연결강도 및 바이어스의 변화량은

$$\Delta W_l = -\eta S_l (a_{l-1})^T \tag{3.1}$$

$$\Delta b_l = -\eta S_l \tag{3.2}$$

이고, 이때 감도 S_l 은

$$S_l = \frac{\partial \bar{F}}{\partial n_l} = F'_l(n_l)(W_{l+1})S_{l+1} \tag{3.3}$$

이며, 시그모이드 함수의 미분은

$$f(x)' = \frac{\beta x e^{-\beta x}}{(1 + e^{-\beta x})^2} \tag{3.4}$$

이므로, 감도 S_l 은 식(3.4)에 의해 이득 β 에 영향을 받는 것을 알 수 있다. 즉 뉴런의 활성화 함수인 시그모이드 함수는 이득 β 에 의해 특성이 결정되고, 식(3.4)의 시그모이드 함수의 미분값은 오차의 감도 신호를 만드는

구성 요소이므로 이 두 요소는 연결강도 및 바이어스 변화에 영향을 미친다.

따라서 역전파 알고리즘의 탐색과정에서는 오차함수의 감도에 크게 의존하고, 감도는 활성화 함수의 이득에도 많은 영향을 받으므로 성능개선을 하기 위하여서는 활성화 함수의 이득을 고정하는 것 보다는 가변하는 것이 바람직하다고 본다. 이에 대한 영향 분석을 하기 위하여 다음 장에서의 패리티문제, 함수 근사화 및 문자 인식에 대하여 오차함수의 감도와 시그모이드 함수의 이득에 대하여 시뮬레이션한다. 오차함수의 감도와 오차 함수는 비례하여 유사한 영향을 나타내므로 시그모이드 함수의 이득에 변화에 대한 오차함수의 변화에 대한 개략도는 그림 3-1과 같다.

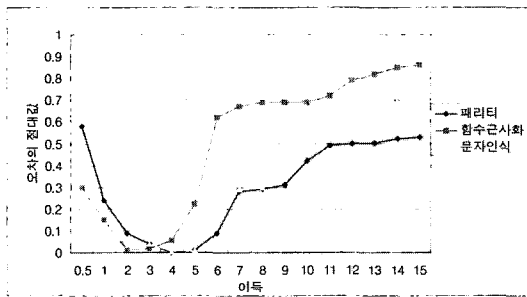


그림 3-1. 오차 대 이득에 대한 그래프
Fig. 3-1. The graph of error vs gain.

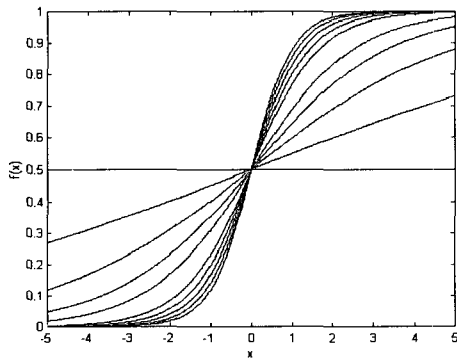


그림 3-2. 활성화 함수의 이득 변화
Fig. 3-2. The change of the gain of activation function.

그림 3-1에서 이득에 변화에 대하여 오차는 감소하다 증가하는 것을 알 수 있으며, 세가지 경우가 유사한 경향이 있음을 알 수 있으며 패리티 문제와 문자 인식은 더욱 유사함을 알 수 있다. 즉 시그모이드 함수의 이득 자동 조절은 역전파 알고리즘의 성능을 개선시킬

수 있다.

그림 3-2는 시그모이드 함수의 이득 β 의 변화에 따른 시그모이드 함수의 형상으로 이득 β 가 무한대로 발산할 때, 연속적인 함수는 이산 함수의 모양에 가까워지며 이득 β 가 감소하면 선형화 된다.

이득 β 의 변화에 따른 미분치의 결과는 그림 3-3과 같다.

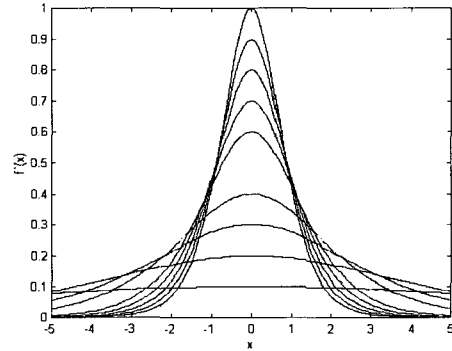


그림 3-3. 미분한 활성화 함수의 이득 변화
Fig. 3-3. The change of the gain of the differential activation function.

그림에서 보는 바와 같이, 미분치는 이득 β 값이 일정할 때 중간 대역의 값이 가장 크고 입력값이 일정할 때 이득 β 값이 클수록 커진다.

3-2. 시그모이드 함수의 이득 자동 조절 방식

다층 신경회로망의 역전파 알고리즘에는 연결강도, 바이어스, 학습율, 활성화 함수, 네트워크 구조 및 활성화 함수의 이득 등의 여러 파라미터들을 가지고 있으나, 활성화 함수의 이득의 변화는 연결강도, 학습율 및 바이어스 변화와 등가이다. 그러므로 이득의 가변화는 다층 신경회로망의 성능개선에 큰 효과를 가질 수 있다. 본 논문에서는 일반적인 역전파 알고리즘의 여러 가지 문제점들을 개선하기 위하여 활성화 함수의 이득을 퍼지 로직 시스템을 이용하여 자동으로 조절하는 방식을 제안한다. 제안하는 퍼지 논리 시스템에 의한 다층 신경회로망의 활성화 함수의 이득을 자동 조절하는 시스템(FGBP : Fuzzy logic system for automatic tuning of the Gain of the activation function in the BackPropagation)의 블록 선도는 그림 3.4 와 같다.

그림에서 퍼지 논리 시스템의 입력인 S1은 출력층의 감도이며, S2는 은닉층에 대한 평균감도이다. 일반적인 퍼지 시스템은 그 시스템에 대한 전문가나 숙련자들의

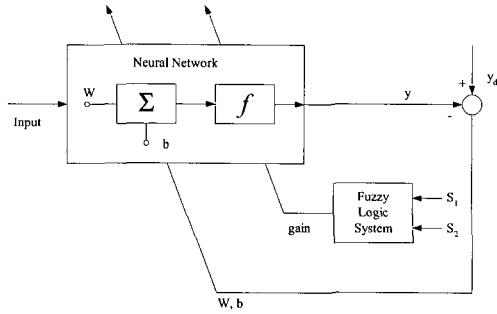


그림 3-4. 제안하는 FGBP 방식의 블록선도
Fig. 3-4. The block diagram of the proposed FGBP method.

지식을 근거로 규칙과 소속함수를 만드나 여기서는 GBP의 패리티 문제, 함수 근사화 및 패턴 인식 문제 등의 시뮬레이션을 통한 데이터를 이용하여 규칙과 소속 함수들을 만든다. 여기서 규칙과 소속함수들은 세가지 문제에 대하여 각각 구성하면 더욱 성능개선을 할 수 있으나, 세가지 문제들의 오차특성이 그림 3-1과 같이 유사하여 일반성이 있게 동일하게 만들어 적용한다.

시그모이드 함수의 이득과 오차에 대한 감도 특성을 여러 가지 시뮬레이션 상에서 검토한 결과 S1은 [-3, 3], S2는 [-2, 2]에 편중되어 있고, S1에 비해 S2의 변화폭은 크지 않았다. 따라서 소속함수의 구성은 사용밀도가 높은 지점을 중심으로 S1의 소속 함수는 그림 3.5 와 같이 삼각형 함수 3개와 사다리꼴 함수 2개로 구성 하였고, 상대적으로 변화가 적은 S2는 그림 3.6과 같이 가장 밀도가 넓은 지점에 삼각형 함수 하나, 그 외의 지점에 사다리꼴 함수 2개로 구성하였다. 입력들은 이 소속함수에 의하여 퍼지화된다.

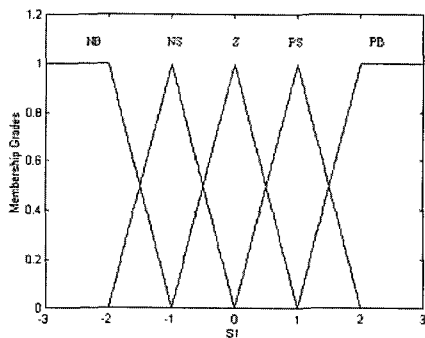


그림 3-5. S1의 소속함수
Fig. 3-5. The membership functions of S1.

퍼지 제어 규칙은 전문가의 경험을 바탕으로 만들어

지나, 여기서는 은닉층을 하나 사용한 경우의 패리티 문제, 함수 근사화 및 패턴 인식 문제 등의 시뮬레이션을 바탕으로 활성화 함수의 이득에 따른 출력에 관계되는 감도와 은닉층에 관계되는 감도에 대한 관계를 표 3-1과 같이 정한다.

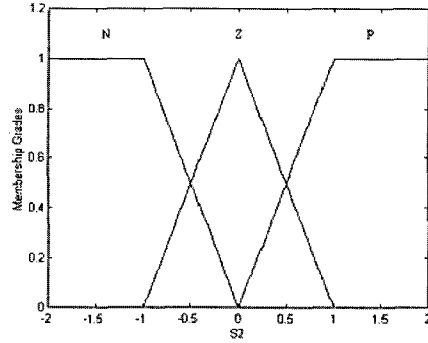


그림 3-6. S2의 소속함수
Fig. 3-6. The membership functions of S2.

표 3-1. 퍼지 제어 규칙
Table 3-1. Fuzzy Control rule.

S1 \ S2	NB	NS	Z	PS	PB
N	L	L	M	M	H
Z	L	M	M	M	H
P	L	M	M	H	H

여기서 S1, S2는 퍼지 논리 시스템의 입력으로 각각 출력층에 관계되는 감도와 은닉층에 관계되는 감도이며, NB(Negative Big), NS(Negative Small), Z(Zero), PS(Positive Small), PB(Positive Big), N(Negative),

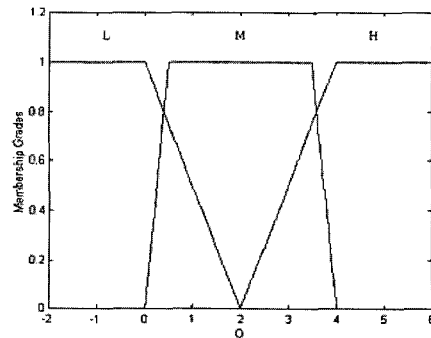


그림 3-7. 출력의 소속 함수
Fig. 3-7. The membership functions of output.

P(Positive)는 S1과 S2의 퍼지 소속 함수이고, 퍼지 제어 규칙 내부의 L(Low), M(Medium), H(High)는 출력 신호인 시그모이드 함수의 이득에 대한 퍼지 소속 함수이다.

출력의 소속 함수는 그림 3-7과 같이 함수 근사, parity 문제, 패턴 인식등에 오차에 감도에 대한 이득의 변화 특성이 비슷하므로 3가지 문제를 일반화 하기 위하여 사다리꼴 함수 3개로 구성한다.

표 3-1의 15개의 퍼지 제어 규칙을 if~then 형식의 퍼지관계로 표현한 것이 식(3.5)이다.

- $R^{(1)} : \text{If } S1 \text{ is NB, } S2 \text{ is N then } u \text{ is L}$
- $R^{(2)} : \text{If } S1 \text{ is NB, } S2 \text{ is Z then } u \text{ is L}$
- $R^{(3)} : \text{If } S1 \text{ is NB, } S2 \text{ is P then } u \text{ is L}$
- $R^{(4)} : \text{If } S1 \text{ is NS, } S2 \text{ is N then } u \text{ is L}$
- $R^{(5)} : \text{If } S1 \text{ is NS, } S2 \text{ is Z then } u \text{ is M}$ (3.5)
- $R^{(6)} : \text{If } S1 \text{ is NS, } S2 \text{ is P then } u \text{ is M}$
- $R^{(7)} : \text{If } S1 \text{ is Z, } S2 \text{ is N then } u \text{ is M}$
- $R^{(8)} : \text{If } S1 \text{ is Z, } S2 \text{ is Z then } u \text{ is M}$
- $R^{(9)} : \text{If } S1 \text{ is Z, } S2 \text{ is P then } u \text{ is M}$
- $R^{(10)} : \text{If } S1 \text{ is PS, } S2 \text{ is N then } u \text{ is M}$
- $R^{(11)} : \text{If } S1 \text{ is PS, } S2 \text{ is Z then } u \text{ is M}$
- $R^{(12)} : \text{If } S1 \text{ is PS, } S2 \text{ is P then } u \text{ is H}$
- $R^{(13)} : \text{If } S1 \text{ is PB, } S2 \text{ is N then } u \text{ is H}$
- $R^{(14)} : \text{If } S1 \text{ is PB, } S2 \text{ is Z then } u \text{ is H}$
- $R^{(15)} : \text{If } S1 \text{ is PB, } S2 \text{ is P then } u \text{ is H}$

식 (3.5)에서 15개의 퍼지관계 $R^{(i)}, i=1\sim 15$ 에서 S1 과 S2는 퍼지 논리 시스템의 입력변수이고, u는 시그모이드 함수의 이득을 나타내는 출력변수이다. 식 (3.5)의 퍼지 제어규칙을 이용하여 퍼지 추론을 하는 방법으로는 최소-최대 연산, 적-대수합 연산 등이 있으나 여기서는 임의로 최소-최대 연산을 사용한다. 퍼지추론의 결과인 퍼지값을 단일 실수값으로 변환시키는 비퍼지화(defuzifier)는 최대값방법(maximum criterion method), 최대평균값(mean of maximum method), 무게중심법(center of area method) 등이 있으며, 여기서도 일반적으로 많이 사용하는 무게중심법을 사용한다^[14].

IV. 시뮬레이션

4-1. N-패리티문제

4-1-1. N=2(XOR)

먼저 교사(supervised) 학습 알고리즘 테스트에 벤치마크(bench mark)로 잘 알려진 N=2인 XOR 문제를 한다. 학습에는 입력층의 뉴런수는 2, 은닉층의 뉴런수는 m, 출력층의 뉴런수는 1이며, 활성화함수는 식 (2.1)의 $\gamma = 1$ 인 시그모이드 함수를 사용한 신경회로망을 이용한다. 훈련 패턴은 (0,0), (1,0), (0,1), (1,1)의 4패턴이고, 훈련 패턴의 1의 개수의 총합이 홀수인 경우에는 목표신호를 1, 짝수의 경우에는 0로 한다. 수렴판정은 모든 패턴에 대하여 출력과 목표신호의 차의 절대치가 0.01 이하로 하고 최대 iteration은 3000회 한다.

결과는 수렴율(C : Convergence rate), 평균학습 iteration(A : Average learning iteration), 가장 작은 학습iteration(L : Lowest learning iteration) 및 최대 학습 iteration(U : Maximum learning iteration)으로 한다.

역전파 알고리즘의 파라미터 값들은 시그모이드 함수의 이득은 1로 놓고 시행 착오에 의한 시뮬레이션에서 가장 우수한 결과를 나타낸 값으로 정하였다. 초기 학습율은 0.5, 초기 연결강도와 바이어스는 (-1,1)/10 에서 랜덤한 값을 사용하였으며, 은닉층의 뉴런 수 m을 2, 5, 10, 15로 변화시켰을 때 시그모이드 함수의 이득을 1로 고정한 GBP 알고리즘을 이용한 방식과 제안한 FGBP 알고리즘을 이용한 방식의 시뮬레이션 결과는 표 4-1과 같다.

표 4-1. N=2 인 경우의 시뮬레이션 결과

Table 4-1. Simulation results of N=2.

	m=2	m=5	m=10	m=15
	C A L U	C A L U	C A L U	C A L U
GBP	0.66 1184 11 2600	0.78 997 12 2400	0.90 745 11 1520	0.82 1010 12 2455
FGBP	1.0 146 4 320	1.0 139 4 302	1.0 96 4 180	1.0 141 4 305

표 4-1의 결과에서 제안한 FGBP 방식의 수렴율은 100%이며, 평균 학습 iteration도 GBP 방식에 비해 약 87% 정도 감소함을 알 수 있으며, 뉴런수가 10개인 경우의 평균 학습 iteration이 가장 적으므로 주어진 파라미터 값에서는 최적의 구조로 본다.

4-1-2. N=5

학습에는 입력층의 뉴런수는 5, 은닉층의 뉴런수는 m, 출력층의 뉴런수는 1이며, 활성화 함수는 $\gamma = 1$ 인

시그모이드 함수를 사용한 신경회로망을 이용한다. 학습 패턴은 32개이고, 입력 패턴의 1의 개수의 총합이 홀수인 경우에는 목표신호를 1, 짝수의 경우에는 0으로 한다. 수렴판정은 모든 패턴에 대하여 출력과 목표신호의 차의 절대치가 0.01 이하로 하고, 최대 iteration을 3000회 하여 위의 경우와 같이 수렴율(C : Convergence rate), 평균학습 iteration(A : Average learning iteration), 가장 작은 학습iteration(L : Lowest learning iteration) 및 최대 학습 iteration(U : Maximum learning iteration)을 구한다

여기서도 시그모이드 함수의 이득을 1로 놓고, 시행착오적으로 우수한 값인 초기 학습율 0.5, 초기 연결강도와 바이어스는 (-1,1)/10에서 랜덤한 값을 사용하고, 은닉층의 뉴런 수 m을 5, 10, 15, 20으로 변화시켰을 때, 활성화 함수의 이득을 1로 고정한 GBP 방식과 제안한 FGBP 방식의 시뮬레이션 결과는 표 4-2와 같다.

표 4-2. N=5 인 경우의 시뮬레이션 결과

Table 4-2. Simulation results of N=5.

	m=5		m=10		m=15		m=20	
	C	A L U	C	A L U	C	A L U	C	A L U
GBP	0.75	733 11 1800	0.83	588 10 1650	0.92	538 8 1450	0.85	575 10 1623
FGBP	1.0	71 3 195	1.0	52 3 190	1.0	44 4 130	1.0	55 4 192

표 4-2 에서도 제안한 FGBP 방식의 수렴율은 100%이며, 평균 학습 iteration도 약 91% 정도 감소됨을 알 수 있으며, 이 경우에는 뉴런 수가 15개인 경우의 평균 학습 iteration이 가장 적으므로 주어진 파라미터 값에서는 최적의 구성으로 본다. 이상과 같이 N-Parity 문제에서는 N=2 나 N=5 에서 유사한 결과가 나오므로 N에 수의 관계없이 제안한 방식의 성능이 개선됨을 확인하였다.

4-2 함수 근사화 문제

함수 근사화 문제를 위한 신경회로망의 구성은 입력층의 뉴런수는 5, 은닉층의 뉴런수는 10, 출력층의 뉴런수는 1이며, 활성화함수는 $\gamma = 1$ 인 시그모이드 함수를 사용한다. 목표 함수는 식(4.1)과 같다.

$$d(t) = \sin\left(\frac{2\pi t}{250}\right) \quad (4.1)$$

역전파 알고리즘의 파라미터 값은 시그모이드 함수의

이득을 1로 놓고 시행착오적으로 우수한 값인 초기 학습율 0.9, 초기 연결강도와 바이어스는 (-1,1)/10에서 랜덤한 값을 사용한다. 시그모이드 함수의 이득을 1로 고정한 GBP 방식의 시뮬레이션 결과는 그림 4-1과 같다.

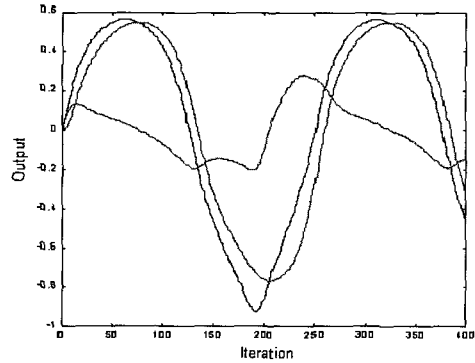


그림 4-1. GBP의 응답 및 오차 곡선
Fig. 4-1. The response and error curve of the GBP.

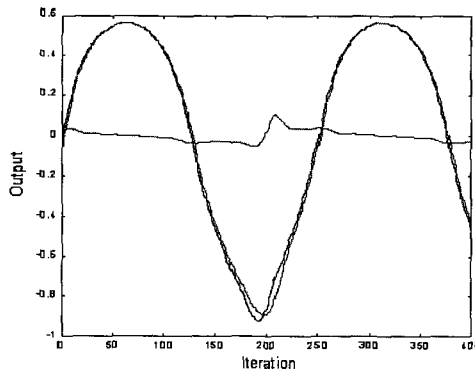


그림 4-2. FGBP의 응답 및 오차 곡선
Fig. 4-2. The response and error curve of FGBP.

그림 4-2는 제안한 FGBP 방식의 시뮬레이션 결과이고, 그림 4-3은 퍼지 출력으로 활성화 함수의 이득의 변화를 나타낸다.

그림과 같이 제안한 방식의 오차가 현저하게 개선됨을 알 수 있다. 즉 제안한 방식은 함수 근사화에서 정밀도가 현저하게 개선됨을 알았다.

GBP는 학습율, 연결강도 및 바이어스의 초기값에 크게 영향을 받고, 제안한 방식은 활성화 함수의 이득의 변화가 학습율, 연결강도 및 바이어스에 관계가 있으므로 초기값에 영향을 덜 받는 것을 알아보기 위하여 학습율과 초기 연결강도를 변화시켜 본다. 그림 4-4와 4-5는 학습율을 0.1로 변화한 경우에 GBP와 제안한 FGBP의 시뮬레이션 결과이다.

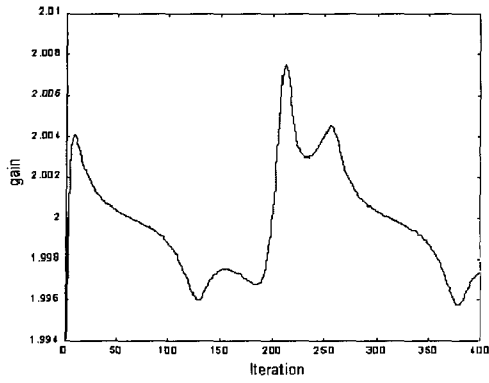


그림 4-3. 활성화 함수의 이득 변화 곡선
Fig. 4-3. The gain change curve of the active function.

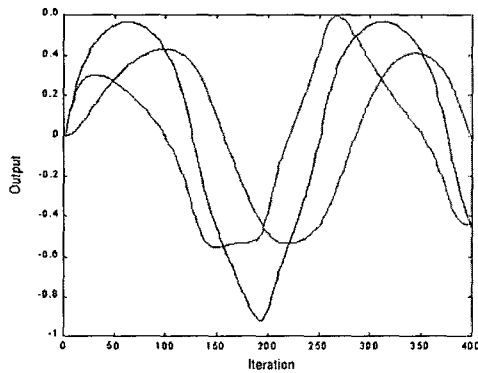


그림 4-4. 학습율을 0.1로 한 GBP의 응답 및 오차 곡선
Fig. 4-4. The response and error curve of the GBP with learning rate 0.1.

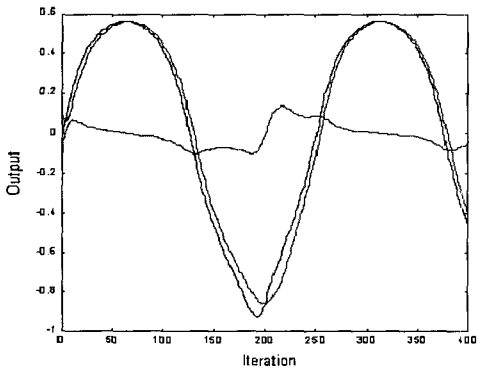


그림 4-5. 학습율을 0.1로 한 FGBP의 응답 및 오차 곡선
Fig. 4-5. The response and error curve of the FGBP with learning rate 0.1.

그림에서와 같이 GBP는 초기 학습율에 크게 영향을 받는 것을 알 수 있으나 제안한 FGBP방식은 영향을

덜 받는 것을 알 수 있다.

그림 4-6과 4-7은 초기 연결강도를 (-1,1)/100으로 변화 시켰을 때의 GBP와 제안한 FGBP의 시뮬레이션 결과이다.

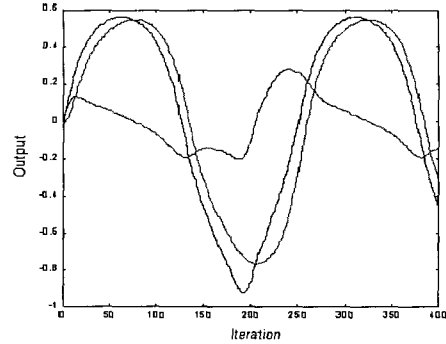


그림 4-6. 초기 연결강도가 (-1,1)/100로 랜덤한 경우의 GBP의 응답 및 오차 곡선
Fig. 4-6. The response and error curve of the GBP with the random initial weight value (-1,1)/100.

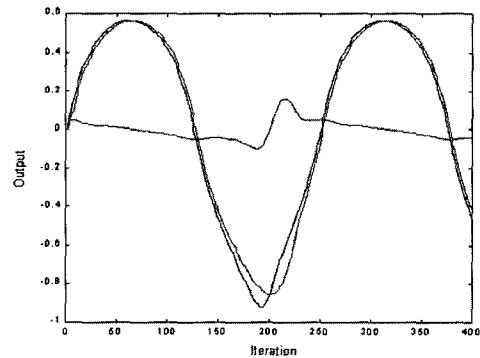


그림 4-7. 초기 연결강도가 (-1,1)/100로 랜덤한 경우의 FGBP의 응답 및 오차 곡선
Fig. 4-7. The response and error curve of the FGBP with the random initial weight value (-1,1)/100.

그림에서와 같이 GBP는 초기 연결 강도에 영향을 많이 받으나 제안한 방식은 영향을 덜 받는 것을 알 수 있다.

또한 GBP에서 새로운 입력 신호가 주어질 때 원하는 오차 범위 안에서의 출력신호를 보장할 수 없는 문제에 대하여 조사한다. 새로운 입력 신호로서 식(4.2)와 같이 주어진다.

$$d(t) = \sin\left(\frac{2\pi t}{250}\right) + 0.5 \cos(4\pi t) \quad (4.2)$$

그림 4-8은 GBP의 응답곡선 및 오차 곡선이고, 그림 4-9는 FGBP의 응답곡선 및 오차 곡선이다.

그림 4-8과 그림 4-9에서 GBP 알고리즘은 새로운 입력에 원하는 오차 범위의 응답을 내고 있지 않으나 제안한 FGBP 방식은 원하는 응답을 내고 있음을 알 수 있다. 이는 새로운 입력에도 원하는 오차 범위의 출력을 보장할 수 있다고 본다.

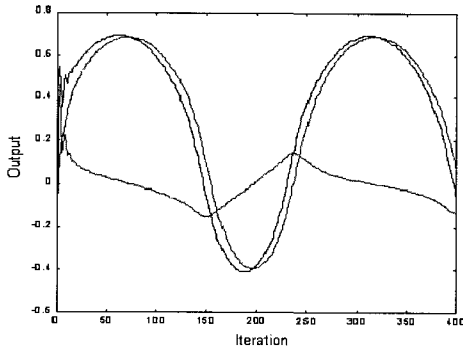


그림 4-8. 새로운 입력의 GBP의 응답 및 오차 곡선
Fig. 4-8. The response and error curve of GBP of the new input.

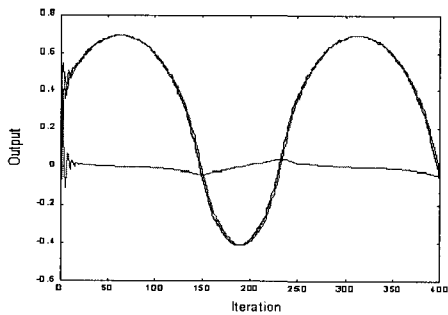


그림 4-9. 새로운 입력의 FGBP의 응답 및 오차 곡선
Fig. 4-9. The response and error curve of FGBP of the new input.

4-3. 패턴 인식 문제

패턴 인식 문제는 그림 4-10에 나타난 아라비아 숫자 0에서 9까지의 패턴들을 학습 신호로 사용한다. 구성된 신경회로망은 입력층의 뉴런 수는 100, 은닉층의 뉴런수는 m, 출력층의 뉴런 수는 4로 하고 활성화 함수는 위와 동일한 시그모이드 함수를 사용하고, 출력의 목표 신호들은 표 4.3과 같다.

수렴판정은 모든 패턴에 대하여 출력과 목표신호의 차의 절대치가 0.001 이하로 하고, 최대 iteration을 500회 하여 평균학습 iteration(A : Average learning

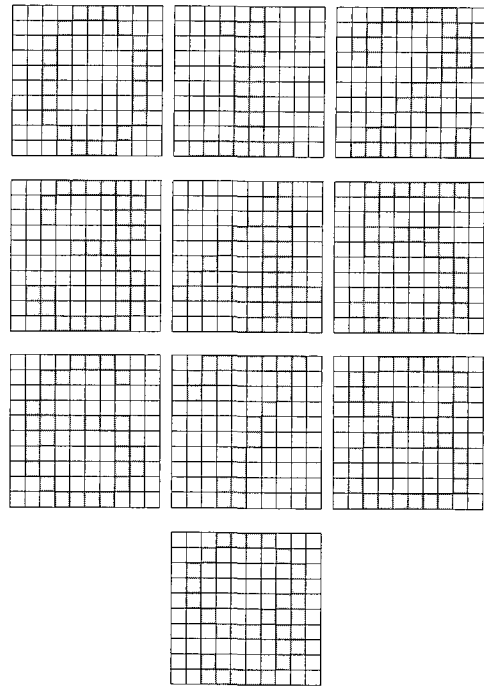


그림 4-10. 학습 패턴을 위한 아라비아 숫자(10x10)
Fig. 4-10. Arabic numerals(10x10) as learning patterns.

표 4-3. 목표 신호들의 출력

Table 4-3. Output of target signals.

숫자	0	1	2	3	4	5	6	7	8	9
목표 신호	0	0	0	0	0	1	1	1	1	1
	0	0	1	1	1	0	0	1	1	1
	0	1	0	1	1	1	1	0	1	1
	1	0	0	0	1	0	1	0	0	1

iteration) 구한다.

역전파 알고리즘의 파라미터 값은 시그모이드 함수의 이득을 1로 놓고 시행착오적으로 우수한 값인 초기 학습율 0.5, 초기 연결강도와 바이어스는 (-1,1)/10의 랜덤한 값을 사용하고, 은닉층의 뉴런 수 m을 25, 50, 100, 150 으로 변화시켰을 때 시그모이드 함수의 이득을 1로 고정한 GBP와 제안한 FGBP방식의 시뮬레이션 결과는 표 4-4와 같다.

표에서 F는 성공하지 못한 경우이다. 기존의 방식은 수렴을 못하는 경우가 있으나 연결강도와 바이어스의 초기값을 조절하거나 학습율을 조절하면 수렴할 수 있

다고 본다. 그러나 제안한 방식은 평균 약 285 학습 iteration으로 100% 성공함을 알 수 있으며 뉴런수가 100개인 경우의 평균 학습 iteration이 가장 적으므로 주어진 파라미터 값에서는 최적의 구조이다.

표 4-4. 아라비아 숫자들의 시뮬레이션 결과
Table 4-4. The comparison results of learning iterations of arabic numerals.

숫자	m=25		m=50		m=100		m=150	
	GBP	FGBP	GBP	FGBP	GBP	FGBP	GBP	FGBP
0	3052	321	2916	318	2615	302	3010	320
1	3006	329	2808	321	2483	279	2902	319
2	2909	312	2503	287	2174	266	2806	303
3	2558	292	2116	261	2108	257	2452	295
4	F	319	2776	287	2655	269	F	298
5	2555	301	2501	289	2438	277	2522	290
6	F	314	2788	284	2721	268	2922	304
7	2995	318	2245	304	2098	285	2665	312
8	F	327	F	309	2876	278	F	308
9	2304	289	1518	266	1191	251	2072	271

V. 결 론

본 논문에서는 역전파 알고리즘의 파라미터인 학습율, 연결강도 및 바이어스의 초기값, 활성화 함수의 이득을 고정한 일반적인 역전파 알고리즘의 여러 가지 문제점들을 개선하기 위하여 활성화 함수(activation function)의 이득을 퍼지 논리 시스템으로 자동 조정하는 간단한 방식을 제안하였다. 퍼지 논리 시스템을 구성하기 위하여 먼저 활성화 함수로서 시그모이드 함수 이득의 변화가 학습율, 연결강도 및 바이어스의 변화와 등가인 관계를 조사하였다. 퍼지 논리 시스템의 입력은 일반적인 역전파 알고리즘의 오차신호의 감도로서 출력층과 관계되는 감도와 은닉층과 관계되는 감도를 사용하고, 퍼지 논리 시스템의 출력은 시그모이드 함수의 이득을 사용하였다. 제안한 방식의 유용성을 확인하기 위하여 활성화 함수로서 시그모이드 함수를 사용하여, N-패리티(parity)문제, 함수 근사화 및 패턴 인식 문제 등에 대하여 적용하여, 기존의 시그모이드 함수의 이득을 고정한 일반적인 역전파 알고리즘 방식과 시뮬레이션을 통해 성능을 비교 검토한 결과는 다음과 같다.

- 一. 퍼지 규칙의 출력 파라미터는 시그모이드 함수의 이득 하나이고 퍼지규칙도 15개로 간단하다.
- 一. 시그모이드 함수의 이득과 학습율, 연결강도 및 바이어스가 등가 관계가 있어 이득 하나의 가변화로 다른 파라미터의 가변화 효과도 얻을 수 있다. 즉 이득이 변화하면 학습율, 연결강도 및 바이어스도 변화한다.
- 一. 수렴율이 N-Parity, 패턴 인식 등에서 100%로 현저히 개선됨을 알 수 있다.
- 一. 평균 학습 iteration이 N-Parity, 함수 근사화, 패턴 인식 등에서 크게 감소됨을 알 수 있다.
- 一. 함수 근사화에서 정밀도가 현저히 개선됨을 알 수 있다.
- 一. 새로운 입력에도 원하는 오차 범위의 출력을 얻을 수 있다.

참 고 문 헌

- [1] Peiman G. Maghami and Dean W.Sparks, "Design of Neural Networks for Fast Convergence and Accuracy : Dynamics and Control", IEEE Transactions on Neural Networks, Vol.11, No.11, pp.113~123, January 2000.
- [2] Martin T. Hagan, Howard B. Demuth, Mark Beale, Neural Network Design, PWS Publishing, Boston, 1995.
- [3] Tokumitsu Fujita, Takao Watanebe and Keiichiro Yasuda, "A Study on Improvement in Learning Efficiency of Multilayered Neural Networks based on Dynamical System", T.IEE Japan, Vol.117-C, No.12, pp.1848~1855, 1997.
- [4] A. Sperduti and A.Starita, "Speed up Learning and Network Optimization with Extended Back-propagation", Neural Networks, Vol.6, pp. 365~383, 1993.
- [5] B.Hassibi, D.G.Stork and G.H.Wolff, "Optimal brain surgeon and general network purging", Proceedings of the IEEE International Joint Conference on Neural Networks, vol.2, pp.441~444, 1992.
- [6] A.K.Rigler, J.M. Irvine and T.P. Vogl,

“Rescaling of variables in backpropagation learning”, *Neural Networks*, vol.3, no. 5, pp. 561~573, 1990.

[7] T.H.Cho, R.W.Conners and P.A. Araman, “Fast Back-propagation Learning using Steep Activation Functions and Automatic Weight Reinitialization”, *Proc. IEEE Int. Conf. Systems, Man, Cybernetics : Decision Aiding for Complex Syst.* 3, pp. 1587~1592, 1991.

[8] R. A. Jacobs, “Increased rates of convergence through learning rate adaptation,” *Neural Networks*, vol.1, no.4, pp. 295~308, 1988.

[9] M. T. Hagan and M. Menhaj, “Training feedforward networks with the Marquardt algorithm,” *IEEE Tran. on Neural Networks*, vol.5, no.6, 1994.

[10] C. Charalambous, “Conjugate gradient algorithm for efficient training of artificial neural networks,” *IEE Proceedings*, vol.139, no.3, pp. 301~310, 1992.

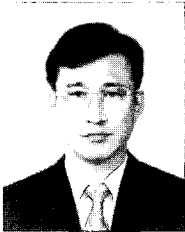
[11] Sanossian, H. Y. Y. and Evans, D. J., “Gradient Range-Based Heuristic Method for Accelerating Neural Network Convergence,” *Integrated Computer-Aided Engineering*, Vol.2(2), pp.147~152, 1995.

[12] C. Codrington and M. Tenorio, “Adaptive gain networks”, *Proc. IEEE, Int. Conf. Neural Networks*, ICNN94, 1, pp. 339~344, 1994.

[13] J. M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, St. Paul, MN, 1992.

[14] W. Pedrycz, *Fuzzy control and Fuzzy systems*, John Wiley& Sons, Inc., 1992.

저 자 소 개



鄭盛夫(正會員)

1979년 2월 : 동국대학교 전자공학과 졸업(공학학사). 1981년 2월 : 동국대학교 전자공학과 대학원 석사졸업(공학석사). 1987년 9월~현재 : 서일 대학 전자과 교수. 1999년 2월 : 동국대학교 전자공학과 대학원 박사수료. <주관심분야> 퍼지, 신경회로망, 적응제어, system identification 등임



李賢寬(正會員)

1991년 5월 : 인디애나주립대학 전자공학과 졸업(공학학사). 1992년 12월 : 인디애나주립대학 전자공학과 대학원 석사졸업(공학석사). 1999년 2월 : 동국대학교 전자공학과 대학원 박사수료. 2001년 3월~현재 : 호남 대학교 인터넷프로그래밍학과 교수. <주관심분야> 퍼지, 신경회로망, 적응제어, system identification 등임

嚴基煥(正會員) 第38卷 SC編 第4號 參照