

중복 학습 방지에 의한 역전파 학습 알고리즘

(Back-Propagation Algorithm through Omitting Redundant Learning)

白 俊 豪,* 金 有 信,** 孫 慶 植**

(Jun Ho Baek, Yoo Sin Kim, and Kyung Sik Son)

要 約

본 논문에서는 각 패턴들의 오차의 평균치를 수렴 기준치보다 작을 때 학습된 것으로 간주하고, 각 패턴 중에서 이미 학습된 것은 역전파 과정에서 가중치 수정을 생략하여 학습 속도를 개선하는 역전파 알고리즘을 제안했다. 제안된 방법을 XOR, Parity 검사 및 필기체 숫자인식에 적용했다. 필기체 숫자인식에서는 은닉층의 뉴우런 수를 변화시켜가며 제안된 방법을 적용하여 학습속도를 2배 이상 개선했고 학습이 진행함에 따라 학습해야 할 패턴이 학습초기부터 감소됨을 확인했다. 학습패턴 수가 증가함에 따라 그리고 은닉층의 뉴우런 수가 증가함에 따라 그 효과는 커짐을 보였다. 제안된 알고리즘은 인식률과 학습률에 있어서 기존의 알고리즘과 거의 같았다.

Abstract

In this paper the back-propagation algorithm through omitting redundant learning has been proposed to improve learning speed. The proposed algorithm has been applied to XOR, Parity check and pattern recognition of hand-written numbers. The decrease of the number of patterns to be learned has been confirmed as learning proceeds even in early learning stage. The learning speed in pattern recognition of hand-written numbers is improved more than 2 times in various cases of hidden neuron numbers. It is observed that the improvement of learning speed becomes better as the number of patterns and the number of hidden numbers increase. The recognition rate of the proposed algorithm is nearly the same as that conventional method.

I. 서 론

다층 퍼셉트론의 학습이론으로서 Rumelhart가 제

안한 역전파 학습 알고리즘^[1]은 여러 분야에서 이용되어 왔다. 이 학습 방법의 문제점 중의 하나는 장시간의 학습시간을 요하는 것이다. 특히 학습 데이터가 많이 필요한 패턴 인식에 역전파 알고리즘을 적용할 경우 엄청난 학습 시간을 요한다. 학습 속도를 개선시키기 위해 여러가지 방법이 제안되어 왔다. 이들 중에는 학습 파라메타 활성화(activation)함수^[2, 3, 4], 오차함수^[5] 등의 변경이 있다. 최근에 Higawara^[6] 및 Kung^[7]은 초기에 어느 정도 많은 뉴우런을 가지고 시작하여 학

* 準會員, (株) 三星電管
(Samsung Electron Devices Co., Ltd)

** 正會員, 釜山大學校 電子工學科
(Dept. of Elec. Eng., Pusan Nat'l Univ.)

接受日字: 1991年 11月 28日

습이 진행됨에 따라 은닉층 뉴우런을 제거하는 방법을 이용하여 수렴 속도 개선을 시도하고 있다. 학습 데이터의 처리 과정 자체를 변경시켜 학습 속도를 개선시키려고 Sawai^[8] 및 전운호^[9]는 시도 하였다. Sawai 등은 적절한 최소 오차를 정해서 각 패턴의 오차가 이 오차 보다 적으면 역방향 과정은 생략하여 학습의 마지막 단계에서 75%의 학습 패턴이 역방향 과정을 거치지 않음으로 학습 속도가 증가함을 보였다. 전운호 등은 이 방법을 적용하여 학습 시간을 1/2 정도 줄였다. 그러나 이 방법은 기준되는 최소 오차를 전체오차와 비교하기 때문에 학습이 진행함에 따라 학습 되어야 하는 패턴 수가 감소하는 과정을 분석 하는데 효과가 적었고 효과적으로 학습속도도 개선시키지 못했다. 학습 데이터 처리 과정 자체를 수정하여 이중 학습 내지 중복학습을 제거하지 않고서는 학습속도 개선은 한계가 있다.

본 논문에는 수렴을 결정할 때 사용하는 학습 오차를 각 패턴들의 오차의 평균 오차를 사용하여 이 평균 오차가 수렴 기준으로 설정한 값보다 적으면 수렴한 것으로 인정하며, 학습이 진행됨에 따라 각 패턴의 오차를 구해 이것을 기준치와 비교하여 그보다 적으면 역 방향 과정에서 가중치를 수정하지 않음으로써 중복 학습을 피하여 학습 속도를 증가시키는 역전파 알고리즘을 제안했다. 이를 편의상 학습 패턴 감소 역전파 알고리즘이라고 부르기로 한다. 또한 이 알고리즘을 분석하여, 학습을 마친 패턴 수가 반복횟수(iteration)가 변함에 따라 어떻게 변하는지 검토했다. 본 논문에서 제안한 알고리즘과 Sawai 알고리즘과의 차이점을 이론적으로 비교 분석하였다. 제안된 알고리즘은 인식률과 학습률에 있어서 기존의 알고리즘과 거의 같았다.

II. 학습 패턴 감소 역전파 알고리즘

다층 퍼셉트론의 구조는 입력층과 출력층 그리고 그 사이에 은닉층으로 구성된다. Rumelhart가 제안한 오차 역전파 알고리즘(error backpropagation)은 가중치에 대한 오차함수의 급하강법(steepest descent method)을 사용하여 반복적 계산으로 오차를 최소화 하는 것이다. 이의 학습을 위한 역전파 알고리즘은 다음과 같다.

$$E = \sum_p E_p = \frac{1}{2} \sum_p \sum_k (t_{pk} - O_{pk})^2 \quad (1)$$

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial W_{kj}} \quad (2)$$

$$\delta_{pk} = (t_{pk} - O_{pk}) f'(net_{pk}) \quad (3)$$

$$\delta_{pj} = f'(net_{pj}) \sum_k \delta_{pk} W_{kj} \quad (4)$$

$$\Delta_p W_{kj}(n) = \eta \delta_{pk} O_{pj} + \alpha \Delta_p W_{kj}(n-1) \quad (5)$$

k : 출력층의 k번째 뉴우런

j : 은닉층의 j번째 뉴우런

p : p번째 패턴

각 층은 뉴우런으로 구성되고 각 층에 있는 뉴우런은 인접층과 서로 연결되어 각 연결은 가중치를 가진다. 그 구조는 다음 그림과 같다.

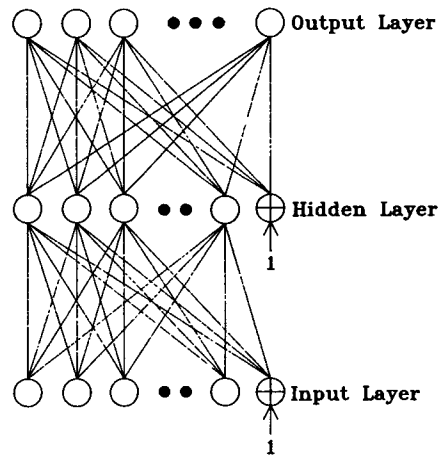


그림 1. 다층 퍼셉트론의 구조.

Fig. 1. The structure of multi-layer perceptron.

여기서 t_{pk} 는 p번째 패턴에 대한 k번째 출력 뉴우런의 원하는 출력이고 O_{pk} 는 실제 나오는 출력이다.

학습시켜야 하는 총 패턴 수를 n_p 라고 하자. 종래의 방법은 모든 학습패턴의 오차 합계가 어떤 ϵ 보다 작을 때까지 모든 학습패턴을 매 학습 때마다 학습시키게 된다. 따라서 이미 학습이 되었다고 생각되는 것까지 중복해서 학습과정을 거치게 된다.

본 논문에서 제안하는 알고리즘은 이 중복학습 과정을 피하는 것이다. 모든 학습시켜야할 패턴을 한번 신경회로망에 입력시켜 출력을 계산하여 가중치를 수정하는 과정을 반복횟수 1회라고 정의한다. 식 (2)에서 계산된 각 패턴의 오차의 합을 패턴 수로 나눈 것을 평균오차라 정의한다. 그 식은 다음과 같다.

$$E = \frac{1}{2} \sum_{\tau}^{np} (tpk - opk)^2 \quad (6)$$

$$\bar{E} = \frac{1}{n_p} \sum_{\tau}^{np} E_i$$

여기서 \bar{E} 는 평균오차이고 E_i 각 패턴의 오차이다.

학습은 윗층에서 아래층으로 오차가 δ_{pk} 에 의해 역전파하여 식(5)에 의해 가중치를 수정함으로서 이루어진다. 각 패턴을 학습시킬 때 평균오차는 주어진 기준치(ϵ)보다 적지 않지만 각각의 오차를 E_p 라 했을 때 어떤 패턴들은 ϵ 보다 작을 경우가 생긴다. 이 패턴들을 p_1, p_2, \dots, p_l 이라 하고 패턴들의 오차를 E_{p1}, \dots, E_{pl} 이라고 하자.

$$E_{p1}, \dots, E_{pl} < \epsilon \quad (8)$$

이 패턴 p_1, p_2, \dots, p_l 에 대해서는 식 (4)와 (5)에 따라 δ_{pk} 값이 무척 작기 때문에 식 (1)에 의해 가중치 변화에 영향을 거의 미치지 않는다. 그러나 종래의 역전파 알고리즘의 경우에는 이러한 패턴들도 매번 학습시키는 불필요한 과정을 수행한다.

본 논문에서 제안한 학습 패턴 감소 알고리즘에서는 이러한 패턴 p_1, p_2, \dots, p_l 의 학습 과정을 생략한다. 그 알고리즘은 다음과 같다.

먼저 순방향 과정을 통하여 입력된 각 패턴의 출력을 발생시키고 각 패턴의 오차를 구한다. 그리고 각 패턴의 오차가 주어진 ϵ 보다 작게되면 그 패턴은 이미 학습된 것으로 생각하고 그 패턴에 대해서는 역전파를 통한 가중치 수정 과정을 생략한다. 그리고 이 패턴들에 의한 출력층의 값은 0 또는 1로 수렴한 상태이므로 식 (4)와 (5)에 의해 가중치 변화에 거의 영향을 주지 않는다. 따라서 위의 생략 과정은 타당하다. 학습이 진행함에 따라 이러한 과정을 각 반복횟수마다 수행할 경우 역방향을 거치는 패턴 수는 점점 줄어들 것이다. 따라서 학습시간은 단축된다. 여기에 대한 반론이 제기될 수도 있다. 즉 패턴 p_1, p_2, \dots, p_l 이 특정한 n 번째의 반복횟수를 거칠때 가중치의 집합 $\{W_{ki}^n, \dots, W_{ki}^n\}$ 에 의해 수렴되었다고 하더라도 다른 패턴들의 학습에 의해 수정된 가중치들에 의해서도 여전히 수렴 되는지는 문제가 된다. 이 문제를 해결하기 위해서 이미 학습되었다고 여겨지는 패턴들은 반복횟수마다 순방향 과정을 거치면서 오차를 다시 구하고 ϵ 보다 큰 패턴들이 생기면 이들은 다시 역방향 과정을 거치면서 가중치 수정을 행한다. 따라서 위에서 제기된 문제는 해결된다. 이를 요약하면 본 논문에서 제안한 학습 패턴 감소

역전파 알고리즘은 참고문헌 (2), (3)의 논문과는 달리 전체 오차를 이미 설명한 수렴 기준치와 비교하지 않고 평균 오차를 기준치와 비교하도록 하는 것 외에는 순방향 과정은 종래의 역전파 알고리즘과 같은 방식으로 수행하되 이미 학습된 패턴은 역방향 과정을 거쳐 가중치 수정을 행하는 불필요한 과정을 생략하는 것이다.

III. Sawai 알고리즘과의 비교 분석

Sawai 알고리즘은 패턴들의 전체 오차 E 가 임의로 정한 ϵ' 보다 작을 때 학습을 멈추지만, 중복 학습을 방지하기 위해서 어떤 최소 오차 δ 를 임의로 정해 각 패턴의 오차 E_p 가 δ 보다 작으면 그 패턴을 학습된 것으로 간주하고 학습을 생략한다. 여기에서는 BP 알고리즘의 난제인 ϵ' 를 결정하는 것 외에 δ 를 결정해야하는 어려움이 있다.

Sawai의 알고리즘을 본 논문에서 제안한 알고리즘과 비교하기 위해서는 임의로 정해서 사용하는 파라미터에 대해서는 공통의 기준이 있어야 한다. Sawai는 학습을 멈추기 위하여 사용한 전체 오차에 대한 수렴 기준치를 ϵ' 이라고 두었기 때문에 이와 같은 기준을 사용하기 위해서는 본 논문의 알고리즘의 평균 오차에 대한 수렴 기준치인 ϵ 을 ϵ'/N (N 은 학습시킬 총패턴 수)와 같은 값으로 한다. Sawai의 알고리즘에서는 δ 를 결정해야 하지만, 본 논문에서 제안한 알고리즘에서는 δ 에 해당하는 것을 다시 결정해야하는 어려움이 없이 ϵ 을 그대로 사용한다. 이러한 조건에서 두 알고리즘을 비교하면 다음과 같은 3가지 경우가 생긴다. 즉 결정된 δ 값이 ϵ 보다 클 경우, 작을 경우 및 같은 경우이다.

1) δ 가 ϵ 보다 클 경우

δ 와 ϵ 의 차이를 $\alpha(>0)$ 라고 두면 Sawai 알고리즘에서는 p 번째 패턴의 오차 E_p 는 $E_p < (\epsilon + \alpha)$ 이다. $E_p < \epsilon$ 경우에 비하면 중복 학습을 더 감소시킬 수 있다. 따라서 학습 속도는 빨라질 것이라고 볼 수 있다. 그러나 개별 패턴의 오차가 $E_p < \epsilon$ 경우에 비해 큰 데도 불구하고 학습된 것으로 간주하는 것으로 보아 인식을 떨어질 것으로 생각된다. 그런데 $\epsilon < E_p < (\epsilon + \alpha)$ 를 만족하는 패턴을 k 개라 두면 경우 전체 오차 E 는 $E = \sum_p E_p = \sum_{p=1} E_p + E_{p=k} E_p$ 로 표시할 수 있다. 따라서 E 는 $k\epsilon < E < (N\epsilon + k\alpha)$ 로서 학습을 생략하기 위한 기준이 처음 정한 것 보다 크게 된다. 그러나 우리가 정해진 기준인 $E < \epsilon'$ ($= N\epsilon$)을 만족해야 되기 때문에 학습은 계속 이루어져야 한다. k 가 일반적으로 0

이 아니고 $\epsilon < E_p < (\epsilon + \alpha)$ 를 만족하는 패턴은 학습된 것으로 간주되었기 때문에 학습을 더 이상하지 않는다. 따라서 전체 오차가 우리가 정해준 기준인 $E < \epsilon'$ ($= N\epsilon$)을 만족할 수가 없기 때문에 결과적으로 학습이 이루어지지 않는다. 그러므로 우리는 δ 가 ϵ 보다 커서는 안된다는 결론을 얻을 수 있다.

2) δ 가 ϵ 와 같을 경우

$\alpha = 0$ 이므로 본 논문의 경우와 동일하다.

3) δ 가 ϵ 보다 적을 경우

δ 와 ϵ 의 차이를 $\alpha (> 0)$ 라고 두면 Sawai 알고리즘에서는 p 번째 패턴의 오차 E_p 는 $E_p < (\epsilon - \alpha)$ 이다. 즉 학습을 생략하기 위한 기준이 본 논문의 경우 보다 작다. $E_p < \epsilon$ 경우에 비하면 중복 학습은 훨씬 많이 이루어진다. 따라서 학습 속도는 늦어질 것이라고 볼 수 있다. 그러나 개별 패턴의 오차가 $E_p < \epsilon$ 경우에 비해 작아야만 학습된 것으로 간주하기 때문에 인식률이 클 것으로 생각된다. 그런데 $(\epsilon - \alpha) < E_p < \epsilon$ 을 만족하는 패턴을 k' 개라 두는 경우 전체 오차 E 는 $E = \sum_p E_p = \sum_{p=1}^{k'} E_p + \sum_k E_p$ 로 표시된다. $k'(\epsilon - \alpha) < E < \{k'\epsilon + (N - k')(\epsilon - \alpha)\}$ 이고, 이는 $k'(\epsilon - \alpha) < E < \{N\epsilon - (N - k')\alpha\}$ 이다. 그런데 k 는 N 보다 작기 때문에 학습을 생략하기 위한 기준이 처음 정한 것 보다 작게 된다. 이것은 우리가 정해준 기준인 $E < \epsilon'$ ($= N\epsilon$)을 충분히 만족시키면서 전체 오차는 더 적기 때문에 인식율은 더 높을 것으로 생각된다. 여기에서는 학습 시간과 인식율 사이에서 본 논문과 비교할 때 trade-off가 있다고 생각된다. 그런데 처음에 정한 수렴 기준치 ϵ' 은 학습하는데 최적에 가까운 값이라고 정한 것이다. 따라서 처음에 정한 수렴 기준치 ϵ' 보다 작은 값에서 수렴하도록 한다는 것은 결국 불필요하게 더 많은 학습을 요구하는 셈이다. Sawai는 δ 를 매우 작은 값으로 잡았다고 하며, 자신의 실험에서 마지막 단계에서 75%의 패턴들이 역전과를 생략한 걸로 이야기하는 것을 미루어 보아, 아마 ϵ 보다 작았음에 틀림없다. 그렇기 때문에 그의 실험에서는 중복 학습 생략 효과가 학습의 마지막 단계에서 나타난 것이라고 추정된다.

요약하면 Sawai 알고리즘과 본 논문의 알고리즘의 차이는 δ 를 정하는 것이 본 논문에서는 평균오차(또는 전체오차)에 대한 수렴 기준이 주어지면 δ 의 값은 이것과 동일하게 자동적으로 정해지지만 Sawai의 경우는 δ 의 값을 정하는 기준이 없고 임의로 최소 값을 정하는데 만약 δ 의 값이 ϵ 보다 크면 수렴이 이루어지지 않기 때문에 학습 시키는데 많은 어려움이 있다.

IV. 실험 및 결과

학습 패턴 감소 알고리즘을 시험하기 위해 3가지 실험을 했다. 이것은 XOR와 Parity 검사와 보다 복잡한 필기체 숫자 인식이다. 이 알고리즘의 순서도는 그림 2와 같다.

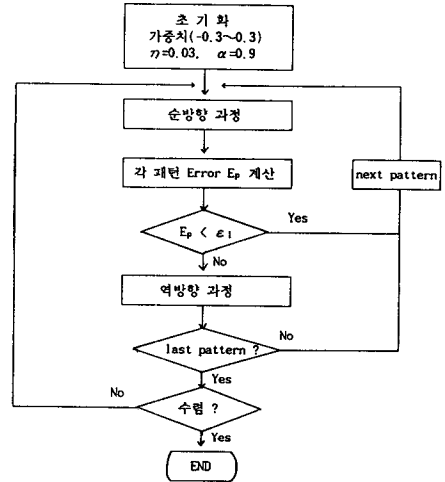


그림 2. 제안한 알고리즘의 순서도.
Fig. 2. Flow chart of the proposed algorithm.

1. XOR 학습

XOR 학습은 단순하고 비선형적 신경회로망 실험에 자주 사용된다. 종래의 방법과 본 논문에서 제안한 방법 모두 동일한 학습 상수를 사용했다. 학습 패턴 감소 알고리즘에 의하면 학습이 진행함에 따라 학습 패턴 수는 줄어든다. 그것은 그림 3에 있다.

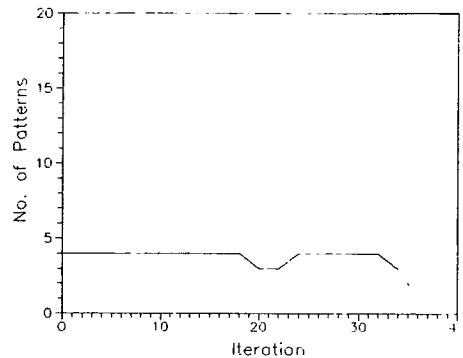


그림 3. XOR에서 학습패턴 수와 반복횟수.
Fig. 3. Number of learning patterns vs. iteration in XOR.

그림 4에서는 17~18번 정도의 반복횟수에서 학습 패턴 수가 1개 감소 되다가 다시 20번의 반복횟수를 넘어서면서 학습시키는 패턴 수는 다시 4개로 증가한다. 이것은 2장에서 설명했듯이 이미 학습된 패턴은 역방향 과정에서 가중치 수정은 생략할지라도, 다른 패턴들의 학습에 의한 수정된 가중치에 의해 순방향 과정을 거치면서 오차를 계산한다. 이 오차는 미리 설정한 기준값보다 커졌기 때문에 재차 학습을 해야하기 때문이다. 그러나 반복횟수가 30번이 넘어서면서 급격하게 학습해야 될 패턴 수가 줄어들었다. 학습속도는 이 알고리즘이 종래의 방법보다 훨씬 빠르다. XOR의 경우 학습패턴의 수는 4개이다. 실험결과는 표 1에 나타났다. 기존의 역전파 알고리즘과 제안한 학습 알고리즘은 수렴속도는 서로 비슷하다. 수렴하지 못한 경우에 역전파 알고리즘의 학습시간과 제안한 학습 알고리즘의 학습시간을 비교하면 제안한 학습 알고리즘이 같거나 조금 더 많은 학습시간이 요구된다. 이것은 각 학습패턴의 수렴여부를 판정하기 위해 요구되는 시간에 비해 학습 패턴이 대단히 적어 중복 학습 생략 효과가 미미하기 때문인 것이다.

표 1. XOR 실험결과

- ① 종래의 BP 알고리즘
- ② 학습 패턴 감소 알고리즘

Table 1. Experimental results of XOR.

- ① conventional BP algorithm,
- ② proposed BP algorithm.

	#1		#2		#3		#4		#5	
	①	②	①	②	①	②	①	②	①	②
반복횟수	36	37	193	188	37	38	31	43	5000	5000
수렴여부	0	0	0	0	0	0	0	0	X	X
학습시간	1	1	3	3	1	1	1	2	30	65

2. Parity 검사

Parity 3은 학습패턴 8개를 가지는 Parity검사 실험이다. 실험결과는 표 2에 나타내었다. 기존의 역전파 알고리즘과 제안한 학습 알고리즘과의 수렴시간 비교는 서로 비슷하였다. XOR의 경우 보다 패턴 수가 8개로서 조금 많기 때문에 중복 학습 생략 효과에 의한 학습속도는 XOR 보다 조금 낮지만 여전히 개선되지는 않았다. 이 때의 학습패턴과 반복횟수의 그래프는 그림4에 잘 나타나 있다.

약 50번 정도의 반복횟수가 지나면서 이미 학습된 패턴 수가 나타나기 시작한다. 50번 이상에서 보이는

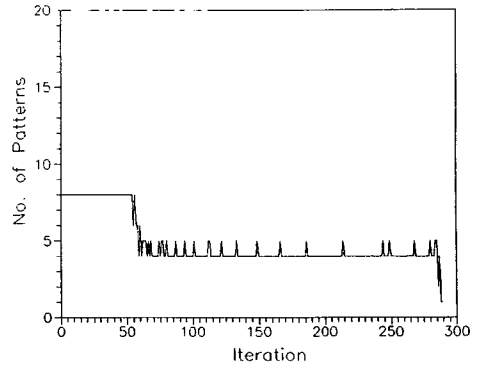


그림 4. Parity 3에서 학습패턴 수와 반복횟수
Fig. 4. Number of learning patterns vs. iteration in Parity 3.

임펄스 같은 모습은 이미 학습된 값이 다른 패턴의 학습에 의해 수정된 가중치에 의해 새로 계산된 오차가 커졌기 때문에 학습되지 않은 것으로 간주되어 다시 역방향 과정의 학습을 해야 한다. 이러한 현상 때문에 그림 4에서 임펄스 같은 것이 보이는 이유이다.

표 2. Parity 3 실험결과

- ① 종래의 BP 알고리즘
- ② 학습 패턴 감소 알고리즘

Table 2. Experimental results of Parity 3.

- ① conventional BP algorithm,
- ② proposed BP algorithm.

	#1		#2		#3		#4		#5	
	①	②	①	②	①	②	①	②	①	②
반복횟수	2321	1971	2614	1346	3684	2140	2407	1373	5000	5000
수렴여부	0	0	0	0	0	0	0	0	X	X
학습시간	87	64	98	44	138	68	91	44	187	165

Parity 4는 학습패턴이 16개인 Parity 검사 실험이다. 실험결과는 표 3에 나타내었다. 결과는 기존의 역전파 알고리즘보다 제안한 학습 알고리즘의 수렴시간이 향상되었음을 알 수 있다. 수렴하지 못한 실험결과에서도 기존의 역전파 알고리즘보다 제안한 학습 알고리즘이 수렴속도가 빠르다.

그러므로 지금까지의 세가지 실험결과에서 추론할 수 있듯이 학습패턴 수가 증가함에 따라 기존의 역전파 알고리즘보다 제안한 학습 알고리즘의 수렴속도가 더욱 향상될 것이다. 이는 학습초기에 많은 패턴이 이

미 수렴되고 나머지 남은 패턴만이 수렴하지 못하고 학습이 이루어지기 때문이다. 따라서 수렴이 끝난 패턴의 중복학습을 피함으로써 학습속도를 향상시킬 수 있다. 그림 5에서는 학습패턴 감소와 반복횟수의 관계가 나타나 있다. 학습 초기부터 그 효과가 나타난다.

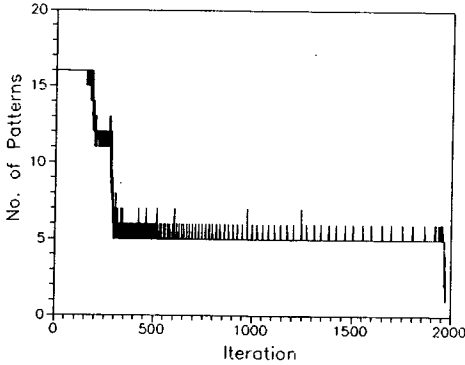


그림 5. Parity 4에서 학습패턴 수와 반복횟수
Fig. 5. Number of learning patterns vs. iteration in Parity 4.

표 3. Parity 4 실험결과

- ① 종래의 BP 알고리즘
- ② 학습 패턴 감소 알고리즘

Table 3. Experimental results of Parity 4.

- ① conventional BP algorithm,
- ② proposed BP algorithm.

	#1		#2		#3		#4		#5	
	①	②	①	②	①	②	①	②	①	②
반복횟수	818	270	232	231	210	218	213	217	5000	5000
수렴여부	0	0	0	0	0	0	0	0	X	X
학습시간	16	6	4	4	4	4	4	4	77	81

3. 필기체 숫자 인식

0에서 9까지의 필기체 숫자 50종류를 종래의 역전파 알고리즘으로 학습시킨 경우와 제안된 알고리즘으로 학습시킨 경우와 학습속도와 인식률을 비교했다. 사용한 다층 퍼셉트론의 구조는 출력층 뉴우런 수는 10개 이고 입력층 뉴우런 수는 22개 은닉층 뉴우런 수는 7개에서 10개까지 변화시켰다. 이것은 은닉층의 뉴우런 수가 몇개가 최적인지 알지 못하기 때문이다. 종래의 방법과 학습 패턴 감소 방법에 의한 학습속도를 비교한 결과는 그림 6과 같다. 인식률은 종래의 방법과 학습 패턴 감소 방법과 비슷하였다. 학습시킨 패턴에 대

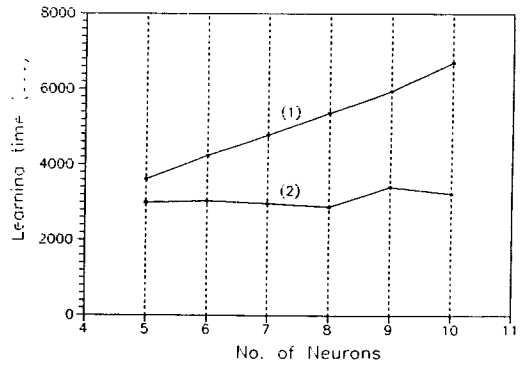


그림 6. 은닉층 뉴우런 수와 학습속도
Fig. 6. No. of neurons of hidden layer vs. learning time.

한 인식률은 98%정도이며, 학습시키지 않은 패턴에 대한 인식률은 95%정도였다.

그림 6에서 볼때 학습 패턴을 감소시켜 나가는 학습 알고리즘에 의한 학습 결과가 수렴까지의 학습시간이 2배 이상 감소하였다. 특히 기존의 학습 알고리즘의 학습시간이 은닉층의 뉴우런 수가 증가함에 따른 학습시간의 증가비율과 학습 패턴 감소 알고리즘의 은닉층의 뉴우런 수가 증가함에 따른 학습시간의 증가비율을 비교하면 기존의 알고리즘에 의한 것보다 학습 패턴 감소 알고리즘의 결과가 비율이 작다는 것을 볼 수 있다. 따라서 은닉층의 뉴우런 수가 증가함에 따라 학습시간은 차이가 더욱 커질 것이다. 이렇게 되는 것은 학습 패턴 감소 알고리즘에서 학습이 진행됨에 따라 역방향 과정을 거치는 패턴의 수가 감소되어 학습 시간을 감소시킬 수 있기 때문이다. 그리고 그림 7은 제안한 학습 패턴 감소 알고리즘에 의하여 학습이 진행됨에 따라 역방향 과정을 거치는 학습 패턴의 수를 나타낸 것이다. 그림 7에서는 학습이 시작하지 불과 20번 만에 학습 패턴의 감소가 나타나기 시작하면서 200번 이전에 80%인 400개의 패턴이 이미 학습된 것을 알 수 있다. 이와같은 학습 패턴 감소가 초기에 나타난 것은 학습속도를 2배 이상 증가 시킨 요인인 것이다. 그리고 그림 7에서 학습이 진행함에 따라 학습 패턴의 수가 진동하는 것을 볼 수 있다. 이것은 이전의 학습에서 수렴한 것으로 판정된 패턴이 다음 학습에는 수렴되지 않은 것으로 판정되는 경우이다. 이러한 현상은 이전에 학습되지 않은 패턴에 의하여 가중치의 변화가 발생하게 되고 이러한 가중치의 변화로 이미 수렴한 것으로 판정된 패턴에 대한 수렴 상태가 바뀐다. 그러

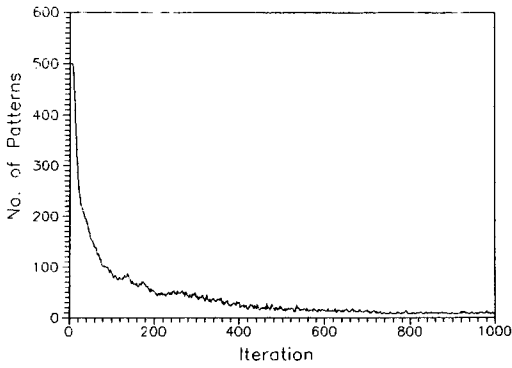


그림 7. 필기체 숫자인식에서 학습패턴 수와 반복횟수.

Fig. 7. Number of learning patterns vs. iteration in hand-written numbers.

나 이러한 과정을 반복하므로써 모든 패턴을 동시에 만족할 수 있는 가중치를 구할 수 있다.

V. 결 론

본 논문에서는 다층 퍼셉트론의 역전파 알고리즘의 학습 속도를 개선하기 위해 가중치를 중복해서 수정을 하지 않도록 하여 학습이 진행함에 따라 학습해야 할 패턴을 감소시켜 가는 알고리즘을 제안했다. 제안한 알고리즘을 Sawai 알고리즘과의 차이점을 이론적으로 비교분석 하였다. 제안한 알고리즘에 의한 중복학습 생략 효과는 일반적으로 생각하는 것처럼 마지막 단계에서 나타나는 것이 아니라, 학습 초기에도 이미 학습된 패턴수가 많이 생겨 역방향 과정을 통한 가중치 수정이 필요없는 경우가 생긴다. 심지어 XOR이나 Parity 검사와 같은 경우에 패턴 수가 매우 적은 경우도 그 효과가 발생한다. 본 논문에서는 수렴을 결정하는 데 사용한 오차를 각 패턴의 오차의 합인 전체 오차를 사용하지 않고 평균 오차를 사용하였기 때문에 각 패턴의 학습시 발생하는 오차가 평균값을 중심으로 분산되어 있기 때문에 학습 초기에도 효과가 나타난 것이다. 학습 패턴 감소의 효과는 학습되는 패턴의 오차들이 오차의 평균치를 중심으로 어떻게 분포하는가에 크게 의존한다. 필기체 숫자 인식이 있어서 500개의 패턴을 사용하고 은닉층의 뉴우런 수를 7개에서 10개 까지 변화 시켰다. 학습 속도는 2.1 배에서 2.3배 정도 까지 증가했다. 은닉층 뉴우런 수가 증가할수록 학습 속도 개선이 잘되었다. 따라서 수많은 패턴을 학습해야 하고 은닉층의 뉴우런 수가 많아질 경우는 그 효과

가 현저하리라고 생각한다. 제안된 알고리즘은 인식률과 학습률에 있어서 기존의 알고리즘과 거의 같았다.

參 考 文 獻

- [1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: Foundations, MIT Press, pp. 318-362, 1986.
- [2] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Acceleration the convergence of the back-propagation method," *Biol. Cybern.* 59, pp. 257-263, 1988.
- [3] J. R. Chen and P. Mars, "Step size variation methods for accelerating the back propagation algorithm," *Proc. IJCNN'90*, vol. 1, pp. 601-710, Jan. 1990.
- [4] A. Rezgui and N. Tepedelenioglu, "The effect of the slop of the activation function on the back propagation algorithm," *Proc. IJCNN'90*, vol. pp. 707-710, Jan. 1990.
- [5] A. Krzyzak, W. Dai and C. Y. Suen, "Classification of large set of handwritten charaters using modified back propagation model." *IJCNN'90*, vol. 3, pp. 225-232, 1990.
- [6] Masafumi Haigawara, "Novel back propagation algorithm for reduction of hidden units and acceleration of convergence using artificial selection." *Proc. ICJN' 90*, vol. 1, pp. 625-630, 1990.
- [7] S. Y. Kung and J. N. Hawang, "An algebraic projection analysis for optimal hidden units size and learning rates in back-propagation searning," *Proc. IJCNN'88*, vol. 1, pp. 363-370, 1988.
- [8] H. Sawai, A. Waibel, P. Haffner, M. Miyatake, K. Shikano, "Parallelism, hierarchy, scaling in time-delay neural networks for spotting japanese phonemes /CV-syllables," *IEEE 3rd int'l Conf. on Neural*

Network, vol. 2, pp. 81-88, 1989.

망을 이용한 필기체 한글문자 인식," 전자공학회
논문지, 제28권, B편, 제 8 호, pp. 85-93, 1991.

[9] 全允豪, 崔悰鎬, "가중치를 공유하는 신경회로

著 者 紹 介

白 俊 豪 (正會員) 第 29 卷 B編 第 4 號 參照
현재 (주)삼성 전관 근무

孫 慶 植 (正會員) 第 29 卷 B編 第 4 號 參照
현재 부산대학교 전자공학과
부교수



金 有 信 (正會員) 第 29 卷 B編 第 4 號 參照
현재 부산대학교 전자공학과
부교수