

다층신경망에서 하이브리드 학습 규칙의 구현에 관한 연구

A Study on the Implementation of Hybrid Learning Rule for Neural Network

송도선*, 김석동**, 이행세***
(Do Sun Song*, Suk Dong Kim*, Haing Sei Lee***)

요약

본 논문에서는 다층구조 순방향 신경회로망에 적용될 수 있는 것으로 입력의 특징 추출기능(Feature Extractor)이 우수한 Hebb 학습 규칙과 패턴 분류 기능(Classifier)이 우수한 BP 알고리즘을 결합한 Hybrid 학습 규칙을 제안하고자 한다.

오차역전파 학습법칙을 적용한 다층구조퍼셉트론(MLP)과는 달리, 다층구조에 오차역전파 학습법칙과 Hebb 학습법칙이 동시에 적용될 수 있는 Hybrid(Hebbian + BP) 학습법칙은 학습시에 출력층의 연결강도를 제외한 모든 연결강도 계산에 적용되며 출력층에는 기존의 오차역전파법칙만이 적용된다. 출력층에 Hebb 학습법칙을 제외시킨것은 다층구조학습시에 학습의 수렴성에 대한 보장이 주어지지 않기 때문이다.

제안된 Hybrid 학습법칙의 성능평가를 위해 몇가지의 영역구분 문제에 적용한 결과 제안된 학습법이 기존의 BP보다 우수함을 보였다. 학습속도면에서는 기존의 BP법칙에 비해 훨씬 빠른 수렴속도를 보여 주었는데, 그중 한가지 예를 보면 제안된 Hybrid법칙에 의한 학습은 기존의 BP의 학습회수의 2/10만으로도 가능함을 보여주었다. 인식률에서도 제안된 법칙에 의한 결과가 BP에 의한 결과보다 최고 약 0.77% 우수하다.

ABSTRACT

In this paper we propose a new Hybrid learning rule applied to multilayer feedforward neural networks, which is constructed by combining Hebbian learning rule that is a good feature extractor and Back-Propagation(BP) learning rule that is an excellent classifier.

Unlike the BP rule used in multi-layer perceptron(MLP), the proposed Hybrid learning rule is used for update of all connection weights except for output connection weights because the Hebbian learning in output layer does not guarantee learning convergence.

To evaluate the performance, the proposed hybrid rule is applied to classifier problems in two dimensional space and shows better performance than the one applied only by the BP rule.

In terms of learning speed the proposed rule converges faster than the conventional BP.

For example, the learning of the proposed Hybrid can be done in 2/10 of the iterations that are required for BP, while the recognition rate of the proposed Hybrid is improved by about 0.77% at the peak.

*중경 공업 전문대학 전자계산기과

**호서 대학교 전자계산학과

***아주 대학교 전자공학과

접수일자: 1994년 5월 17일

I. 서 론

컴퓨터가 인간의 음성이나 문자를 인식하는 연구는 오래전부터 연구되어져 왔으며, 그 중에서 인간의 뇌의 신경 구조와 유사한 신경망을 이용한 음성인식이나 문자인식이 미래에 우수한 인식 방법의 한 종류가 될 것이라 예견되어 이 분야에 대한 많은 연구가 되어져 왔다[1, 2].

신경망을 이용한 음성 및 문자인식은 대부분 지도 학습인 오차역전파(BP: Back-Propagation)기법을 이용한 것으로 BP알고리즘은 입력패턴을 분류하는 기능(Classifier)은 매우 우수하나 학습 시간이 느린 단점이 있다[3, 4].

자유 학습인 Hebb학습법은 이 알고리즘을 신경망에 적용한 Hopfield 메모리와 Kosko가 제안한 BAM(Bidirectional Associative Memory)등이 있다[5, 6, 7].

그러나 위의 두 신경망모델은 단층 구조(Single Layer)로 구성되므로 학습속도나 특징 추출 기능(Feature Extractor)은 우수하지만 학습패턴의 저장능력이 매우 제한되고 선형적 분리가 되지 않는 문제에는 적용할 수 없는 단점이 있다[8, 9, 10, 11].

최근에는 BP와 Hebb 학습법을 결합한 hybrid 학습법으로 BP학습시의 우수한 패턴 분류능력과 Hebb 학습시의 특징 추출 기능의 장점을 동시에 살리기 위하여 연구가 되어지고 있으며 R.J.Wood가 제안한 hybrid학습법칙은 실험결과가 만족할 만한 정도가 아니며 특히 출력층까지 hybrid법칙이 적용되어 있다[12].

다층구조의 출력층에 기존의 Hebb법칙이 적용될 경우 학습의 수렴성이 항상 보장된다고는 할 수 없다.

이러한 단점을 보완하기 위해 Competitive Hebbian Rule이 제안되었으나 이것은 출력층의 각 비트 사이에 패턴에 대한 직교성(orthogonality)이 되게 학습을 시켜야 하며 이러한 직교성이 만족되도록 학습을 시키려면 학습패턴의 수(저장능력)가 상당히 제한될 수 있다.

A/D 변환된 음성 데이터는 데이터의 양이 너무 크기 때문에 신경망의 입력 데이터는 변환된 데이터로부터 LPC 및 PARCOR 등의 특징 벡터를 추출하여 신경망의 데이터로 사용한다. 특히 이러한 특징 벡터는 연속된 신호원들이기 때문에 그 특징 벡터들 내부에도 이것들로 이루어지는 특징이 있을 것이다. Hybrid학습법을 연구한 이유는 음성의 특징 벡터 즉

PARCOR를 신경망의 입력 데이터로 사용하여 BP 학습에 의해 인식하지만 잡음이 있거나 연속음성에서는 만족할만한 수렴속도나 인식율을 얻지 못했다. 그러나 신경망 입력층과 은닉층에 BP의 분류기능과 Hebbian의 특징 추출기능을 동시에 만족되게 하면 PARCOR에 의해 주어지는 특징 벡터들에서 또다시 신경망의 입력층과 은닉층에서 특징 추출 및 분류 기능이 가능하게 하여 은닉층과 출력층에서 BP로 분류하게 하는 기능을 수행하게 하여 성능을 향상시켜보자는 것이다.

따라서 본 연구는 잡음이 있는 음성인식과 연속음성 인식을 하기위한 기초 연구로서 학습시간도 짧고 인식률도 BP보다 우수한 Hybrid학습의 알고리즘을 제안하는 것으로서 2장에서는 오차 역전파 학습법칙을 적용한 다층구조퍼셉트론(MLP)과는 달리 오차 역전파 학습법칙과 Hebb학습법칙을 동시에 적용할 수 있는 Hybrid학습법칙을 제안 하였다.

실험 및 결과에서는 수학적으로 증명하기 어려운 문제 즉 랜덤데이터를 임의의 알고리즘으로 처리하고 그 처리 결과를 확률적으로 고려하는 방법인 몬테칼로(Monte Carlo) 방법으로 랜덤함수에서 발생하는 좌표점을 이용하여 곡선의 변화가 느린 두개의 원이 조금 겹치게 한 후 겹치는 영역과 나머지 영역을 구분하는 문제와 곡선의 변화가 큰 삼각형 4개의 영역을 구분하는 문제 및 두개의 사각형이 겹치게 한 후 영역을 구분하는 기하학적 문제를 선정하여 BP와 Hybrid의 성능을 각각 분석 비교하였다.

제안된 학습법 구현은 학습시 입력층과 은닉층 사이의 연결강도 계산에만 Hybrid(Hebbian+BP) 학습 규칙을 적용하고 은닉층에서 출력층까지는 BP 학습 규칙을 구현하였으며 이것은 Hebb법칙이 출력층에 적용될 경우 학습시의 수렴성에 어려움을 줄 가능성이 있기 때문이다.

결론에서는 본 연구에 대한 일반적인 결과를 서술 하였다.

II. 하이브리드 학습법(Hybrid Learning Rule)

제안하고자 하는 hybrid 학습법칙은 임의의 다층 구조신경망에 적용이 가능하나 설명이 편의를 위해 2층구조(은닉층 1개 포함)에 대해 적용하여 보기로 한다. 기존의 오차역전파 알고리즘은 출력층에서 정의 되는 전체 오차 E 즉,

$$E = \frac{1}{2} \sum_s \sum_j (t_j^s - O_j(x^s))^2$$

을 최소화하는 방향으로 네트워크를 최적화시키는 것으로, (0~1) 형태의 시그모이드 함수를 갖는 경우에 연결강도의 Update는 식 (1)~(3)으로 표현된다.

$$\Delta W_{ji} = \eta \delta_j O_i \quad (1)$$

$$\delta_j = O_j(1-O_j)E_j \quad (2)$$

$$E_j = (t_j^s - O_j) \quad (\text{출력층}) \quad (3.1)$$

$$E_j = \sum_k \delta_k W_{kj} \quad (\text{은닉층}) \quad (3.2)$$

여기서, O_i 와 O_j 는 시냅스 W_{ji} 에 연결되어 있는 뉴런의 실제값(activation value)으로, 각각 입력에 학습패턴 벡터 x^s 를 인가하였을 때 W_{ji} 의 앞단에 연결되어 있는 뉴런의 값 $O_i(x^s)$ 와 W_{ji} 의 뒷단에 연결되어 있는 뉴런의 값 $O_j(x^s)$ 를 나타내는 것으로 $O_j = f(\sum_i w_{ji} O_i + \theta_j)$ 의 관계가 있다. $f(O_j)$ 는 비선형 시그모이드함수이며 그 미분치 $f'(O_j)$ 는 $O_j(1-O_j)$ 의 값을 갖는다. 식 (3.2)에서 W_{kj} 는 은닉층의 j 번째 뉴런과 출력층의 k 번째 뉴런사이의 연결강도를 나타낸다. 좀더 자세히 살펴보면, 출력층의 경우에는 식(1)~(3.1)이 적용되는 것으로 j 번째 출력뉴런과 i 번째 은닉층 뉴런사이의 weight update는 식 (2)와 (3.1)로 결합되는 출력층 오차 δ_j 와 은닉층의 뉴런값(activation) O_i 에 의해 이루어지며, 은닉층의 경우에는 식 (1), (2), 그리고 (3.2)가 적용되는 것으로 j 번째 은닉층 뉴런과 i 번째 입력뉴런사이의 weight update는 식(2)와 (3.2)로 구성되는 오차 δ_j (출력층의 경우와 다른데 여기서는 출력층 오차가 은닉층에 역전파되어온 오차의 양임)와 입력의 i 번째 뉴런에 적용되는 학습패턴의 값 x^s 에 의해 이루어진다. 단, $x^s = [x_1^s, x_2^s, \dots, x_N^s]^T$ 로 $[\cdot]^T$ 는 $[\cdot]$ 의 transpose를, 그리고 N 은 입력뉴런의 갯수를 나타낸다. 한편, (-1~+1)의 시그모이드 함수를 사용하는 경우에는 식(2)에서 $O_j(1-O_j)$ 대신에 $(a/2)(1-O_j)^2$ 를 사용하여야 하며, a 는 시그모이드함수의 exponential slope 즉 $O_j = 2/(1 + e^{-a(\sum_i w_{ji} O_i + \theta_j)}) - 1$ 에서 exponential 부분의 상수이다. 그리고, 식(1)에서 η 는 학습률을 나타낸다. 한편, 일반적인 Hebb 학습법은 식 (4)로 표현되는데

$$\Delta W_{ji} = K_a + K_b(O_i - \theta_i)(O_j - \theta_j) \quad (4)$$

여기서, K_a , K_b , θ_i 는 상수다[8, 9, 12]. Hopfield 모델과 Kosko의 BAM 모델에 적용되는 법칙은 Hebb 법칙의 특수한 경우이나 식(4)가 더욱 일반적인 Hebb 학습법칙으로 잘 알려져 있다. 구체적으로 weight update를 살펴보면 BP법칙의 경우에 식(1)과 식(2)를 결합하면 다음식 (5)가 되며, Hebb법칙의 경우에서 $K_a = 0$, $K_b = \eta \theta_i \theta_j = 0$ 으로 두면 식(6)이 된다.

$$\Delta W_{ji} = \eta O_i O_j (1-O_j) E_j \quad (5)$$

$$\Delta W_{ji} = \eta O_i O_j \quad (6)$$

식(5)와 식(6)을 결합하면 식(7)이 되는데

$$\Delta W_{ji} = \eta O_i O_j [(1-\beta) + \beta] + \beta E(1-O_j) \quad (7)$$

여기서, β 는 hybrid 학습시 사용되어지는 BP와 Hebb update 양의 상대적 크기를 결정해 준다[12]. 식(7)에서는 출력층에 적용될 때에는 Hebb Update 항에 의해 Weight가 발산할 수도 있으며 또한 정확한 목표치 값에 수렴한다는 보장이 없다. 따라서, 본 논문에서는 이러한 문제점을 해결하기 위해 출력층에는 BP법칙만 적용하고 은닉층의 모든 weight에는 Hebb 법칙과 BP 법칙이 동시에 적용되는 hybrid 학습법칙을 제안하고자 한다. (0~1)의 시그모이드 함수를 갖는 경우에도 같은 원리가 적용되나 여기에서는 (-1~1)의 시그모이드 함수의 경우에 대해 적용했으며, 그 Update식 ΔW_H 는 다음과 같다.

$$\begin{aligned} \Delta W_H &= \Delta W_{ji} = \Delta W_{BP} + \Delta W_{HEBB} \\ &= \eta O_i f'(O_j) E_j + \alpha O_i O_j |\delta_j| \end{aligned} \quad (8)$$

여기서, 시그모이드 함수의 exponential slope $a = 1$ 을 사용하였으며, $f'(O_j) = (1-O_j^2)/2$ 이고, $|\cdot|$ 는 절대값을 나타내며, 그리고 α 는 Hebb법칙의 Hebb를(Hebb 학습률)이라 하자. 식(8)의 오른쪽 부분의 두번째 항 즉, Hebb법칙에 대응되는 항의 $|\delta_j|$ 는 오차역전파에서 적용되는 오차항의 절대값으로 Hebb 법칙의 적용에 의해 weight가 무한대로 발산되는 것을 제어하는 동시에 학습이 점점 이루어짐에 따라 일정값으로 weight가 수렴하도록 하는 역할을 한다. 그

러나, 여기서 주의해서 살펴보아야 할 것은 Hebb항의 $|\delta_i|$ 가 오차역전파로부터 인용되어 왔으나 절대값을 취하였기 때문에 BP의 학습법칙을 따르지 않는다는 점이며, 단지 위에 말한바와 같이 Hebb항에 기인한 weight의 무한대 증가를 배제시켜 주며 점점 학습이 이루어짐에 따라 weight의 수렴을 보장해 주는 역할을 한다. 아울러 중요한 사실은 Hebb법칙은 패턴의 특징 추출기능이 있다는 것이 널리 잘 알려져 있는 사실이기 때문에, 식(8)의 hybrid 학습법칙에 의해 신경망을 학습시키면 학습속도의 향상과 인식 성능의 향상을 가져오리라 기대된다.

제안된 hybrid법칙에서 Hebb항에 관계되는 Hebb를 α 가 너무 작으면 BP특성만이 나타나고 ($\alpha=0$ 이면 BP만의 학습임), 너무 크면 Hebb 특성만이 나타나게 됨을 식(8)에서 알 수 있다.

III. 실험 및 결과 분석

(1) 신경망 구성

식(1)의 오차역전파 학습과 식(8)에서 제안된 하이브리드 학습을 구현하기 위한 신경망의 구조는 MLP이며 입력층 은닉층 출력층으로 하였다.

Structure는 입력층 뉴런은 2개, 은닉층 뉴런은 10개, 출력층 뉴런은 원과 사각형은 3개이고 삼각형은 4개이며 시그모이드 함수는 -1과 1의 출력이 나타나는 함수를 사용했다.

제안된 알고리즘 학습시 입력층과 은닉층은 Hybrid 학습을 그리고 출력층은 BP만 적용하였다. BP 학습과 제안된 알고리즘 학습법시 전체오차 E가 허용오차 범위에 있으면 학습된 것으로 하였다.

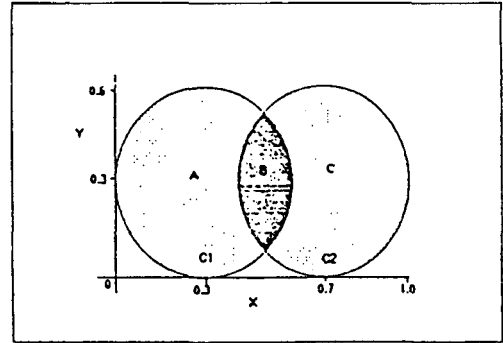
(2) 실험 데이터 구성

신경망에 적용된 실험 데이터는 두개의 랜덤 함수에서 생성되는 2차원 평면상의 점들로서 구성되었다. 즉, $X = \text{RAND}(\text{seed1})$, $Y = \text{RAND}(\text{seed2})$ 이다. 두개의 랜덤 함수에서 발생되는 X, Y의 두수를 직교좌표의 좌표점(X, Y)이라 하자. 랜덤함수를 연속적으로 사용하여 좌표점(X_0, Y_0), (X_1, Y_1).....(X_n, Y_n)를 산출한다. 랜덤함수에서 산출되는 좌표점들을 A, B, C, D그룹으로 분류하여 신경망의 학습데이터로 산출하여 이점들을 A, B, C 그리고 D 그룹으로 분류하는 문제에 적용하여 보았다.

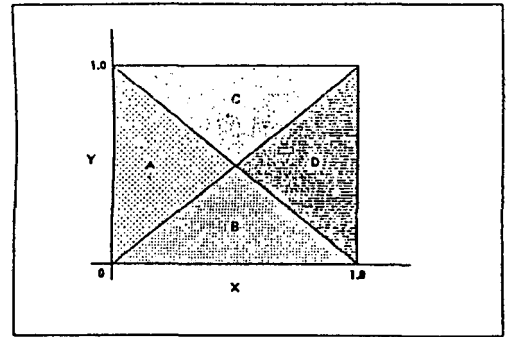
A, B, C, D그룹으로 분류하는 경계 조건은 그림 1의

(a), (b), (c)와 같다.

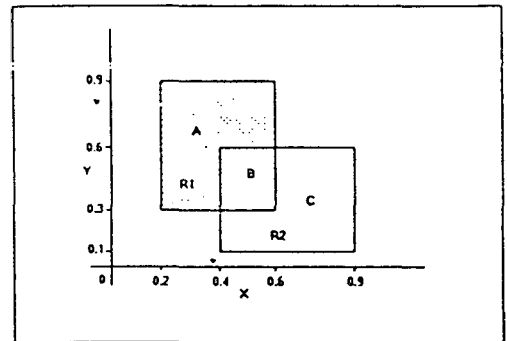
그림 1의 (a)는 두개의 원을 각각 C1, C2라 하고 겹치게 하였다.



(a) 원의 경계 조건
(a) Boundary Condition of the Circle



(b) 삼각형의 경계 조건
(b) Boundary Condition of the Triangle



(c) 사각형의 경계 조건
(c) Boundary Condition of the Rectangle

그림 1. A, B, C, D영역 구분을 위한 경계 조건
Fig 1. Boundary Conditions for the A, B, C, D Groups

원의 중심점의 좌표가 C1는 (0.3, 0.3)이고 C2는 (0.7, 0.3)이며 반지름은 0.3이다.

원 C1과 원 C2가 겹치는 부분을 B영역이라 하고 나머지는 각각 A, B영역으로 하였다.

그림 (b)는 가로와 세로가 1인 정사각형으로서 영역 구분은 좌표(0, 0)에서 좌표(1, 1)로 그린 대각선과 좌표(1, 0)에서 좌표(0, 1)로 그린 대각선으로 구분된 4개의 삼각형 A, B, C, D의 영역이다.

그림 (c)는 가로 세로의 길이가 다른 두개의 사각형으로 구성되어 사각형 R1은 4개의 좌표가 각각 (0.2, 0.3), (0.2, 0.9), (0.6, 0.3), (0.6, 0.9)이고 사각형 R2는 (0.4, 0.1), (0.4, 0.6), (0.9, 0.1), (0.9, 0.6)으로 이루어지며 두개의 사각형이 겹치는 부분을 B 나머지를 A와 B로 구분하였다.

그림1의(a), (b), (c)의 영역구분시 선 및 곡선에 나타나는 좌표점은 제외했다.

(3) 구현 및 결과 분석

본 연구에서 제안된 식 (8)의 성능을 평가 하기 위해 그림 1의 A, B, C, D영역 구분을 위한 몬테칼로 방법은 다음과 같다.

첫째, A, B, C 및 D영역으로 분류되는 데이터 400 개씩을 산출한다.

둘째, A, B, C 및 D영역에서 각각 100개의 데이터를 A, B, C, D영역구분을 위한 신경망의 학습 데이터로 한다.

셋째, 학습된 데이터를 이용하여 학습에 참여하지 않은 각 영역의 300개 데이터로 인식한다.

위의 방법을 실현하기 위한 학습법은 BP와 Hybrid이며 사용된 컴퓨터는 INTEL 486 PC로서 단 하나의 컴퓨터만 실험하였다.

첫번째로 모우멘텀(Momentum)을 사용하지 않는 BP와 Hybrid를 평가하였다.

실험조건은 학습 계수 η 가 변화할때 BP와 hybrid의 성능변화를 고찰하기 위하여 η 를 0.1과 0.05로, 출력허용오차(tolerance)는 0.1, α 는 자기 인식시 그림 1의 세가지 패턴중 자기오인식이 나타나지 않는 α 의 최소 범위중에서 0.001, 0.01, 0.02의 세가지로 제한했으며 최대 학습 횟수(Learning Iteration)는 30000으로 하였다. $\eta=0.1$ 일때 실험 결과는 표 1과 같다.

각 표에서 학습 횟수는 학습을 완료한 횟수이며 자기오인식은 학습데이터의 자기인식 결과이고 오인식

표 1. η 가 0.1일때 모우멘텀을 사용하지 않은 실험결과

Table 1. The Result of Experiment without Momentum($\eta=0.1$)

	$\eta=0.1$ M=0.0	BP	Hybrid		
			$\alpha=0.001$	$\alpha=0.01$	$\alpha=0.02$
원	학습 시간	18m : 03s	18m : 48s	13m : 43s	13m : 32s
	학습 횟수	5184	5469	3990	3970
	자기 오인식	300(0)	300(0)	300(0)	300(0)
	오인식	896(4)	897(3)	897(3)	896(4)
	인식율	99.555	99.666	99.666	99.555
삼각형	학습 시간	2h : 36m : 23s	2h : 33m : 57s	38m : 36s	1h : 13m : 50s
	학습 횟수	30001	29698	7391	14156
	자기 오인식	397(3)	400(0)	400(0)	400(0)
	오인식	1185(15)	1187(13)	1187(13)	1181(19)
	인식율	98.75	98.916	98.916	98.416
사각형	학습 시간	1h : 43m : 34s	1h : 34m : 45s	36m : 33s	48m : 43s
	학습 횟수	30001	27354	10579	14099
	자기 오인식	299(1)	300(0)	300(0)	300(0)
	오인식	877(23)	876(24)	884(16)	867(33)
	인식율	97.444	97.333	98.222	96.333

은 학습에 참여하지 않은 데이터를 오인식한 결과이다.

오인식시 팔호안은 인식 못한 데이터 갯수, 팔호 밖에 인식한 데이터 갯수를 의미한다. 인식율은 학습에 참여하지 않은 데이터를 인식한 결과이며 소수점 4째 자리 이하는 생략하였다.

두 학습 법칙의 비교를 위해 학습속도에서는 학습 시간의 관점과 학습횟수의 관점인 두가지 측면에서 고려되었다. 신경망의 고유특징을 충분히 살리기 위해서는 신경망 전용하드웨어의 구현이 동반되어 신경망의 초대규모 병렬처리특징이 살려져야 하는데 이러한 관점에서 보면 학습횟수 비교가 더 바람직하므로 학습횟수에 대하여 서술한다.

원은 BP와 Hybrid가 전부 학습이 완료되었으며 학습횟수는 α 가 크면 BP보다 Hybrid가 현저히 적어진다. α 의 크기에 따라 학습 속도가 비례하며 0.02보다 더 큰 값에서도 수렴할 수 있음을 나타낸다. 즉 Hebbian의 특성이 더 강화되어질수록 수렴속도가 빨라질 수 있음을 나타낸다.

삼각형과 사각형 학습시 BP는 학습이 완료되지 않았지만 Hybrid는 학습이 되었으며 학습속도는 α 가 0.001일때 제일느리고 α 가 0.01일때 제일 빠르며 0.02일때는 중간이다.

즉 α 가 0.001에서부터 0.01까지는 수렴속도가 빨라지고 0.01보다 더 크면 수렴 속도가 느려질 것이며 매우 크면 BP와 마찬가지로 학습이 완료되지 않을 것이다. 즉 삼각형과 사각형은 BP에서도 수렴이 잘 안되지만 Hybrid에서는 α 가 0.01에서 최고의 수렴 속도가 되어 α 에 의한 수렴 영역($0.001 \leq \alpha \leq 0.02$) 즉 α 의 최대 크기가 제한된다.

BP와 제한된 학습법의 인식률을 고찰하면 원에서 BP가 99.555%(896/900)*100이고 Hybrid에서는 99.666%(0.001 \leq α \leq 0.01)이므로 제안된 학습법에 의한 인식율이 0.111% 더 우수하다. 삼각형에서는 BP는 98.75%이나 Hybrid는 98.916%(0.001 \leq α \leq 0.01)로서 0.166% 더 우수하다.

제안된 학습법에서 α 의 변화에 의한 전체적인 인식률은 $\alpha=0.001$ 일때 BP보다 우수하고 $\alpha=0.01$ 일때는 제일 우수하다.

그러나 사각형 학습시 α 가 0.02일때 학습은 되었지만 인식 못한 데이터가 BP보다 많은 것은 α 에 대한 일반화(Generalization)가 연구되어져야 함을 의미한다.

BP학습시의 성능이 $\eta=0.1$ 일때보다 우수하지 못

표 2. η 가 0.05 일때 모멘텀을 사용하지 않은 실험결과
Table 2. The Result of Experiment without Momentum ($\eta=0.05$)

	$\eta=0.05$ M=0.0	BP	Hybrid		
			$\alpha=0.001$	$\alpha=0.01$	$\alpha=0.02$
원	학습 시간	45m : 13s	42m : 48s	28m : 37s	12m : 38s
	학습 횟수	13039	12609	8473	3759
	자기 오인식	300(0)	300(0)	300(0)	300(0)
	오인식	895(5)	897(3)	895(5)	896(4)
	인식율	99.444	99.666	99.444	99.5555
삼각형	학습 시간	2h : 36m : 12s	1h : 34m : 42s	1h : 12m : 30s	54m : 45s
	학습 횟수	30001	18428	13976	10504
	자기 오인식	397(3)	400(0)	400(0)	400(0)
	오인식	1186(14)	1190(10)	1182(18)	1189(11)
	인식율	98.833	99.166	98.5	99.083
사각형	학습 시간	1h : 44m : 1s	1h : 31m : 49s	1h : 34m : 27s	1h : 40m : 2s
	학습 횟수	30001	26813	27320	29067
	자기 오인식	289(11)	300(0)	300(0)	300(0)
	오인식	874(26)	870(30)	871(29)	876(24)
	인식율	97.111	96.666	96.777	97.333

하지만 $\eta=0.05$ 일때의 성능을 비교하여 표2에 나타내었다.

표2의 수렴속도는 원에서 Hybrid($\alpha=0.001$)가 BP보다 빠르지만 α 가 클수록 매우 빠름을 알 수 있다. 즉 Hybrid의 수렴은 BP 수렴속도의 약 29%(3759/13039)*100 정도에서 수렴된다.

삼각형과 사각형에서 BP는 학습이 완료되지 않았지만 Hybrid는 완료되었고 삼각형에서는 α 가 클수록 학습속도가 빨라지지만 사각형에서는 α 가 클수록 학습속도가 느려진다.

즉 원과 삼각형은 α 의 크기에 비례하지만 사각형은 반비례한다. 사각형이 반비례하는 것은 수렴되지 않은 BP의 크기에 Hebbian의 미소 벡터량만 강화되면 더 우수한 수렴 속도를 가질 수 있음을 나타내고 0.02보다 더 커지면 수렴이 BP만큼 느려질 수 있음을 나타낸다.

BP보다 hybrid의 인식율은 원에서 최고 0.222%($\alpha=0.01$) 더 높고 삼각형은 0.333%($\alpha=0.001$) 사각형은 0.222%($\alpha=0.02$) 더 높다.

사각형에서 표에 제시하지 않았지만 $\alpha=0.03$ 에서 학습은 되지않았지만(자기오인식 1개) 8개의 오인식

밖에 없지만 BP에서는(자기오인식 11)26개 오인식이다.

표1과 표2에서 Hybrid의 특성은 η 가 적어지면 일반적으로 수렴 속도는 동일 α 의 범위에서 저하되지만 수렴 영역인 α 의 크기가 더 커질 수 있다.

즉, η 가 적어지면 α 를 크게 하여 수렴 속도를 보상할 수 있음을 나타낸다. 그러나 최고 인식율은 대체적으로 비슷하다.

원의 BP와 Hybrid($\eta=0.1, \alpha=0.01$) 학습 결과에서 매 Learning Epoch에 대한 출력층의 전체오차 E를 그림 2에 보였다.

BP와 Hybrid가 수렴되는 과정의 오차를 보기 위하여 오차의 크기를 LOG(E)로 나타내었다.

그림 2에서 학습횟수가 500일때부터 Hybrid가 BP보다 오차가 적어지며 전체적으로 오차가 적어지는 물이 BP보다 크다.

$\eta=0.1$ 일때 모우멘텀을 사용한 실험 결과는 표 3과 같다. 모우멘텀은 0.6이며 최대 학습횟수는 20000이며 다른 조건은 첫번째 실험과 같다.

표 3에서는 BP에서 원은 학습시 최소점(local minimum)에 빠져 B영역은 전체 학습이 되지 않았다.

표 3. η 가 0.1일때 모우멘텀을 사용한 실험 결과
Table 3. The Result of Experiment with Momentum($\eta=0.1$)

	$\eta=0.1$ M=0.6	BP	Hybrid		
			$\alpha=0.001$	$\alpha=0.01$	$\alpha=0.02$
원	학습 시간	1h : 10m : 47	10m : 11s	6m : 14s	3m : 48s
	학습 횟수	20001	2893	1771	1068
	자기 오인식	200(100)	300(0)	300(0)	300(0)
	오인식	598(302)	896(4)	898(2)	897(3)
	인식율	66.444	99.555	99.777	99.666
삼각형	학습 시간	1h : 47m : 57s	1h : 46m : 33s	22m : 19s	40m : 39s
	학습 횟수	20001	20001	4161	7445
	자기 오인식	399(1)	399(1)	400(0)	400(0)
	오인식	1188(12)	1196(4)	11888(12)	1188(12)
	인식율	99	99.666	99	99
사각형	학습 시간	1h : 11m : 7s	1h : 10m : 45	1h : 12s	22m : 19s
	학습 횟수	20001	20001	16914	6288
	자기 오인식	299(1)	299(1)	300(0)	300(0)
	오인식	877(23)	877(23)	879(21)	876(24)
	인식율	97.444	97.444	97.666	97.333

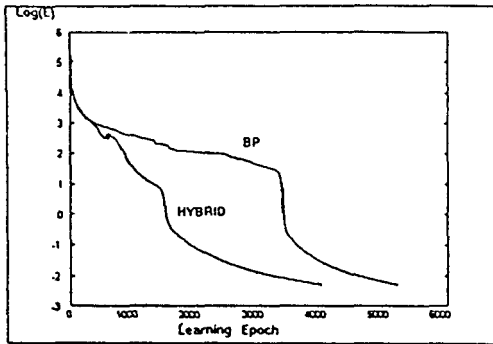


그림 2. Learning Epoch에 대한 BP, Hybrid의 학습 오차
Fig 2. The BP and the Hybrid Learning Error on Learning Epoch.

Hybrid 학습시 α 가 크면 학습속도가 크게 개선되고 본 연구에 제시되지 않았지만 α 가 0.1일 때는 학습횟수가 560이고 수렴속도가 1분 60초나되고 인식률은 α 가 0.02와 같았다.

또, 삼각형과 사각형은 BP뿐만 아니라 Hybrid($\alpha = 0.001$)가 학습이 완료되지 않았고 수렴속도도 BP와 비슷하다. 삼각형은 α 가 0.01일때의 학습속도가 0.02보다 빠르지만 사각형은 α 가 0.01일때보다 0.02일때가 매우 빠르다.

원과 사각형은 학습 속도가 α 의 크기에 반비례해서 빨라지며 삼각형에서는 α 가 0.001부터 0.01까지 수렴속도가 개선되지만 0.01부터는 수렴 속도가 느려진다. BP보다 Hybrid의 인식률은 삼각형에서 최고 0.666%($\alpha = 0.001$) 사각형에서는 0.222%($\alpha = 0.01$) 더 높다.

Hybrid 학습법도 BP와 마찬가지로 모우멘텀항을 사용하면 성능이 향상됨을 알 수 있다.

각 표들에서 원과 삼각형 및 사각형 문제를 Hybrid와 BP의 성능을 비교하면 다음과 같다.

첫째, 원에서 BP의 최고 학습의 수는 5184이고 Hybrid는 모우멘텀항 결과에서 1068($\alpha = 0.02$)이다. 즉 BP 수렴속도의 20% 정도에서 수렴된다. 인식률은 BP는 η 가 0.1에서 99.555%이고 Hybrid는 모우멘텀항 결과에서 99.777%($\alpha = 0.01$)이므로 제한된 학습법이 0.222% 더 우수하다.

둘째, 삼각형에서 BP는 수렴되지 않았지만 Hybrid는 모우멘텀항 결과에서 4161($\alpha = 0.01$)이다. 인식률은 BP는 모우멘텀항 결과에서 99%이고 Hybrid는 모우멘텀항 결과에서 99.666%($\alpha = 0.001$)이므로 0.

666% 더 우수하다.

셋째, 사각형에서 BP는 수렴되지 않았지만 Hybrid는 모우멘텀항 결과에서 6288($\alpha = 0.02$)이다. 인식률은 BP는 모우멘텀항 결과에서 97.444%이고 Hybrid는 η 가 0.1일때 98.222%($\alpha = 0.01$)이므로 0.778% 더 우수하다.

Hybrid의 단점은 첫째, 수렴속도의 빠르기와 인식률이 비례하지 않는다. 즉, η 가 0.1일때 원에서 수렴속도가 0.02일때 더 우수하지만 인식률은 0.001이나 0.01이 더 우수하다.

둘째, 모든 패턴의 수렴에 적용될 수 있는 α 의 크기가 제한된다. 즉, 원, 삼각형, 사각형 문제에서 공통으로 수렴될 수 있는 최소 영역이 $0 < \alpha \leq 0.02$ 이다. 그러나 BP로 수렴이 잘 되거나 η 가 적어질때는 α 의 크기(α 의 영역)가 커진다.

IV. 결 론

본 연구에서는 Hebb 학습법칙과 오차역전파 기법을 동시에 적용하는 새로운 Hybrid 학습법을 제안하였다.

몬테 카를로 방법으로 제안된 학습법을 다층구조신경망에 적용하여 컴퓨터 시뮬레이션 한 결과를 BP 학습법으로 시뮬레이션 한 결과와 비교하는 실험을 하여 보았다.

BP 알고리즘이 적용된 기존의 MLP에 대해 동일 네트워크 구조에 Hybrid 학습법칙을 적용하였을 때는 학습속도는 매우 큰 개선이 이루어졌고 인식율이 소폭이지만 개선되었다.

이것은 Hebb 법칙에 의한 입력의 특징 추출기능이 강화되어 매우 빠른 학습속도와 소폭이지만 인식률의 향상을 보였다고 생각된다.

특히, Hebb 학습법칙에 도입되는 학습률 α 가 적은 값일때는 BP와 성능의 차이가 별로 없었는데 이것은 α 가 아주 적은 경우에는 BP의 특성이 지배적이기 때문이다.

앞으로 Hybrid 알고리즘의 일반적 특성에 관한 연구와 여러 분야에서 응용이 기대된다.

참 고 문 헌

1. 송도선, 김석동, 이행세, "Spatio-temporal 방법을 이용한 우리말 모음 인식에 관한 연구," 한국 음향학회지 제

12권 제4호, 1993.

2. K. Fukushima, S.Miyake, T.Ito, & T.Kouno, Handwritten numeral recognition by the algorithm of the neocognitron, An experimental system using a microcomputer(in Japanese) Transactions of the information Processing Society of Japan, 28(6). pp. 627-635, 1987.
3. D.E.Rumelhark, G.E.Hinton, and R.J.Williams, Learning Internal Representation by Error Propagation in Parallel Distributed Processing(Eds. Davud E.Rumelhark and Jame L.McClland), The MIT Press, vol.1. 1986.
4. R.Linsker, "From basic network principles to neural architecture : Emergence of spatial-opponent cells," Proceedings of the National Academy of Sciences USA, 83, 1986.
5. D.O.Hebb, The Organization of Behavior, Introduction and Chapter 4, The first stage of Perception growth of the assembly, pp. xi-xix, 60-78, New York : Wiley, 1949.
6. J.J.Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of National academic science, Vol.79, pp.2554-2588, biophysics, April 1982.
7. J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," Proceedings of Mational Academic Science, Vol. 81, pp.3088-3092, Biophysics, May 1984.
8. Terence David Sanger, Optimal unsupervised learning in a single-layer linear feedforward neural network. Neural Networks, 2 : 459-473, 1989.
9. Terence David Sanger, Optimal unsupervised learning in feedforward neural networks. MIT AI Lab Tech, Report 1086. 1989.
10. T. Kohonen, Self-organization and Associative Memory. Springer-Verlag, Berlin, 1984.
11. T. J. Sejnowski and C. R. Rosenberg, Net talk : A parallel network that learns to read aloud. Johns Hopkins University EE & CC Technical Report JHU/EECS-86/01, Jan. 1986.
12. Richard J. Wood and Michale A Gennert, "A Neural Network that uses a Hebbian/Backpropagation Hybrid Learning Rule," IJCNN, Boltimore, Maryland June 7-11, 1992.

▲송도선(正會員)

1956年 4月 14日生



1977年 : 영남대 전자공학과 졸업
 1981年 : 고려대학 대학원 전자공학과 졸업
 1990年 : 아주대학교 박사과정 수료
 1981年 ~ 현재 : 중경공업전문대학 전자계산기과 부교수
 ※주관심분야 : 인공지능 및 신호처리

▲金錫東

1957년 8월 23일생



1982년 2월 : 아주대학교 전자공학과(공학사)
 1984년 2월 : 아주대학교 대학원 전자공학과(공학석사)
 1993년 2월 : 아주대학교 대학원 전자공학과(공학박사)

1984년 3월 ~ 현재 : 호서대학교 전자계산학과 부교수
 ※주관심분야 : 음성인식, 신경회로망, 음성신호처리

▲李幸世(正會員)

1943年 8月 29日生



1966年 : 전북대학교 전기공학과 공학사
 1972年 : 서울대학교 전자공학과 공학석사
 1984年 : 고려대학교 전자공학과 공학박사
 1968年 ~ 1970年 : 해군사관학교 전자공학 교관

1973年 ~ 현재 : 아주대학교 교수

1982年 ~ 1983年 : 미국 Columbia Univ.(N.Y.) 객원교수

1987年 ~ 1988年 : 프랑스 INRIA(Paris) 객원교수

1992年 ~ 현재 : 거제전문대학 학장

1973年 ~ 현재 : IEEE 회원

※주관심분야 : 문자 및 음성인식, 인간 기계 인터페이스, 인공지능, 신경회로망, VLSI 회로, 디지털 신호처리 등임