

적응 역 전파 신경회로망의 초기 연결강도 설정에 관한 연구

홍봉화*

요 약

본 논문에서는 역 전파 신경회로망의 학습파라미터를 발생한 오차에 따라서 유동성 있게 갱신할 수 있고 이 학습알고리즘의 효율을 향상시킬 수 있는 초기연결강도 설정 방법을 제안하였다. 제안한 알고리즘은 국소 점을 벗어날 수 있는 것으로 기대되고, 수렴환경에 알맞은 초기 연결강도 발생을 설정할 수 있다. 모의실험에서는 세 가지의 학습패턴을 가지고 실험하였다. 첫 번째 3-패리티 문제에 대한 학습을 수행하였고, 두 번째는 7x5 알파벳 영문자 폰트에 에 대한 학습이고 세 번째는 필기체 숫자 및 한글의 기본 획에 적용하였다. 모의실험결과, 제안된 방법은 기존의 표준 역 전파 알고리즘에 비하여 약 27%~57.2정도 학습효율이 향상됨을 고찰하였다.

1. 서론

오늘날 반도체 기술의 발달로 인하여 신경회로망의 구현 가능성이 높아짐에 따라서 신경회로망에 대한 연구가 활발히 진행되고 있다. 신경회로망의 알고리즘 중에서 일반적으로 이용되는 역 전파 알고리즘은 반복적인 학습을 수행시, 연결강도의 갱신과정에 있어 발생한 오차에 상관없이 고정된 상수 값을 학습파라미터로 하여 연결강도를 일률적으로 조정하는 불합리성을 지니고 있다.

역 전파 알고리즘의 학습과정에서 학습 파라미터를 부적절하게 선택할 경우, 필요이상의 반복횟수를 요하는 문제점과 종종 국소적인 해로 빠지는 문제 등이 야기된다.[1][2][3]

역 전파 신경회로망의 과대한 학습시간의 소요문제 및 국소 해로 빠지는 문제점을 해결하고

학습 알고리즘을 최적화 하여 실시간으로 처리를 수행할 수 있는 신경회로망에 대한 연구가 요구된다. 또한, 신경회로망의 초기 연결강도는 임의 실험 값으로 여러 번의 시도에 의해 결정한다. 이러한 문제를 해결할 수 있도록 하기 위하여 동적으로 초기 연결강도를 설정할 수 있는 연구가 요구된다.

따라서, 본 논문에서는 역 전파 알고리즘의 학습수행 시, 연결강도 갱신과정에 영향을 미치는 학습계수와 관성 계수 값을 고정된 상수 값이 아닌 유동적인 값을 가질 수 있도록 지수함수(Exponential Function)의 성질을 도입하여 발생한 오차에 따라서 학습계수와 관성계수를 유동적으로 변경시킴으로써 반복횟수를 줄일 수 있고 안정성을 향상시킬 수 있는 학습 알고리즘[9]과 이 알고리즘의 학습 효율을 향상시킬 수 있는 초기 연결강도 설정 알고리즘을 제안하였다.

제안된 알고리즘의 효과를 검증하기 위하여 3-패리티 문제, 7x5 알파벳 영문자폰트, 필기체

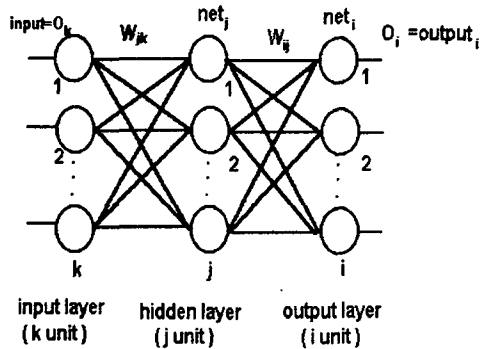
* 세명대학교 컴퓨터수리정보학과 교수

숫자 및 한글의 기본 획의 학습에 적용하였다.[3][4][5][6]

II. 역 전파 알고리즘

역 전파 학습이론은 미분의 반복규칙을 여러 번 반복 적용하여 확률 근사치를 유도해내는 알고리즘으로서 연결강도 공간의 결합 점에서 오차의 표면이 기울기를 따라가는 방법이며, 일반화된 기울기 감소법(Generalized Delta Rule)이라고도 한다.[1][2][3]

역 전파 신경회로망은 그림 1에 나타낸 바와 같이 입력, 은닉 및 출력 층으로 구성된 계층형 구조를 지닌 다층 신경회로망으로 1986년 Rumelhart가 교사 있는 학습 이론인 역 전파 학습 알고리즘을 적용하여 제안한 신경회로망이다.[1][2][3]



(그림 1) 역 전파 신경회로망
(Fig. 1) The structure of back propagation neural networks

그림 1의 학습 층에 있어서 \$k\$ 번째 층의 노드

를 \$p^k\$, 노드 수를 \$L\$이라 하면 \$k+1\$번째 층의 노드 입력 \$u_p^{k+1}\$는 \$k\$ 번째 층 시그모이드 함수 \$a_p^k\$의 출력과 연결강도 \$w_p^k\$의 내적으로서 구하며, 그 과정은 다음과 같이 기술 할 수 있다.

$$U_p^{k+1} = \sum_{p=1}^L W_p^k a_p^k \quad (1)$$

식(1)에서 \$u_p^{k+1}\$는 \$k\$층의 \$p^k\$번째 노드와 \$K+1\$층의 \$p^{k+1}\$번째 노드사이의 가중치 \$w_p^k\$에 전단의 출력 \$a_p^k\$를 내적 한 값이다. 노드 내부에 있어서 시그모이드 함수에 의해 출력되는 값은 식 (2)와 같이 기술할 수 있다.

$$a_p^k = \frac{1}{1 + \exp(-U_p^k)} \quad (2)$$

역 전파 신경회로망은 실제출력 \$a_i\$와 목표출력 \$T_i\$사이 평균출력오차의 총합으로 정의된 에너지를 최소화하기 위하여 모든 연결 강도를 갱신한다.

$$E = \frac{1}{2} \sum_{i=1}^n (T_i - a_i^L)^2 \quad (3)$$

연결강도의 갱신 문제는 입력과 훈련패턴의 집합을 기준으로 하여 식 (3)의 오차 평가함수 \$E\$의 크기에 따라 연결강도를 갱신하며, 연결강도의 갱신 법칙은 식 (4)와 같다.

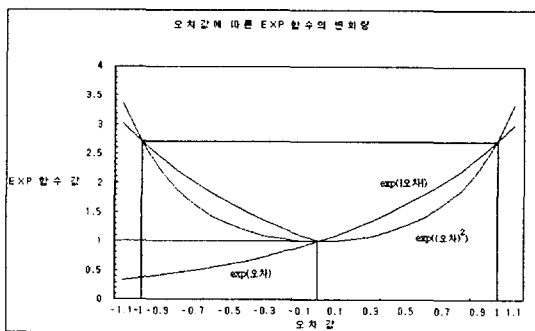
$$\begin{aligned} \Delta W_p^{k+1}(t+1) &= \eta \delta_p^{k+1} a_p^{k-1} + \alpha \Delta W_p^k(t) \\ \delta_p^k &= \sum_{i=1}^L \delta_p^{k+1} W_p^{k+1} \quad : \text{ 학습층} \\ \delta_p^L &= a_p^L - T_p \quad : \text{ 출력층} \end{aligned} \quad (4)$$

식 (4)에서 t는 반복횟수를 나타내며, 관성계수(α), 학습계수(η)는 고정된 상수 값으로서 1회의 학습에 따른 변화량을 계산하기 위하여 이용된다.

역 전파 학습 알고리즘은, 학습 시 반복횟수의 증가로 인한 수렴속도의 저하문제가 야기되므로, 발생한 오차 값에 비례하여 학습파라미터를 유동적으로 갱신함으로써 학습효율과 수렴속도의 향상을 기할 수 있다.

III. 적응 학습 알고리즘

식 (4)의 학습계수와 관성계수는 연결강도의 갱신과정에 중요한 역할을 한다. 그러나 기존에는 이들 파라미터 값을 고정시킴으로서 오차 값이 크던 작던 간에 일정한 고정된 비율에 의하여 연결강도가 갱신되는 불합리성을 지니고 있다.^{[1][2][3]}



(그림 2) 오차에 따른 지수함수의 성질
(Fig. 2) The property of exponential function for the error

본 논문에서는 그림 2와 같은 지수(Expo-

ponential)함수의 성질을 이용하여 연결강도의 갱신에 중요한 역할을 담당하는 학습계수와 관성계수를 각각의 오차 값에 따라서 갱신할 수 있는 알고리즘을 다음과 같이 제안하였다.

신경회로망에 입력이 주어질 경우, 전 방향 처리과정을 거쳐 출력 층에서 식 (5)와 같이 오차를 산출하며, 식 (3)의 에너지를 최소화하기 위하여 학습과정을 반복한다. 이 경우, 발생된 오차는 양의 오차와 음의 오차로 표현된다. 신경회로망에 발생한 오차를 지수 함수적으로 증가 또는 감소 함수형태로 나타내면 식 (5)와 같다.

$$h(\delta_{ij}) = A \exp(\delta_{ij}) \tag{5}$$

그러나, 식 (5)에 있어 오차가 음으로 증가될 경우, "0"로 근접하는 문제가 발생한다.

따라서, 음의 오차를 양의 오차가 갖는 성질로 하기 위한 방법으로 식 (6)과 같이 오차의 절대치를 이용하거나 오차를 제곱하여 나타낼 수 있다.

$$\left(\begin{array}{l} h(\delta_{ij}^2) = A \exp(\delta_{ij}^2) \quad \delta_{ij} > 1, \delta_{ij} < -1 \\ h(|\delta_{ij}|) = A \exp(|\delta_{ij}|) \quad -1 \leq \delta_{ij} \leq 1 \\ A : \text{상수} \end{array} \right) \tag{6}$$

식 (6)에서 발생한 오차에 의한 두 식의 크기를 비교하면 표1과 같다.

<표 1> 오차에 따른 절대치 함수 값과 오차 값을 제공한 함수 값의 비교

<Table 1> Comparing Absolute value and square value of error

발생한 오차(δ)	대수 비교
$-1 < \delta < 1$	$E_{\text{pow}} < E_{\text{abs}}$
$\delta = \pm 1$	$E_{\text{pow}} = E_{\text{abs}}$
$\delta > 1, \delta < -1$	$E_{\text{pow}} > E_{\text{abs}}$

표 1에서 알 수 있듯이 발생한 오차가 1보다 크거나 -1보다 작을 경우, 오차의 제곱에 대한 지수함수 값이 오차의 절대치 함수 값보다 클수록 연결강도의 갱신시 연결강도의 변동량이 커진다. 역으로, 오차가 -1과 1사이의 값을 가질 경우에는 오차의 제곱에 대한 지수함수 값이 오차의 절대치에 대한 지수함수 값보다 작게 된다. 따라서, 연결강도의 갱신시, 연결강도의 변동폭을 작게 함으로써 시스템의 안정도를 향상시킬 수 있으므로 오차의 제곱에 대한 지수함수 값을 학습파라미터로 이용하는 것이 더욱 효율적이다. 식(6)을 학습계수에 적용하여 기술하면 식 (7),(8)과 같다.

$$\eta_{ij}(t+1) = \eta_{ij}(t) + \Delta \eta_{ij}(t) \quad (7)$$

$$\begin{aligned} \Delta \eta_{ij}(n) &= \begin{pmatrix} h(\delta_{ij}^2) & \delta_{ij} > 1, \delta_{ij} < -1 \\ h(|\delta_{ij}|) & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \\ &= A \begin{pmatrix} \exp(\delta_{ij}^2) & \delta_{ij} > 1, \delta_{ij} < -1 \\ \exp(|\delta_{ij}|) & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \end{aligned} \quad (8)$$

또한, 식 (4)에서 학습계수와 관성계수는 반비례 관계가 성립되므로 관성계수에 대응된 관계식은 다음과 같이 나타낼 수 있다.

$$\alpha_{ij}(t+1) = \alpha_{ij}(t) + \Delta \alpha_{ij}(t) \quad (9)$$

$$\begin{aligned} \Delta \alpha_{ij}(t) &= \begin{pmatrix} \frac{1}{h(\delta_{ij}^2)} & \delta_{ij} > 1, \delta_{ij} < -1 \\ \frac{1}{h(|\delta_{ij}|)} & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \\ &= B \begin{pmatrix} \frac{1}{\exp(\delta_{ij}^2)} & \delta_{ij} > 1, \delta_{ij} < -1 \\ \frac{1}{\exp(|\delta_{ij}|)} & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \end{aligned} \quad (10)$$

따라서, 역 전파 신경회로망의 연결강도 갱신 과정은 식 (7), (8),(9),(10)를 식(4)에 대입하여 식(11)과 같이 나타낼 수 있다.

$$\begin{aligned} \Delta W_{ij}(t+1) &= a_i(t) \eta_j(t) \delta_{ij} + \alpha_{ij}(t) W_{ij}(t) \\ &= \begin{pmatrix} h(\delta_{ij}^2) \delta_{ij} + \frac{1}{h(\delta_{ij}^2)} W_{ij}(t), & \delta_{ij} > 1, \delta_{ij} < -1 \\ h(|\delta_{ij}|) \delta_{ij} + \frac{1}{h(|\delta_{ij}|)} W_{ij}(t), & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \\ &= \begin{pmatrix} A \exp(\delta_{ij}^2) \delta_{ij} + \frac{1}{A \exp(\delta_{ij}^2)} W_{ij}(t), & \delta_{ij} > 1, \delta_{ij} < -1 \\ A \exp(|\delta_{ij}|) \delta_{ij} + \frac{1}{A \exp(|\delta_{ij}|)} W_{ij}(t), & -1 \leq \delta_{ij} \leq 1 \end{pmatrix} \end{aligned} \quad (11)$$

식 (11)을 이용하여 기존의 역 전파 신경회로망에 비하여 20~40%정도 학습효율의 향상을 시킬 수 있다. 또한, 본 논문에서는 역 전파 신경회로망에 영향을 미치는 중요한 요소인 초기 연결강도를 유동적으로 설정하고자 한다.

IV 초기 연결강도의 설정

초기 연결강도의 설정은 역 전파 신경회로망의 학습에 있어 최종적인 수렴상태를 결정하는 중요한 요소이다. 그러나, 기존의 방법들은 지도 학습의 본질인 교사신호와 무관하게 실험 값으로 설정하여 학습을 수행함으로써 학습효율 및 일반화가 저하되는 문제점을 갖는다.

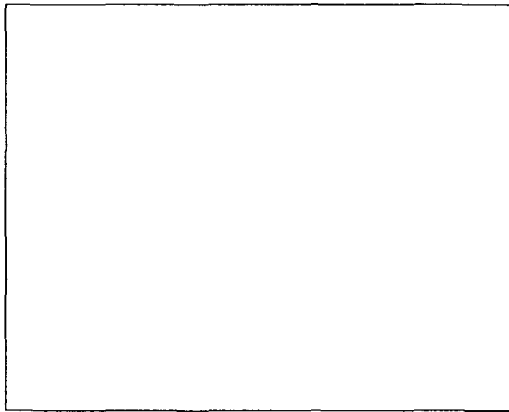
따라서, 본 논문에서는 역 전파 학습알고리즘의 효율성 및 일반화를 향상시키기 위하여 그림 3에 나타낸 바와 같이 교사신호 차이를 초기 연결강도의 발생함수로 설정할 수 있는 방법을 제안하였다.

그림 3에서 초기연결 강도를 W_{init} 라 하면 역 전파 신경회로망의 초기 연결강도의 발생 범위는 식 (12)와 같이 나타낼 수 있다.

$$W_{init} = F(\text{Range}), \quad -W \leq \text{Range} \leq W \quad (12)$$

$$(W = \text{Target}_{max} - \text{Target}_{min})$$

식 (12)에서 함수 F()는 연결강도 발생함수를 나타내며, "Range" 범위내의 난수 발생함수이다. 이 "Range"는 교사신호의 최대 값인 Target_{max}와 최소 값인 Target_{min}의 차이(W)를 나타낸다.



(그림 3) 초기 연결강도 설정을 위한 역 전파 신경회로망의 구조

(Fig. 3) The structure of back propagation neural networks for the configuration of initial weight

V. 모의실험 및 고찰

본 논문에서 제안한 알고리즘의 타당성을 검증하기 위하여 3 종류의 입력 패턴을 가지고 모의실험을 수행하였다.

첫 번째는 3-패리티(Parity)문제에 대하여 학습을 수행하였고 두 번째로는 보다 복잡한 예로

써 7x5 영문자 알파벳 폰트(alphabetic font)의 학습을 수행하였다. 끝으로, 한글 및 숫자 온라인 필기체 기본 획 14자의 분류를 수행하였다.

비교를 위하여 표준적인 역 전파 알고리즘(SBP: Standard Back Propagation)과 기존에 제안된 방법(EPBP: Exponential Parameter Back Propagation)과의 반복학습 횟수를 비교 분석하였다.

실험에 사용한 역 전파 신경회로망의 구조는 입력, 은닉 그리고 출력 층을 갖는 3층 구조를 갖는 신경회로망을 사용하였으며, 초기 연결강도는 교사 신호의 최대 값과 최소 값의 차이를 초기연결강도 범위로 설정하여 이산 적으로 발생하였다.

실험을 위한 제한조건으로 최대수렴오차를 10⁻⁵로 하였고 최대 반복학습 횟수로 배타적 논리합과 3-패리티 문제는 5000번으로, 7×5 영문자 폰트의 학습과 온라인 필기체 한글 및 숫자 기본 획의 학습은 32000번으로 제한하여 실험하였다.

본 논문에서는 역 전파 신경회로망의 학습효율을 향상시키기 위하여 초기연결강도발생 범위를 목표벡터의 최대 값과 최소 값의 차이로 설정하였을 경우, 표준 역 전파 알고리즘과 본 논문에서 제안한 알고리즘과의 비교를 통하여 학습의 효율을 검증하였다.

표 2는 3-패리티 문제의 학습에 대하여 표준 역 전파 알고리즘과 본 논문에서 제안한 방법과의 반복학습횟수의 비교를 나타낸다.

표 2에서 SBP(Standard Back-Propagation)는 표준 역 전파 알고리즘을 나타내며, SBP+WG (Weighted Generating)는 표준 역 전파 알고리즘에 초기연결강도 발생 방법을 부가한 경우를 나타낸다.

〈표 2〉 3-패리티 문제의 학습에 대한 반복학습
횟수의 비교

〈Table 2〉 The Comparison of the number
of learning iteration for the
3-parity problem

SBP	SBP+WG	EPBP	EPBP+WG
1848회	1514회	1192회	1097회

EPBP(Exponential Parameter Back-Propagation)는 역 전파 신경회로망에서 발생한 오차를 지수 함수화 하여 유동적으로 학습파라미터를 갱신할 수 있는 알고리즘을 의미하고 EPBP+WG는 EPPB에 초기연결강도 발생방법을 적용한 경우를 의미한다.

표 2에서 알 수 있듯이 표준 역 전파 알고리즘(SBP)은 초기 연결강도 설정 알고리즘을 적용하였을 경우, 약 18.1%의 학습효율이 향상되었고 본 논문에서 제안된 알고리즘의 경우에도 약 8%~16.1%정도 학습효율이 향상됨을 고찰하였다.

표 3은 영문자 알파벳 폰트 26문자에 대하여 초기연결강도 설정알고리즘을 적용하였을 경우, 본 논문에서 제안된 알고리즘과 표준 역 전파 알고리즘의 반복학습횟수의 비교를 나타낸다.

〈표 3〉 7×5 알파벳 폰트의 학습에 대한 반복
학습 횟수의 비교

〈Table 3〉 The Comparison of the number
of learning iteration for 7 by 5
alphabet font

SBP	SBP+WG	EPBP	EPBP+WG
26069회	23467회	19209회	19029회

표 3에서 알 수 있듯이 알파벳 영문자 폰트 26자의 학습에 있어 초기 연결강도 설정 알고리

즘을 적용하였을 경우, 표준 역 전파 알고리즘(SBP)은 약 10%정도 학습효율이 향상되었고, 본 논문에서 제안한 알고리즘의 경우에는 약 0.9%~38%정도 학습효율이 향상됨을 고찰하였다. 표 4는 필기체 한글 및 숫자의 기본 획 14자에 대하여 초기연결강도 알고리즘을 적용하였을 경우, 표준 역 전파 알고리즘과 본 논문에서 제안한 알고리즘의 반복학습횟수의 비교를 나타낸다.

〈표 4〉 필기체 한글 및 숫자 기본 획의 학습에
대한 반복학습 횟수의 비교

〈Table 4〉 The Comparison of the number
of learning iteration for the
primitive strokes

SBP	SBP+WG	EPBP	EPBP+WG
30162	25173	18520	12908

표 4에서 알 수 있듯이 필기체 한글 및 숫자의 경우의 학습에 있어 3-패리티 문제와 알파벳 영문자 26자의 학습의 경우와 마찬가지로 초기연결강도 설정알고리즘을 적용하였을 경우, 표준 역 전파 알고리즘은 약 16.5%정도 학습효율이 향상되었고 본 논문에서 제안한 알고리즘의 경우에는 약 24.1%~30.3%정도 학습효율이 향상됨을 고찰하였다. 또한, 분류 율 면에서도 표 5에 나타낸바와 같이 표준 역 전파 알고리즘(SBP)은 93.63%의 분류 율을 나타내었고 SBP에 초기연결강도 설정 알고리즘을 적용한 방법(SBP+WG)은 약 95.8%의 분류 율을 나타내었다. 또한 본 논문에서 제안한 지수함수를 학습 파라미터로 이용한 역 전파 알고리즘(EPBP)은 97.7%의 분류 율을 나타내었고, EPBP에 초기연결강도 설정알고리즘을 적용한 경우(EPBP+WG)는 98.5%의 높은 분류 율을 나타냄을 고찰하였다.

〈표 5〉 초기연결강도 설정알고리즘을 적용한 경우
 우에 대한 필기체 기본 획의 분류율 비교
 〈Table 5〉 The comparison of classification rate in the case of the apply to configurative algorithm of the initial weight

알고리즘	SBP	SBP+WG	EPBP	EPBP+WG
분류율	93.63%	95.8%	97.7%	98.5%

위의 결과로부터 입력패턴이 이진 패턴이 아니라 실수 값을 갖는 경우, 초기연결강도 설정 알고리즘은 더욱더 효율이 우수함을 고찰하였다.

IV. 결론

본 논문에서는 지도학습 신경회로망의 대표적인 학습알고리즘인 역 전파 알고리즘에 있어, 부적절한 학습파라미터 및 초기연결강도 설정에 의한 지연현상을 줄이고 시스템의 안정도를 향상시킬 수 있는 학습파라미터의 설정방법 및 초기 연결강도 설정방법을 제안하였다.

제안된 알고리즘을 검증하기 위하여 3-패리티 문제, 알파벳 영문자 및 필기체 한글의 기본 획 14자에 적용한 결과 표 2, 표 3, 표 4 및 표 5와 같은 결과를 얻었다.

SBP(Standard Back- Propagation)+WG(Weighted Generating)의 경우, SBP에 비하여 약 10%~18%정도 반복학습 횟수의 감소와 2.7%의 분류 효율이 향상되었고, 지수함수를 학습파라미터로 적용한 알고리즘에 초기연결강도 설정방법을 적용한 경우(EPBP+WG)는 EPBP에 비하여 약 0.9%~33.3%정도의 반복학습 횟수의 감소와 0.8%정도 분류효율이 향상됨을 고찰하였다.

앞으로의 연구 방향은 본 논문에서 제안한 학습방법을 이용하여 온라인 필기체 영상처리, 문자인식 및 패턴인식 등의 응용분야에서 응용될 수 있도록 실험데이터의 확장과 하드웨어의 구현 등이 이루어져야 할 것이다.

참고문헌

- [1] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. *Learning representations by error propagation*. In D. E. 1986.
- [2] Rumelhart, J. L. McClelland and the PDP Reserch Group(Eds.), *parallel distributed processing Vol.1*, pp.318~536. Cambridge, MA : MIT Press.
- [3] Rumelhart, D. E., Hinton, G.E., & Williams, R. J.. *Learning representation by back propagation errors*. *Nature*, 323, 533~536, 1986
- [4] Sejnowski, T. J. & Rosenberg, C. R. *Parallel networks that learn to pronounce English text*. *Complex Systems*, 1, 14 5~168, 1987.
- [5] Yoh-Han Pao, *Adative Pattern Recognition and Neural Networks*. Addison Wesley pp.113~139, 1989.
- [6] James A. Freeman, David M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques*. Addison Wesley pp.89~125, 1991.
- [7] Jacek M. Zurada, *Introduction to Artificial Neural System*. West, pp.163~248, 1992.

- [8] 홍봉화, 탁형근, 조원경 “역전과 학습알고리즘의 최적화를 위한 파라미터에 관한 연구”, 대한 전자공학회 추계종합 학술대회 논문집, vol.15, no.1, pp.725~727 1992.
- [9] 홍봉화, 이승주, 조원경 “역전과 신경회로망의 수렴속도 개선을 위한 학습파라미터 설정에 관한 연구”, 대한 전자 공학회 논문지, 제33권 B편 제11호, pp.159~165, 1996

On the Configuration of initial weight value for the Adaptive back propagation neural network

Bong-Wha, Hong*

Abstract

This paper presents an adaptive back propagation algorithm that update the learning parameter by the generated error, adaptively and configuration of the range for the initial connecting weight according to the different maximum target value from minimum target value.

This algorithm is expected to escaping from the local minimum and make the best environment for the convergence. On the simulation tested this algorithm on three learning pattern. The first was 3-parity problem learning, the second was 7×5 dot alphabetic font learning and the third was handwritten primitive strokes learning. In three examples, the probability of becoming trapped in local minimum was reduce. Furthermore, in the alphabetic font and handwritten primitive strokes learning, the neural network enhanced to learning efficient about 27%~57.2% for the standard back propagation(SBP).

* Semyung Univ. Dept. of Computer Aided Mathematical Information Science, processor