

論文95-32B-1-20

일정 학습계수와 이진 강화함수를 가진 자기 조직화 형상지도 신경회로망

(Self-Organizing Feature Map with Constant Learning Rate and Binary Reinforcement)

趙成元*, 石鎮旭*
(Seongwon Cho, Jinwuk Seok)

요약

이진 강화함수와 일정 적응 이득을 사용한 수정된 Kohonen의 자기조직화 형상지도 알고리즘을 제안한다. Kohonen의 자기조직화 형상지도 알고리즘에서 사용되는 시간에 비례하여 단조감소하는 적응 이득 대신 일정한 값의 적응 이득을 사용하고, 일정 적응이득에 의해 발생하는 국소 최소점을 보상하기 위해 이진 강화 함수를 결합하였다. 이 알고리즘에서는 일정 적응이득을 사용함에 따라 곱셈연산이 불필요하므로 디지털 하드웨어 구성에 보다 용이하다. 제안한 알고리즘은 8 Class데이터에 대해 기존의 알고리즘과 성능비교를 하였으며, 알고리즘의 약 수렴성을 논하였다.

Abstract

A modified Kohonen's self-organizing feature map (SOFM) algorithm which has binary reinforcement function and a constant learning rate is proposed. In contrast to the time-varying adaptation gain of the original Kohonen's SOFM algorithm, the proposed algorithm uses a constant adaptation gain, and adds a binary reinforcement function in order to compensate for the lowered learning ability of SOFM due to the constant learning rate. Since the proposed algorithm does not have the complicated multiplication, it's digital hardware implementation is much easier than that of the original SOFM.

I. 서론

신경회로망의 하드웨어 구현은 그 가능성과 대규모 병렬처리라는 이점에도 불구하고 일반적이고도 상용

적인 수준에까지는 이르지 못했다. 그 이유는 신경회로망의 각 알고리즘이 분명 대규모 병렬처리의 특징을 가지고 있음에도 불구하고 뉴런 하나가 수행 해야하는 학습 알고리즘 연산의 복잡성에 기인한다.

현재까지 신경회로망의 하드웨어 구현의 경우 이와 같은 문제점을 극복하기 위해 대부분 아날로그 방식 혹은 Stochastic Pulse Stream방식으로 이루어져 왔다. 하지만 아날로그 방식의 하드웨어에서는 CMOS 내 PMOS 트랜지스터와 NMOS 트랜지스터의 특성

* 正會員, 弘益大學校 電氣制御工學科
(Dept. of Electrical and Control Eng., Hong Ik University)
接受日字 : 1994年 8月 24日

불일치, 이상적인 아날로그 메모리 소자 구현의 어려움, A/D, D/A 변환기 내장시 이로인한 정확도 손실등의 문제점이 있고, Stochastic Pulse Stream의 경우 충분한 정확도를 얻기 위해서는 연산에 필요한 시간이 지나치게 길어진다는 것과 데이터의 동기 확보가 역시 문제점으로 지적된다. 또한 디지털 방식의 경우 연산에 필요한 곱셈기나 나눗셈기가 필요하다는 문제점 때문에 뉴런의 집적도가 아날로그 방식이나 Stochastic Pulse Stream 방식보다 현저하게 떨어진다는 단점이 있다.

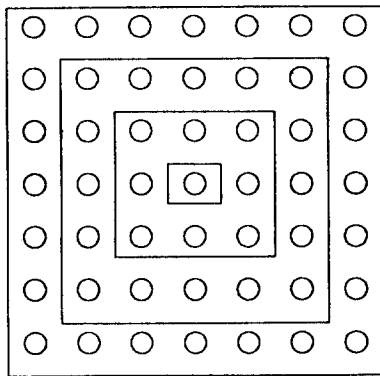
따라서 디지털 방식의 하드웨어 구현에서는 뉴런의 집적도를 높이기 위해 하드웨어 구현에 쉬운 알고리즘의 개발이 필수적인 문제가 되고 이는 또한 신경회로망 알고리즘 개발에 있어 고려 하여야할 가장 중요한 문제중 하나이다.^[1]

본 논문은 Kohonen의 자기조직화 형상지도(Self Organizing Feature Map) 알고리즘을 하드웨어로 구현하기 위하여 학습시간에 따른 단조감소 함수인 학습계수를 일정한 값으로 놓고 이를 보상하기 위하여 이진 강화 함수를 결합한 새로운 학습 알고리즘을 제안하고, 두 가지의 서로 다른 특성을 지닌 데이터를 사용하여 기존의 알고리즘과 성능비교를 하였다.

II. SOFM 신경회로망

1. SOFM 신경회로망의 특징

SOFM 신경회로망은 비지도학습법(Unsupervised Learning)을 이용하는 대표적인 신경회로망으로서 각 뉴런이 다른 뉴런들 사이에 의해 자기조직에 의한 유클리드 거리에 따른 조직적인 지도를 형성한다.



SOFM신경회로망의 학습 알고리즘은 다음과 같다.

$$w_r(t+1) = w_r(t) + \epsilon(t) h_{ns}(t)(v(t) - w_r(t)) \quad (1)$$

for $\forall r \in A, s \in A$

여기서 $\epsilon(t)$ 는 학습계수(learning rate), $h_{ns}(t)$ 는 근접 상호작용 함수 혹은 집합(Neighborhood Interaction Function or Set), $w_r(t)$ 는 시간 t에서의 뉴런 r에서의 Weight Vector, $v(t)$ 는 입력 벡터이며, A는 뉴런이 위치하는 Compact 공간을 의미한다.^[2]
[3]

특히 $\epsilon(t)$ 의 경우 시간에 따른 단조감소 함수로서 Kohonen은 다음과 같은 감소 함수를 제안 하였다.^[4]

$$\epsilon(t) = 0.9 \cdot (1 - \frac{t}{\text{Number of Iteration}})$$

이 함수는 일반적인 회귀 확률 알고리즘(Recursive Stochastic Algorithm)의 수렴조건 (2)를 반드시 만족하는 것은 아니다.(Glady Shev Theorem)^[5]

$$\sum_1^{\infty} \epsilon(t) = \infty \quad (2)$$

$$\sum_1^{\infty} \epsilon(t)^2 < \infty$$

Kohonen은 최근 위에서 소개한 Glady Shev Theorem을 만족하도록 $\epsilon(t)$ 에 동일한 제한조건을 부과하였으나 국내외에서 위 조건을 만족하지 않더라도 학습이 정확히 잘 이루어짐을 실험적으로 혹은 이론적으로 보이고 있다.^{[6] [7] [8]}

특히 K.Schulten 등은 SOFM의 Fokker Plank 방정식을 유도하여 다음의 수렴조건을 유도, Kohonen의 $\epsilon(t)$ 가 타당함을 보였다.^[8]

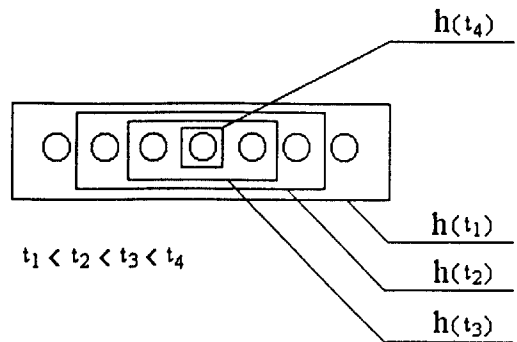


그림 1. 근접 상호작용 함수 $h_n(t)$ 의 작용에 관한 개념도

Fig. 1. The concept figure of Neighborhood function $h_n(t)$.

$$\lim_{t \rightarrow \infty} \int_0^t \varepsilon(\tau) d\tau = \infty \tag{3}$$

$$\lim_{t \rightarrow \infty} \varepsilon(t) = 0 \tag{4}$$

근접 상호작용 함수 $h_{rs}(t)$ 는 시간에 대한 단조 감소 함수로서 승자 뉴런주위의 어떤 범위까지의 뉴런을 학습 시킬 것인가를 정의하는 함수이다.

또한 이 함수는 SOFM의 특징인 각 뉴런들 사이의 자기 조직화를 이룰수 있도록하는 성질을 가지고 있어 일단 각 뉴런에 대응하는 Weight Vector(혹은 Code Book Vector)들이 형식 뉴런 공간위에 위상적 배열을 갖도록 하고 형성된 위상에 대해서는 시간에 대한 단조감소 특성을 통해 그것을 보존하도록 한다. 그림 1은 2차원으로 SOFM을 구성할 때(좌측 그림)와 1차원으로 SOFM을 구성할 때(우측 그림) 근접 상호작용 함수의 시간에 따른 작용 범위를 나타낸다. 우측 그림에서 알 수 있듯 $h_{rs}(t)$ 의 t가 커질 수록 Update범위는 작아져 최종적으로 Winner Neuron만이 Update된다.

2. 일정 학습계수 특성과 이진 강화학습

앞에서 알아본 바와 같이 학습계수 $\varepsilon(t)$ 는 시간에 따라 단조감소하는 함수로서 특히 각 패턴 데이터 ($v(t)$)와 평형상태 대표벡터 $\bar{w}(t)$ 간의 평균제곱 오차를 최소화 하기 위한 필요충분 조건이다^[2].

여기에 대해 학습계수를 시간에 따라 단조 감소하지 않고 항상 일정한 값을 갖도록 놓게 되면 일반적인 회귀 확률 알고리즘과는 달리 SOFM자체의 특징에 따라 일정 학습계수를 가졌다 하더라도 평균제곱 오차를 최소화 하는 방향으로 학습이 진행되게 되며 일단 조직화가 이루어 지게 되면 비록 평균제곱오차가 최소값은 아닐지라도 자기조직화 형상지도를 구성하게 된다.^[4]

이를 H.Ritter와 K.Schulten에 의해서 유도된 상관행렬과 전파 방정식을 통해 구체적으로 살펴보면

먼저 전파 방정식에서 평형상태 대표벡터 $\bar{w}(t)$ 와 비평형 상태 벡터 $w(t)$ 간의 변량벡터 $u(t)$ 를 정의하면 다음관계를 가지게 되고

$$u(t) = w(t) - \bar{w}(t)$$

변량벡터 $u(t)$ 의 평형상태 벡터의 전파 방정식은 다음과 같이 주어지므로

$$\bar{u}_r(t) = Y(t)\bar{u}_r(0) \tag{5}$$

$$Y(t) = \exp(-B \int_0^t \varepsilon(\tau) d\tau) \tag{10}$$

$$\bar{u}(t) = 0 \quad t \rightarrow \infty$$

$$Y(t) = \text{Propagator}$$

$$B_{rmt;n} = \frac{\partial V_{rm}(w)}{\partial w_{r;n}} \Big|_{w=\bar{w}}$$

$$V_{rm} = \sum_s (w_{rm} - \bar{v}_{sm}) h_{rs} \bar{f}_s(w)$$

이어야 하므로 위 (3)식을 만족하여야 하고 (3)식은 ε 을 일정하게 놓으면 만족하므로 ε 일정이 타당한 조건이 될 수 있음을 알 수 있다.

또한 변량벡터 $u(t)$ 의 평형상태 벡터 $\bar{u}(t)$ 와의 상관 행렬의 표준(Norm)을 살펴보면

$$\begin{aligned} \|C(t)\|^2 &\leq \gamma \int_0^t \varepsilon(t')^2 \exp\left(\int_{t'}^t -\beta \varepsilon(t'') dt''\right) dt' \\ &\leq \frac{\gamma \varepsilon^2}{\beta} \end{aligned} \tag{6}$$

(n 은 벡터의 차수)

상관 행렬의 표준(Norm)이 0으로 수렴 하는것이 아닌 적응이득의 제곱에 비례하는 만큼의 오차를 지니게 된다. (그러나 학습계수가 차수에 비해 매우 작다면 상관행렬의 표준의 제곱 $\|C(t)\|^2$ 은 매우 작은 값이 된다.) 이러한 오차는 평균 제곱 오차를 최소화 하는 평형상태 $\bar{w}(t)$ 주변에 일정적응이득이 적용될 경우의 평형상태 벡터 $\bar{w}_c(t)$ 가 분포함을 의미한다.따라서 이러한 오차가 작으면 작을 수록 $\bar{w}_c(t)$ 가 $\bar{w}(t)$ 에 가깝게 분포함을 의미한다.^[11]

이진강화 함수는 $w(t)$ 를 보다 평균제곱 오차를 최소화하는 대역 최소점(Global Minimum Point)으로 보내게 하는데 쓰인다.

이진강화학습을 이용한 적응 필터 알고리즘은 다음과 같다.^[12]

$$c_{k+1} = c_k - a u_k \text{sgn}[e_k]$$

$$e_k = c_k \cdot u_k - d_k$$

$$u_k = \varphi(x_k, x_{k-1}, \dots)$$

φ : 관찰 과정(Observable Process) x_k 를 데이터 과정 u_k 로 보내는 사상

d_k : 요구 과정(Desirable Process)

a : 일정 적응 이득(Constant Adaptive gain)

여기에서 $u_k = 1$ 로 놓고 $d_k = c^* \cdot u_k + f_k$ 로 놓으

면 f_k 는 데이터 신호가 된다. c_k 와 d_k 가 동일 공간상에 존재한다면 이것은 확률적 최급 강하법의 한 종류가 된다¹¹³

위 방정식을 고쳐쓰면 다음과 같다.

$$\begin{aligned} w_{k+1} &= w_k - \rho(t) \operatorname{sgn}[e_k] \\ e_k &= w_k - x_k \\ x_k &= w^* + n_k \end{aligned} \tag{7}$$

x_k : 데이터 벡터 열

w^* : w 의 최적 평형 벡터

n_k : 오차 혹은 잡음 벡터 열

$\rho(t)$: 시간에 따라 1의 발생확률이 감소하는 랜덤변수 $\rho(t) \rightarrow 0$ as $t \rightarrow \infty$

이진강화함수의 수렴성 판별은 다음식에 의하여 구한다¹¹³¹

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} \sum_{k=1}^n E|e_k| \leq E|e^*| + E\rho(t)h \quad \rho(t) \rightarrow 0 \text{ as } n \rightarrow \infty \leq E|e^*|$$

즉 평균 절대오차의 상한이 평균 최적오차보다 같거나 작으므로 위 알고리즘은 확률 1로 최적 평형상태 $\bar{w}(t)$ 로 수렴해 들어간다는 것을 의미한다.

III. 이진 강화함수와 일정 학습계수가 결합된 SOFM의 제안

앞 절의 식 (6)에서 알 수 있듯이 학습계수 $\epsilon(t)$ 가 시간에 관계없이 일정하면 $w(t)$ 는 평균제곱 오차를 최소화 할 수 있는 평형상태 벡터 $\bar{w}(t)$ 주변에 분포하게 된다. 반면 이진 강화 함수는 평균 제곱 오차를 최소화 할 수 있는 평형상태 벡터 $\bar{w}(t)$ 에 수렴한다. 이진 강화 함수는 그러나 수렴에 많은 시간을 필요로 하므로 수렴속도와 수렴사이에 Trade-off를 발생시킨다¹¹²¹. 그러므로 초기값이 대역 최소점 부근에 분포하게 된다면 이러한 문제는 사라지게 될 것이다.

이진 강화함수와 일정 학습계수가 결합된 SOFM은 일정적응이득을 가진 확률회귀 알고리즘과 이진 강화 함수의 결합을 통해 양자의 문제점을 상호 보완하고 디지털 방식의 신경망 하드웨어 구현에 보다 적합한 형태로 Kohonen의 SOFM알고리즘을 보완하는데 그 목적이 있다.

이진 강화함수와 일정 학습계수가 결합된 SOFM은

다음과 같다.

1. 뉴런 r 이 winner 이면

$$w_r(t+1) = w_r(t) + \epsilon h_{rs}(t)(v(t) - w_r(t)) + \rho(t) \operatorname{sgn}(v(t) - w_r(t)) \tag{8}$$

2. 뉴런 r 이 winner는 아니나 $r \in h_{rs}$ 이면

$$w_r(t+1) = w_r(t) + \epsilon h_{rs}(t)(v(t) - w_r(t)) \tag{9}$$

3. 뉴런 r 이 winner가 아니고 $r \notin h_{rs}$ 이면

$$w_r(t+1) = w_r(t) \tag{10}$$

제안된 SOFM알고리즘은 오차 $\bar{u}_k(t)$ 가 R^r 공간내 수열인 특성을 이용하여 이것이 이진 강화 함수에 의해 일정 학습계수로 결정되는 어떤 영역 K_s 에 $\delta > 0$ 인 임의의 양수에 대하여 다음을 만족하도록 하는 것이다.

$$P\{\bar{u}_k(t) \in K_s\} \geq 1 - \delta$$

이렇게 되면 $\bar{u}_k(t)$ 는 “긴밀히 되었다”(be tight)고 하고 분포내에 수렴하는 부분수열이 있음을 의미한다¹¹⁴. 이진 강화함수와 일정 학습계수가 결합된 SOFM 알고리즘에는 이진 강화 함수가 수렴하는 부분 수열이므로 이는 타당하다 하겠다.

[증 명2] 이진 강화 함수에 의한 약 수렴성

(5)식에서

$$\bar{u}_k(t) = Y(t) \bar{u}_k(0)$$

$$Y(t) = \exp(-B \int_0^t \epsilon(\tau) d\tau)$$

여기서 $\bar{u}_k(t) \in R^r[0, \infty)$ 이기 위해서는 $\int_0^\infty |\bar{u}_k(t)|' dt < \infty$ 이어야 한다.

$r=2$ 에 대하여 살펴보면

$$\begin{aligned} \lim_{t \rightarrow \infty} \int_0^t |\bar{u}_k(\tau)|^2 d\tau &= \lim_{t \rightarrow \infty} \int_0^t Y(\tau)^2 \bar{u}(0)^2 d\tau \\ &= \lim_{t \rightarrow \infty} \int_0^t \exp(-2B \int_0^\tau \epsilon(\tau') d\tau') \bar{u}(0)^2 d\tau \\ &= \lim_{t \rightarrow \infty} \bar{u}(0)^2 \frac{1}{2B\epsilon} (1 - \exp(-2B\epsilon t)) < \infty \end{aligned}$$

이므로 $\bar{u}_k(t) \in R^r[0, \infty)$

그리고 식 (6)에서 $\|C\| \leq n\epsilon^2$ 이므로 $|\bar{u}(t) - u(t)| \propto \|C\|$ 인 어떤 값을 항상 유지하게 된다. $\lim_{t \rightarrow \infty} |\bar{u}(t) - u(t)| = K_{\max}$ 로 놓고 K_{\max} 를 상 극한으로 하는 영

역 K_δ 을 생각하자. (K_δ 는 $\|C\|$ 에 의해 결정된다. 여기에서 δ 는 Constant Adaptation Gain일때 $t \rightarrow \infty$ 에서 K_δ 에 $\overline{u}_k(t)$ 가 들어가지 않을 확률. 1보다 매우 작은 임의의 값이 된다.)

이진 강화함수에 의해 상관행렬을 다음과 같이 놓으면

$$C_{r_m r_n} = \langle (u_{r_m} - \overline{u}_{r_m})(u_{r_n} - \overline{u}_{r_n}) \rangle$$

$$u_{r_m} = u_{r_m}^s + u_{r_m}^r \alpha(t)^r \quad u_{r_m}^r = \text{강화함수에 의한 교란 성분}$$

$u_{r_m}^s =$ SOFM에 의한 오차성분

$$\alpha(t) = \text{교란 성분 발생확률 최대 0.5}$$

$$\alpha(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty$$

$$\begin{aligned} C_{r_m r_n} &= \langle (u_{r_m} - \overline{u}_{r_m})(u_{r_n} - \overline{u}_{r_n}) \rangle \\ &= \langle (u_{r_m}^s - \overline{u}_{r_m}^s)(u_{r_n}^s - \overline{u}_{r_n}^s) \rangle \\ &\quad + u_{r_m}^r u_{r_n}^r \alpha(t)^r + u_{r_m}^s u_{r_n}^r \alpha(t)^r \\ &\quad + u_{r_m}^r u_{r_n}^s \alpha(t)^r - \overline{u}_{r_m}^r \overline{u}_{r_n}^r \alpha(t)^r - \overline{u}_{r_m}^r \overline{u}_{r_n}^s \alpha(t)^r \end{aligned}$$

$\alpha(t)$ 의 최대 값은 0.5 이므로 $\alpha(t)^r$, $\alpha(t)^s$ 중 하나는 0이라고 놓을 수 있다. 이에 따라 위 식을 간단히 하면

$$\begin{aligned} C_{r_m r_n} &= \langle (u_{r_m}^s - \overline{u}_{r_m}^s)(u_{r_n}^s - \overline{u}_{r_n}^s) \rangle + u_{r_m}^r u_{r_n}^r \alpha(t)^r - \overline{u}_{r_m}^r \overline{u}_{r_n}^r \alpha(t)^r \\ &\leq \langle (u_{r_m}^s - \overline{u}_{r_m}^s)(u_{r_n}^s - \overline{u}_{r_n}^s) \rangle + \langle u_{r_m}^r u_{r_n}^r \alpha(t)^r - \overline{u}_{r_m}^r \overline{u}_{r_n}^r \alpha(t)^r \rangle \\ &\leq \langle (u_{r_m}^s - \overline{u}_{r_m}^s)(u_{r_n}^s - \overline{u}_{r_n}^s) \rangle + \nu(t) \\ \therefore \nu(t) &= \langle u_{r_m}^r u_{r_n}^r \alpha(t)^r - \overline{u}_{r_m}^r \overline{u}_{r_n}^r \alpha(t)^r \rangle \end{aligned}$$

If $\nu(t) < 0 \quad \|C(t+1)\| < \|C(t)\|$ 이므로 $P(K_{\max} \in K_\delta) \leq 1 - \delta$

If $\nu(t) > 0 \quad \|C(t+1)\| > \|C(t)\|$. 그러나 그림에서 볼 수 있듯이 그 다음 Update과정에서 K_{\max} 는 영역 K_δ 내로 들어 오게 된다.

$$\epsilon(v(t) - w(t)) > \epsilon(v(t) - (w(t) + \Delta w))$$

$\therefore \Delta w$ 는 $\nu(t) > 0$ 에 의하여 증가한 부분

따라서 $\|C(t+2)\| < \|C(t)\| < \|C(t+1)\|$ 이므로 $P(K_{\max} \in K_\delta) \leq 1 - \delta$

$\|C(t)\|$ 는 강화함수의 작용으로 Constant Learning rate가 작용되었을 때 보다 더 작은 영역으로 약 수렴하게 된다. (증명 끝)

Neighborhood Function에 의해 Weight Vector는 Constant Adaptation Gain이 적용되더라도 발산하거나 하지 않고 어떤 Boundary를 따라 분포하게 된다. (Cyclo Ergodic) 이때 Binary Reinforcement Function이 작용하게 되면 두번째 Update부터 Boundary내부로 들어가게 된다. Boundary내부로

들어가게 된 Weight Vector가 Boundary내부를 빠져 나오려면 충분히 큰 Reinforcement Function이 작용하지 않으면 안되는데 Reinforcement Function에서 $t \rightarrow \infty$ 에 따라 $\rho(t) \rightarrow 0$ 이므로 $t \rightarrow \infty$ 에서 Weight Vector는 Boundary내부에 존재하게 된다.

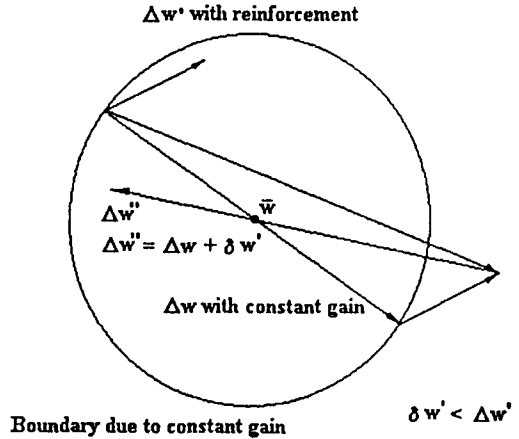


그림 2. Binary Reinforcement Function과 Constant Adaptation Gain이 적용되었을 때의 Weight Vector의 상태 개념도

Fig. 2. The concept figure about the state of weight vector, with Binary Reinforcement function and constant adaptation gain.

이진 강화함수와 일정 학습계수가 결합된 SOFM은 다음과 같은 특징을 가지고 있다.

1) Hill Climbing 특성

그림에서 볼 수 있듯이 회귀 방정식에 의해 진행되는 Weight 벡터의 방향은 대역 최소치부근의 어떤 국소 최소치로 향하게 된다. 여기에 이진 강화 함수에 의해 만들어진 벡터는 대역 최소치 방향으로의 어떤 성분을 지니게 된다. 따라서 확률적 최급강하법(Stochastic Gradient Descent Rule)에 따른 성분에 대해 임의의 벡터가 부가되어 Weight Vector가 원래 방향으로 갈때 보다 약간 더 높은 에너지를 포함하게 되므로 국소 최소점을 빠져 나올 수 있게 된다.

그러므로 이진 강화 함수는 일종의 Hill Climbing 함수로 볼 수 있고 따라서 50% 이상 적용되어서는 안 된다.^[15] 그러므로 $\rho(t)$ 는 다음과 같은 두가지 향으로 이루어 져야 한다.

- $P(t)$: 1을 발생할 확률이 0.5인 랜덤변수, 시간 t 에 대하여 독립이다.
- $\epsilon_{RR}(t)$: 1을 발생할 확률이 시간에 따라 단조감소하

는 랜덤변수 $t=0$ 에서는 확률 1로 1을 발생하고 $t \rightarrow \infty$ 에서는 $P(\epsilon_{RR}(t)=1) \rightarrow 0$ 이다.

2) Momentum항의 부가 특성

SOFM의 Momentum항은 다음과 같이 나타나 질 수 있다.^[16]

$$\Delta w_i(t) = \epsilon(t) h_{rs}(v(t) - w_i(t)) P(v) + \alpha \Delta w_i(t-1)$$

위 방정식은 1차 모멘트 항이 된다. 여기에 대해 n 차 모멘트 항을 가정하게 되면 위 방정식은 다음과 같다.

$$\Delta w_i(t) = \epsilon(t) h_{rs}(v(t) - w_i(t)) P(v) + \sum_{k=1}^n \alpha(k) \Delta w_i(t-k)$$

여기에서 $\alpha(k)$ 에 대한 정보를 명확히 얻을 수 없으므로 $\Delta w_i(t-k)$ 항의 영향을 결정하기는 어렵다. 따라서

$$\sum_{k=1}^n \alpha(k) \Delta w_i(t-k)$$

항이 어떠한 영향을 가져다 줄 것인가를 명확히 할 수 없다면, 어떤 랜덤 벡터를 사용하여 시뮬레이션 할 수 있을 것이다. 제안된 SOFM알고리즘에서는 이진 강화 함수가 이 역할을 하고 있다.

3) Jittered Data Maker특성

입력 데이터 v 가 $P(v)$ 의 확률 분포를 가지고 있다면 j 번째 클러스터를 대표하며 평균 제곱오차를 최소화 하는 대표 벡터는 다음과 같이 주어진다.^[17]

$$\hat{w}_j = \frac{\int_{I(j)} v P(v) dv}{\int_{I(j)} P(v) dv}$$

만일 데이터의 확률분포를 알고있다면 위와 같이 클러스터의 무게중심을 직접 계산할 수 있으므로(대역 최소치는 그러므로 클러스터의 무게중심이 된다) 학습

과정없이 대표벡터를 결정 할 수 있다. 하지만 일반적으로 데이터 전체의 확률분포를 알 수 없으므로 제한된 입력 데이터의 확률분포를 통해 전체 데이터의 확률분포를 추정하여야 한다. 이렇게 추정된 확률 분포 $\hat{P}(l)$ 가 원래의 확률분포와 근사하다면 문제가 없으나 그렇지 않다면 추정된 확률 분포 $\hat{P}(l)$ 에 의해 판별된 대역 최소치는 실제로 하나의 국소 최소치에 지나지 않게 된다. 따라서 이러한 문제점을 극복하기 위해 입력 데이터의 성질을 왜곡하지 않는 범위에서 Jittered 데이터를 뿌려 학습 정확도를 높이기도 한다.

확률적 회귀 알고리즘에 의해 생성된 벡터를 추정된 확률 분포 $\hat{P}(l)$ 에 의한 대역 최소치를 향해 가는 벡터로 본다면 이진 강화함수로 생성된 벡터는 Jittered 데이터에 대한 영향으로 교란된 벡터로 볼 수 있다. 따

라서 실제 데이터보다 더 많은 데이터를 통해 대역 최소치를 찾아 가는 것처럼 되므로 추정된 확률 분포와 실제 확률분포가 크게 다른 "Dirty Data"에 대하여 더 높은 정확도를 얻을 수 있다.

(4) 하드웨어 구성상의 용이성

이 알고리즘은 단 한군데도 곱셈연산을 하지 않으뿐더러 일정 학습계수를 가짐에 따라 발생하는 나뉠셈도 비트 교환을 통해 구현 할 수 있으므로 1개의 덧셈기와 2개의 (혹은 3개의) 데이터 레지스터, 작은 비트의 플래그 레지스터, 1개의 디코더, 그리고 전역 제어기(Global Controller)와 1 비트 파이프라인으로 1개의 뉴런을 구성 할 수 있다. 즉 전역 제어기와 명령 디코더는 일반적인 SOFM알고리즘에도 동일하게 구현되어야 하는 대신 SOFM알고리즘에서는 필수적인 곱셈기와 나뉠셈기를 삭제 할 수 있으므로 보다 간편하게 뉴런을 구성 할 수 있게 된다.

IV. 실험결과

실험에 사용한 데이터는 미국 Indiana주 West Lafayette시 근교 지역을 항공 촬영한 원격 탐사 데이터로서 농작물 재배 지역을 촬영한 것과 미국 Colorado주의 산악 지역의 지형을 촬영한 데이터 2가지이다.

표 1. 8 Class Data

Table 1. The contents of 8 class data.

클래스	영역설명
1	풀밭
2	옥수수
3	귀리
4	붉은클로버
5	콩
6	밀
7	콩터
8	호밀

이중 Indiana주의 농작물 재배 지역을 촬영한 것은 추정된 확률 분포와 실제 확률분포가 근사하여 분류 결과가 비교적 우수한 특징을 지닌다.(따라서 이 데이터를 약하여 "Clean Data"라 부르도록 하겠다.) 반면 Colorado 지역의 지형을 촬영한 데이터는 추정된 확률 분포와 실제 확률분포가 크게 달라 분류 결과가 전자에 비하여 현저히 떨어진다.(따라서 이 데이터를 약하여 "Dirty Data"라 부르도록 하겠다.)

"Clean Data"는 미국 중부지방의 평원 지역을 항공 촬영한 원격 탐사 데이터를 사용하였다. 이 데이터는 측정 주파수 범위에 따라 8-밴드를 가지며 0에서

255사이의 정수값을 가진다. 각 클래스는 200개와 375개의 서로다른 패턴으로 구성되어 있으며 학습에는 200개의 데이터로, 학습 결과의 테스트에는 375개의 데이터를 사용했다. 각 클래스의 특징은 다음과 같다.

표 2. 4 Class Data

Table 2. The contents of 4 class data.

클래스	영역설명
1	푼밭
2	옥수수
3	귀리
4	붉은클로버

표 3. 4 Class $\epsilon = 1 / 16 = 0.0625$

Table 3. The Experimental result with 4 class clean data

EPOCH	Kohonen Original -Training	Kohonen Original -Testing	Proposed Algorithm -Training	Proposed Algorithm -Testing
1000	97.125	93.93	97.25	94.27
900	96.625	94.26	97.125	94.73
800	97.00	93.93	97.25	93.93
700	97.375	94.13	97.625	94.20
600	96.875	94.53	97.25	93.40
500	97.375	94.27	97.00	94.13
400	96.375	94.00	97.125	94.53
300	97.375	94.53	96.875	94.60
200	96.875	94.00	97.25	94.33
100	97.375	94.60	97.125	94.20
Average	97.04	94.218	97.187	94.232

Kohonen Original SOFM의 경우 학습계수는

$$\epsilon(t) = 0.9(1 - \frac{t}{\text{Number of iteration}})$$

$h(t)$: 시간에 따라 감소하는 Topological Neighborhood

표 4. 8 Class $\epsilon = 1 / 16 = 0.0625$

Table 4. The Experimental result with 8 class clean data.

EPOCH	Kohonen Original -Training	Kohonen Original -Testing	Proposed Algorithm -Training	Proposed Algorithm -Testing
1000	94.1875	91.67	94.625	90.87
900	92.8175	89.47	94.375	92.77
800	94.125	92.77	94.5625	92.60
700	94.9375	92.70	94.00	92.43
600	94.00	92.70	93.625	92.73
500	93.8175	91.67	94.125	92.90
400	93.4375	90.63	94.00	90.57
300	94.1875	91.60	95.0625	92.67
200	94.375	92.07	93.5625	90.80
100	92.9375	90.20	94.00	91.73
Average	93.88	91.55	94.19	92.01

이진 강화함수와 일정 학습계수가 결합된 SOFM의 경우 적응이득은 2⁻⁴의 형태로 놓아야 하드웨어 구현에 편리하므로

$$\epsilon = 2^{-4} = \frac{1}{16} = 0.0625$$

$h(t)$: 시간에 따라 감소하는 Topological Neighborhood

표 5. Colorado 지역 데이터에 의한 실험 결과

Table 5. The experimerdal result of Colorado data.

EPOCH	Kohonen's Original -Training	Kohonen's Original -Testing	Proposed Algorithm Training	Proposed Algorithm Testing
100	57.2451	55.6874	67.6587	64.3917
200	56.7460	54.0060	65.7579	64.5895
300	56.5476	54.0060	69.8413	65.7765
400	56.7460	53.1157	65.7738	61.6222
500	54.7619	51.1375	67.4603	62.9080
600	56.5476	53.6103	69.6429	65.6775
700	56.7460	53.8081	71.0317	66.3699
800	55.7540	53.0168	68.8492	64.8863
900	55.2579	52.6212	70.1389	66.4668
1000	56.7460	52.9179	68.3532	64.1939
Average	56.4095	53.3927	68.6508	64.6884

표 6. Colorado 데이터의 각 Class당 데이터 갯수

Table 6. The number of data per class in Colorado data.

클래스	데이터 수
1	302
2	56
3	21
4	23
5	70
6	87
7	35
8	35
9	122
10	147
11	38
12	25
13	50

랜덤값의 경우 시간에 독립적으로 값을 얻어오는(난수표 참조방식)방법을 택했다. 그것은 시간에 가변하는 랜덤값을 취할 경우 각 Iteration에 대하여 서로 다른 정확도를 지니기 때문에 실험결과가 일정치 않다는 단점이 있었다. 대체로 시간에 가변하는 값을 취할때 보통 0.1에서 0.2%의 정확도 향상을 볼 수 있었다.

표에서 알 수 있듯 "Dirty Data"의 경우 11%~12%정도의 정확도 향상이 있었다.

V. 결 론

본 논문에서는 디지털 하드웨어 구현에 편리한 이진 강화함수와 일정 학습계수를 갖는 학습 알고리즘을 제안 하였다. 이진 강화함수를 학습 알고리즘에 적용 가능하도록 변형 하였고 강화함수를 결합함으로써 얻는 이 점에 대해서도 분석하였다. 또한 일정 학습계수를 갖더라도 SOFM의 알고리즘을 따를 경우 시스템이 불안정해 지는 것이 아닌 대역 최소점 부근의 국소 최소점에 빠질 뿐이라는 것을 보였으며 강화함수를 결합할 경우 약 수렴함을 보였다. 이진 강화함수와 일정 학습계수가 결합된 SOFM을 "Clean Data"와 "Dirty Data"에 대해서 실험한 결과 학습계수가 충분히 작을 경우 Kohonen의 SOFM알고리즘보다 더 향상된 성능을 가짐을 실험을 통해서 확인하였다. 또한 하드웨어 구성상에도 다른 알고리즘과는 달리 보다 간단하게 그리고 보다 높은 뉴런 집적도를 가질 수 있음을 보였다.

참 고 문 헌

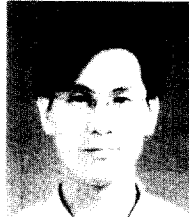
- [1] T.Poggio F.Girosi "Networks for Approximation and Learning", Proc IEEE, vol 78, no 9, Sept 1990, pp 1481-1497.
- [2] z._P.Lo B.Bavarian "On The Convergence in Topology Preserving Neural Networks" Biol.Cybern. 65, 55-63 (1991).
- [3] H.Ritter K.Schulten "Convergence Properties of Kohonen's Topology Conserving Maps: Fluctuations, Stability, and Dimension Selection" Biol.Cybern. 60, 59-71 (1988).
- [4] T.Kohonen "Self Organizing Map" Proc. IEEE.
- [5] Z._P Lo Y.Yu B.Babarian "Convergence Properties of Topology Preserving Neural Networks" IEEE Trans. Neural Networks vol 4, No 2, March 1993.
- [6] T.Kohonen "Generalizations of the Self-Organizing Map" Proc.IJCNN, vol 1, pp457 - 461, 1993.
- [7] 공 성근, "개선된 경쟁학습 알고리즘에 의한 패턴 클러스터 중심의 예측" 제3회 인공지능 신경망 및 퍼지 시스템 종합 학술대회/전시회 논문집 pp 365-368
- [8] E. Domany J.L.van Hemmen K.Schulten (Eds.) Models of Neural Networks Springer-Verlag 2nd printing 1992.
- [9] V.V.Tolat "An Analysis of Kohonen's Self-Organizing Maps Using a System of Energy Functions" Biol.Cybern 64, 155-164 (1990).
- [10] van Kampen Stochastic Processes in Physics and Chemistry North-Holland .1981. Armsterdam.
- [11] C. Darken J.Moody "Note on Learning Rate Schedule for Stochastic Optimization".
- [12] A.Gersho "Adaptive Filtering with Binary Reinforcement" IEEE Trans.Inform. vol IT-30, 2, March 1984.
- [13] L. Botton P.Gallinari "A Framework for the Cooperation of Learning Algorithm" pp781 - 788.
- [14] H.Kushner A.Schwarz "Weak Convergence and Asymptotic Properties of Adaptive Filters with Constant Gains" IEEE Trans.Inform. vol IT-30, NO 2, March 1984.
- [15] S.Y.Kung Digital Neural Networks Prentice Hall, Englewood Cliffs, N.J. 1993.
- [16] M.Hagiwara "Self Organizing Feature Map with a Momentum Term" Proc. IJCNN93, vol 1, pp467-470.
- [17] M.P.Windham "Cluster Validity for the Fuzzy c-Means Clustering Algorithm" IEEE Trans.Pattern Anal.Machine Intell., vol PAMI-4, no.4, July 1982.

저 자 소 개



趙成元(正會員)

1959년 9월 7일 생. 1982년 2월 서울대학교 전기공학과 졸업(공학사). 1987년 2월 미국 Purdue 대학교 대학원 전기공학과 졸업(공학석사). 1992년 8월 미국 Purdue대학교 대학원 전기공학과 졸업(공학박사). 1993년 부터 현재 홍익대학교 전기제어 공학과 조교수. 주 관심분야는 신경회로망, 퍼지 시스템, 인공지능, 음성 인식, 컴퓨터 비전, 데이터 베이스 등임.



石鎮旭(正會員)

1969년 6월 26일생. 1993년 2월 홍익대학교 전기제어 공학과 졸업. 1995년 1월 현재 홍익대학교 대학원 전기제어 공학과 석사과정 재학중. 주 관심 분야는 신경 회로망, 적응제어, 확률 프로세스 등임.