시불변 학습계수와 이진 강화함수를 가진 자기 조직화 형상지도 신경회로망의 동적특성

The Dynamics of Self-Organizing Feature Map with Constant Learning Rate and Binary Reinforcement Function

석 진 욱, 조 성 원
(Jinwuk Seok and Seongwon Cho)

Abstract : We present proofs of the stability and convergence of Self-organizing feature map (SOFM) neural network with time-invarient learning rate and binary reinforcement function. One of the major problems in Self-organizing feature map neural network concerns with learning rate "Kalman Filter" gain in stochastic control field which is monotone decreasing function and converges to 0 for satisfying minimum variance property. In this paper, we show that the stability and convergence of Self-organizing feature map neural network with time-invariant learning rate. The analysis of the proposed algorithm shows that the stability and convergence is guranteed with exponentially stable and weak convergence properties as well.

Keywords : constant learning rate, binary reinforcement function, self-organizing feature map (SOFM), weak convergence

I. 서론

신경회로망의 하드웨어 구현은 그 가능성과 대규모 병렬처리라는 이점에도 불구하고 일반적이고도 상용적인 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 이와같은 문제점을 극복하기 위해 대부분 아날로그 방식 혹은 Stochastic Pulse Stream방식으로 이루어져 왔다.하지만 아날로그 방식의 하드웨어에서는 CMOS내 PMOS 트랜지스터와 NMOS 트랜지스터의 특성 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 데이타의 동기 확보가 역시 문제점으로 지적된다.또한 디지탈 방식의 경우 연산에 필요한 곱셈기나 나눗셈기가 필요하다는 문제점 때문에 뉴런의 집적도가 아날로그 방식이나 Stochastic Pulse Stream방식보다 현저하게 떨어진다는 단점이 있다.

따라서 디지탈 방식의 하드웨어 구현에서는 뉴런의 집적도를 높이기 위해 하드웨어 구현에 쉬운 알고리즘의 개발이 필수적인 문제가 되고 이는 또한신경회로망 알고리즘 개발에 있어 고려 하여야할 가장 중요한 문제중 하나이다[1]. 이같은 문제를 극복 하기 위하여 학습계수를 시불변의 작은 일정한 값으로 하고 이로인해 파생되는 국소 최소점 문제에 대한 해결책으로 이진 강화함수를 부가한 형태의 하드웨어 구현에 용이한 신경회로망 알고리즘을 제안한 바 있다[18].

본 논문은 자기조직화 형상지도(Self Organizing Feature Map)알고리즘이 Exponentially stable함을 보이고 동시에 이진 강화함수를 시 불변 학습율을 가지는 자기조직화 형상지도 알고리즘에 결합 시켜도 마찬가지로 안정성(Stability)이 보장 됨을 보이며, 제안한 알고리즘이 이 특성을 만족시킬 경우 $w_r(n)$이 $n\to\infty$에서 약 수렴(Weak Convergence)함을 증명한다. 본 논문의 구성은 다음과 같다. 먼저 2장에서는 자기조직화 형상지도 알고리즘을 간략히 설명하고 일반적 자기조직화 형상지도 알고리즘이 가지는 수렴 특성을 밝히며, 3장에서는 시불변 학습계수를 가지는 자기조직화 형상지도 알고리즘과 이진 강화함수를 여기에 결합시킨 알고리즘의 Exponentially stable 특성을 증명한다. 4장의 1절에서는 제안한 알고리즘의 약 수렴성을 보이기 위하여 $w_r(n)$의 극점과 $w_r(n)$과의 Covariance를 살펴보고 2절에서는 제안한 알고리즘의 약 수렴성을 보인다. 마지막 5장에서는 본 논문의 결론을 맺는다.

II. 자기 조직화 형상지도(SOFM) 신경회로망

1. 자기 조직화 형상지도(SOFM) 신경회로망 알고리즘

SOFM신경회로망은 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.

$$w_r(t+1) = w_r(t) + \varepsilon(t) h_{rs}(t)(v(t) - w_r(t)) \quad \text{for } \forall\ r \in A,\ s \in A \tag{1}$$

여기서 $\varepsilon(t)$는 학습계수(learning rate), $h_{rs}(t)$는 근접 상호작용 함수 혹은 집합(Neighborhood Interaction Function or Set), $w_r(t)$는 시간 t에서의 뉴런 r 에서의 Weight Vector, $v(t)$는 입력 벡터이며, A는 뉴런이 위치하는 Compact 공간을 의미한다[2][3]. 특히 $\varepsilon(t)$의 경우 시간에 따른 단조감소 함수로서 Kohonen은 다음과 같은 감소 함수를 제안 하였다[4].

$$\varepsilon(t) = 0.9 \cdot \left(1 - \frac{t}{Number\ of\ Iteration}\right)$$

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

$$\sum_{1}^{\infty} \varepsilon(t) = \infty, \qquad \sum_{1}^{\infty} \varepsilon(t)^2 < \infty \tag{2}$$

접수일자 : 1995. 11. 29., 수정완료 : 1995. 4. 30.
석진욱, 조성원 : 홍익대학교 공과대학 전자·전기공학군
※ 본 논문은 1995년도 학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

Kohonen은 최근 위에서 소개한 Glady Shev Theorem을 만족하도록 $\varepsilon(t)$에 동일한 제한조건을 부가하였으나 국내외에서 위 조건을 만족하지 않더라도 학습이 정확히 잘 이루어짐을 실험적으로 혹은 이론적으로 보이고 있다[6][7][8]. 근접 상호작용 함수 $h_{rs}(t)$는 시간에 대한 단조 감소함수로서 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.

2. 자기 조직화 형상지도(SOFM) 신경회로망의 수렴특성

K.Schulten 등에 의해 유도된 SOFM의 Fokker Plank 방정식은 다음과 같다[3][8].

$$\partial_t S(\mathrm{u},t) = \varepsilon(t)\sum_{rmr'n}\frac{\partial}{\partial \mathrm{u}_{rm}}\mathbf{B}_{rmr'n}\mathrm{u}_{r'n}S(\mathrm{u},t) + \frac{\varepsilon(t)^2}{2}\sum_{rmr'n}\mathbf{D}_{rmr'n}\frac{\partial^2 S(\mathrm{u},t)}{\partial \mathrm{u}_{rm}\partial \mathrm{u}_{r'n}} \quad (3)$$

위 (3)에서

$$\mathbf{B}_{rmr'n} = \partial\frac{\sum_S(\mathrm{w}_{rm}-\mathrm{v}_{sm})h_{rs}\widehat{P_s}(\mathrm{w})}{\partial \mathrm{w}_{r'n}}\Big|_{\mathrm{w}=\bar{\mathrm{w}}}$$ 인 Matrix 이며 $\mathbf{D}_{rmr'n}$은 $E[\delta \mathrm{w}_{rm}\delta \mathrm{w}_{r'n}]$ 인 Matrix이다. Fokker Plank 방정식의 독립변수 u_{rm}은 r번째 Weight vector의 m번째 성분 w_{rm}과 그것의 평균값 $\bar{\mathrm{w}}_{rm}$과의 차이다. 따라서 u_{rm}의 Correlation을 구하여 그것의 Matrix Norm을 찾아보면 SOFM 신경회로망의 수렴 특성은 구할 수 있다. H. Ritter등은 다음의 수렴 판별식을 유도하여 SOFM 신경회로망의 수렴 특성을 구하였다.

$$\|\mathbf{C}(t)\|^2 \le \gamma\int_0^t \varepsilon(t')^2\exp\left(-\beta\int_{t'}^t \varepsilon(t'')dt''\right)dt'e \quad (4)$$

위 (4)에서 $\beta>0$는

$TRACE\ \mathbf{C}[\mathbf{B}(\bar{\mathrm{w}})+\mathbf{B}(\bar{\mathrm{w}})^T]\mathbf{C} > \beta\|\mathbf{C}\|^2/2$를 만족하는 값이며 $\gamma>0$는 $\partial_t\|\mathbf{C}\| \le -\varepsilon(t)\beta\|\mathbf{C}\|^2+\varepsilon(t)^2\gamma$를 만족하는 값이다. (4)에서 1차 수렴조건은 다음과 같다.

$$\exists t'>0 \quad t'\in R \quad s.t. \quad \lim_{t\to\infty}\int_{t'}^t \varepsilon(\tau)d\tau = \infty \quad (5)$$

또 다른 수렴조건은 SOFM에 의해 만들어진 수열 $\{\mathrm{w}_r(t)\}_{t=t_0}^{\infty}$이 Cauchy Sequence임을 증명하여 유도할 수 있다. 먼저 SOFM 계열의 신경회로망 알고리즘의 위상적 특징과 SOFM알고리즘의 특징에서 다음의 가정을 유도할 수 있다.

Assumption 1 :

i. 학습율 $\varepsilon(t)$와 $h(\mathrm{r},t)$는 다음을 만족한다.

$$sup_t\ \varepsilon(t) = 1 \qquad sup_{(r,t)}\ h(\mathrm{r},t) = 1$$

ii. $\forall t\in\Lambda\ \exists\rho_t>0\ \&\ \mathrm{v}(t)\ s.t.$
$\mathrm{v}(t)\in B^o(\mathrm{v}(t),\mathrm{w}_r(t)) = \{\mathrm{v}(t)\in R^n | d(\mathrm{v}(t),\mathrm{w}_r(t))<\rho_t\}$

iii. 목적함수 $f(\mathrm{v}(t),\mathrm{w}_r(t)) = \sum_{r=0}^{n}\|\mathrm{v}(t)-\mathrm{w}_r(t)\|^2$의 Gradient $\nabla f(\mathrm{v}(t),\mathrm{w}_r(t))$는 Lipschitz이다.

Theorem 1 : 초기 Weight vector $\mathrm{w}_r(0)=\mathrm{v}(0)$이고 학습율 $\varepsilon(t)$가 $\varepsilon(t)\downarrow 0$이면 $\{\mathrm{w}_r(t)\}_{t=0}^{t<\infty}$은 Cauchy Sequence이다.

Proof : SOFM 신경회로망의 학습 방정식은 다음과 같이 주어진다.

$$\mathrm{w}_r(t+1) - \mathrm{w}_r(t) + \varepsilon(t)h_{rs}(t)(\mathrm{v}(t) - \mathrm{w}_r(t)) \quad (6)$$

위 (6)에서 $\mathrm{w}_r(t)$를 이항하면

$$\|\mathrm{w}_r(t+1)-\mathrm{w}_r(t)\| = \|\varepsilon(t)h_{rs}(t)(\mathrm{v}(t)-\mathrm{w}_r(t))\|$$
$$\le|\varepsilon(t)||h_{rs}(t)|\|\mathrm{v}(t)-\mathrm{w}_r(t)\|$$
$$=|\varepsilon(t)||h_{rs}(t)|\|\mathrm{v}(t)-\mathrm{w}_r(t-1)+\mathrm{w}_r(t-1)-\mathrm{w}_r(t)\|$$
$$\le|\varepsilon(t)||h_{rs}(t)|(\|\mathrm{v}(t)-\mathrm{w}_r(t-1)\|+\|\mathrm{w}_r(t)-\mathrm{w}_r(t-1)\|)$$
$$\le|\varepsilon(t)||h_{rs}(t)|(\|\mathrm{v}(t)-\mathrm{w}_r(0)\|+\sum_{k=0}^{t-1}\|\mathrm{w}_r(k+1)-\mathrm{w}_r(k)\|)$$
$$\le|\varepsilon(t)||h_{rs}(t)|(L\|\mathrm{w}_r(t)-\mathrm{w}_r(0)\|+\sum_{k=0}^{t-1}\|\mathrm{w}_r(k+1)-\mathrm{w}_r(k)\|)$$

By Lipschitz Condition

$$\le|\varepsilon(t)||h(\mathrm{r},t)|(1+L)\sum_{k=0}^{t-1}\|\mathrm{w}_r(k+1)-\mathrm{w}_r(k)\|$$

가정 iii)을 만족하는 $\mathrm{v}(t)$에 대하여 $h_{rs}(t)=1$, 그리고 $inf\ sup_t\ \rho_t$를 만족하는 ρ^M을 잡으면

$$\|\mathrm{w}_r(t+1)-\mathrm{w}_r(t)\| \le|\varepsilon(t)|(1+L)\sum_{k=0}^{t-1}\|\mathrm{w}_r(k+1)-\mathrm{w}_r(k)\|$$
$$\le|\varepsilon(t)|(1+L)t\rho^M$$

그런데 $\varepsilon(t)\downarrow 0$이므로 $|\varepsilon(t)|(1+L)t\rho^M < inf_t\ \rho_t$인 $\varepsilon(t)$를 잡을 수 있다. 따라서 임의의 t 에 대하여

$$\|\mathrm{w}_r(t)-\mathrm{w}_r(t-m)\| \le|\varepsilon(t)|\left(\sum_{j=t-m}^{t-1}\|\mathrm{w}_r(j+1)-\mathrm{w}_r(j)\| + L\sum_{k=0}^{t-m-1}\|\mathrm{w}_r(k+1)-\mathrm{w}_r(k)\|\right)$$
$$\le|\varepsilon(t)|(m+(t-m)L)\rho^M$$

이므로 마찬가지로 $|\varepsilon(t)|(m+(t-m)L)\rho^M < inf_t\ \rho_t$ 인 $\varepsilon(t)$를 잡을 수 있다. 그러므로

$\forall\rho_t>0\ \exists n_0\ s.t.\ n,m>n_0 \Rightarrow \|\mathrm{w}_r(n)-\mathrm{w}_r(m)\|<\rho_t$ ■

Theorem 1에 의하여 SOFM 신경회로망 알고리즘에서 $\{\mathrm{w}_r(t)\}_{t=0}^{\infty}$이 수렴할 필요조건은 $\varepsilon(t)\downarrow 0$임을 알 수 있다. 이상의 두가지 결론을 통하여 SOFM 신경회로망 알고리즘으로 생성된 Sequence $\{\mathrm{w}_r(t)\}_{t=0}^{\infty}$이 수렴할 필요충분 조건은 다음과 같다.

$$\exists t'>0,\ t'\in R\ s.t\ \lim_{t\to\infty}\int_{t'}^t\varepsilon(\tau)d\tau=\infty\ \&\ \lim_{t\to\infty}\varepsilon(t)=0.$$

Ⅲ. 시불변 학습계수와 이진 강화함수를 가진 SOFM 신경회로망의 동적특성

1. 이진 강화함수와 일정 학습계수이 결합된 SOFM 신경회로망

시불변 학습계수와 이진 강화함수가 결합된 SOFM 신경회로망 알고리즘은 다음과 같다[18].

1. 뉴런 r이 $r=\arg{}_r\min_{\mathrm{w}_r(t)} d(\mathrm{v}(t),\mathrm{w}_r(t))$ 이고 $\mathrm{r}\in \mathrm{h}_{rs}$이면

$$\mathrm{w}_r(t+1) = \mathrm{w}_r(t)+\varepsilon h_{rs}(t)(\mathrm{v}(t)-\mathrm{w}_r(t)) + \gamma(t)sgn(\mathrm{v}(t)-\mathrm{w}_r(t)) \quad (7)$$

2. 뉴런 r이 $r\ne\arg{}_r\min_{\mathrm{w}_r(t)} d(\mathrm{v}(t),\mathrm{w}_r(t))$이나 $\mathrm{r}\in \mathrm{h}_{rs}$ 이면

$$\mathrm{w}_r(t+1) = \mathrm{w}_r(t)+\varepsilon h_{rs}(t)(\mathrm{v}(t)-\mathrm{w}_r(t)) \quad (8)$$

3. 뉴런 r이 $r\ne\arg{}_r\min_{\mathrm{w}_r(t)} d(\mathrm{v}(t),\mathrm{w}_r(t))$이고 $\mathrm{r}\notin \mathrm{h}_{rs}$ 이면

$$\mathrm{w}_r(t+1) = \mathrm{w}_r(t) \quad (9)$$

여기서 $\gamma(t)$는 $\delta=\frac{\varepsilon}{2}$인 δ와 다음을 만족하는 random variable $\rho(t)$의 곱이다.

$$\forall t<0 \quad t\in Z^+\ \exists\rho(t)\in\{0,1\} \quad s.t$$
$$P(\rho(t)=1)\le 0.5\ \&\ P(\rho(t)=1)\downarrow 0 \quad as\ t\to\infty$$

제안한 알고리즘에 의해 생성되는 Sequence $\{\mathrm{w}_r(t)\}_{t=0}^{\infty}$

는 그러나 $w_r(t)$가 winner, 즉 $r=\arg_r \min_{w_r(t)} d(v(t), w_r(t))$를 만족 시킬 수 있는 입력 벡터 $v(t)$에 대하여 다음 명제를 성립시킨다.

$$\exists \rho_t > 0 \ s.t \ v(t) \in B^o(v(t), w_r(t)) = \{ v(t) \in R^n | d(v(t), w_r(t)) < \rho_t \}$$

따라서

$$\begin{aligned} \| w_r(t+1) - w_r(t) \| &= \| \varepsilon h_{rs}(t)(v(t) - w_r(t)) + \delta\rho(t) sgn(v(t) - w_r(t)) \| \\ &\le |\varepsilon| |h_{rs}(t)| \| v(t) - w_r(t) \| + \| \delta\rho(t) sgn(v(t) - w_r(t)) \| \\ &\le |\varepsilon| h_{rs}(t) | \rho_t + \frac{\varepsilon}{2} \le \varepsilon(\rho_t + \frac{1}{2}) \end{aligned}$$

이므로 Cauchy sequence가 될 수 없다. 그러므로 학습계수 ε이 어떤 일정한 값일 경우의 SOFM 신경회로망 알고리즘으로 생성된 Sequence $\{ w_r(t) \}_{t=0}^{\infty}$ 는 uniformly convergence되지 않는다.

2. 시불변 학습율과 이진 강화함수가 결합된 SOFM 알고리즘의 Exponentially Stable성질

시불변 학습율을 가진 경쟁 학습 알고리즘의 경우 앞 절에서 살펴 보았듯이 Cauchy Sequence가 될 수 없다. 그러나

$$v(t) \in B^o(v(t), w_r(t)) = \{ v(t) \in R^n | d(v(t), w_r(t)) < \rho_t \}$$

를 만족하는 입력벡터 $v(t)$가 충분히 조밀하게 분포한다면 "Strong Law of Large Numbers"에 의해 다른 결론을 유도 할 수 있다[19].

Assumption 2 :

i. 입력벡터 $v(t)$는 최적 Weight vector w_r^*의 Neighborhood 주위에 충분히 조밀하게 분포하고 있다. i.e.

$$\forall v(t) \in R^n \ \& \ \rho_t > 0 \ s.t. \ v(t) \in B^o(v(t), w_r^*) = \{ v(t) \in R^n | d(v(t), w_r^*) < \rho_t \} \ as \ t \to \infty$$

ii. 입력벡터 $v(t)$는 Gaussian 분포를 따른다.

iii. 입력벡터 $v(t)$는 variance $\|D\| > 0$ 를 가지면서 Identically, Independent Distribution이다.

Assumption 2의 첫 번째 가정은 논의의 대상을 충분히 많은 입력벡터 $v(t)$에 대하여 제한한 것이고 두 번째 가정은 매우 많은 데이터의 경우 일반적으로 나타나는 분포가 Gaussian 분포이기 때문이다[20]. 세 번째 가정은 입력 데이터의 분포가 White Noise에 의해 교란되어 나타나는 것임을 의미한다. 이것은 첫 번째와 두 번째 가정이 성립할 때 일반적으로 나타나는 성질이기도 하다.

증명에 들어가기 앞서 제안한 알고리즘이 가지는 일반적인 성질에 따르는 가정을 두자.

Assumption 3 :

i. $v(t) - w_r(t) \ge 0$일 확률은 $t \to \infty$에서 $P\{ v(t) - w_r(t) \ge 0 \} = 0.5$ 이다.

ii. $t \to \infty$에서 최적 Weight vector w^* 주변에는 강화함수의 영향이 없다. i.e. $\lim_{M \to \infty} \sum_{t=0}^{M} \rho(t) < \infty$

Theorem 2 : 시불변 학습율과 이진 강화함수를 가진 SOFM 알고리즘은 일정시간 이후 w_r^* 근방으로 간다.

Lemma 1 : 이진 강화함수항의 Ensemble합은 최적 Weight vector w_r^*주변에서 유계를 가진다. i.e.

$$\exists R > 0 \ \ s.t. \ \lim_{M \to \infty} E \sum_{t=n_0}^{M} \delta\rho(t) sgn(v(t) - w_r(t)) < R$$

Proof of Lemma 1 :

Assumption 3의 i 의 조건을 약화시켜 $P\{ v(t) - w_r(t) \ge 0 \} = 0.5 + \eta$ 라 하자. 여기서 η는 $0 < |\eta| < 0.5$인 값이다.

$$\begin{aligned} &\lim_{M \to \infty} E \sum_{t=n_0}^{M} \delta\rho(t) sgn(v(t) - w_r(t)) \\ &= \lim_{M \to \infty} \sum_{t=n_0}^{M} \delta\rho(t) [P(v(t) - w_r(t) > 0) - P(v(t) - w_r(t) \le 0)] \\ &= \lim_{M \to \infty} \sum_{t=n_0}^{M} \delta\rho(t) [0.5 + \eta - (1 - 0.5 - \eta)] \\ &= \lim_{M \to \infty} \sum_{t=n_0}^{M} \delta\rho(t) 2\eta \end{aligned}$$

가정 3의 ii조건에서 $\rho(t)$는 유한번 1을 발생한다. 따라서

$$\exists 0 < n < \infty \Rightarrow \lim_{M \to \infty} \sum_{t=n_0}^{M} \delta\rho(t) 2\eta < 2\delta n\eta < \infty \quad \blacksquare$$

Proof of Theorem 2:

제안한 학습 방정식은 다음과 같다.

$$w_r(t+1) = w_r(t) + \varepsilon(t) h_{rs}(t) (v(t) - w_r(t)) + \delta \rho(t) sgn(v(t) - w_r(t))$$

그리고 최적 Weight vector w^* 에서의 학습 방정식은 다음과 같다.

$$w_r^* = w_r^* + \varepsilon(t) h_{rs}(t) (v(t) - w_r^*) + \delta'$$

이 경우 최적 Weight vector w^* 와 시간 t+1에서의 Weight vector $w_r(t+1)$과의 차분치는 다음과 같다.

$$w_r^* - w_r(t+1) = (1 - \varepsilon(t) h_{rs}(t)) (w_r^* - w_r(t)) - \delta \rho(t) sgn(v(t) - w_r(t)) + \delta' \quad (10)$$

여기서 δ'는 $|\delta'| \ll 1$인 유계를 갖는 임의의 작은 값이다. 이 값은 최적 Weight Vector가 $-\nabla_{w_r(t)} E = \varepsilon h_{rs}(t) (v(t) - w_r^*)$의 영향을 받지 않도록 보정 하는 값이다. 따라서 시간 t가 무한대에 접근할 때 δ'가 축적되어 $w_r^* - w_r(t)$이 발산하는가를 알아보면 제안한 알고리즘의 안정성을 알 수 있게 된다. 학습 계수 $\varepsilon(t)$를 일정한 값 ε으로 놓고 (10)을 전개하면 다음식을 얻는다.

$$\begin{aligned} w_r^* - w_r(t+1) &= (1 - \varepsilon h_{rs}(t)) (w_r^* - w_r(t)) - \delta \rho(t) sgn(v(t) - w_r(t)) + \delta' \\ &\le (1 - \varepsilon h_r^M)^{(m+1)} (w_r^* - w_r(t-m)) \\ &\quad - \sum_{j=0}^{m+1} (1 - \varepsilon h_r^M)^j \delta \rho(t) sgn(v(t) - w_r(t)) \\ &\quad + \sum_{k=0}^{m+1} (1 - \varepsilon h_r^M)^k \delta' \\ &\le \sum_{i=0}^{m+1} (-1)^i (\varepsilon h_r^M)^i \frac{(m+1!)}{i!(m+1-i)!} (w_r^* - w_r(t-m)) \\ &\quad - \sum_{j=0}^{m+1} (1 - \varepsilon h_r^M)^j \delta \rho(t) sgn(v(t) - w_r(t) \\ &\quad + \sum_{i=0}^{m+1} (1 - \varepsilon h_r^M)^i \delta' \\ &\le \sum_{i=0}^{m+1} (-1)^i (\varepsilon h_r^M)^i \frac{k!}{i!(k-i)!} (w_r^* - w_r(t-m)) \\ &\quad - \sum_{j=0}^{m+1} (1 - \varepsilon h_r^M)^j \delta \rho(t) sgn(v(t) - w_r(t)) \\ &\quad + \sum_{i=0}^{m+1} (1 - \varepsilon h_r^M)^i \delta' \end{aligned}$$

여기서 $k \triangleq m+1$ 이고,

$$h_r^M = \max \{ h_{rs}(t) \in \{0, 1\} | \| h_{rs}(t) \| \} \quad \forall t \in N$$

이다.

$\frac{k!}{i!(k-i)!} \le \frac{k!}{i!}$ 이므로 위 부등식은 다음과 같이 쓸 수 있다.

$$
\begin{aligned}
& w_r^* - w_r(t+1) \\
& \le \sum_{i=0}^{m+1}(-1)^i(\varepsilon h_r^M)^i \frac{k!}{i!}\,(w_r^* - w_r(t-m)) \\
& \quad - \sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta\rho(t)\mathrm{sgn}(v(t-j)-w_r(t-j)) \\
& \quad + \sum_{i=0}^{m+1}(1-\varepsilon h_r^M)^i \delta' \\
& \le e^{-\varepsilon h_r^M k}(w_r^* - w_r(0)) + O(\varepsilon,\ h_{rs}^M) \\
& \quad - \sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta\rho(t)\mathrm{sgn}(v(t-j)-w_r(t-j)) \\
& \quad + \sum_{i=0}^{m+1}(1-\varepsilon h_r^M)^i \delta'
\end{aligned}
$$

따라서,

$$
\begin{aligned}
& \sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta\rho(t-j)\mathrm{sgn}(v(t-j)-w_r(t-j)) \\
& = \sum_{j=0}^{m+1}\sum_{k=0}^{j}(-1)^k(\varepsilon h_r^M)^k \frac{j!}{k!(j-k)!} \\
& \quad \cdot \delta\rho(t-j)sgn(v(t-j)-w_r(t-j)) \\
& = \delta\sum_{j=0}^{m+1}\rho(t-j)\mathrm{sgn}(v(t-j)-w_r(t-j)) \\
& \quad \cdot \sum_{k=0}^{j}(-1)^k(\varepsilon h_r^M)^k \frac{j!}{k!(j-k)!}
\end{aligned}
$$

여기서 $0<\varepsilon\ll 1$이므로 ε의 2차 이상의 멱급수는 멱수 $k\to\infty$에따라 0으로 수렴해 들어간다. 따라서

$$\sum_{k=2}^{j}(-1)^k(\varepsilon h_r^M)^k \frac{j!}{k!(j-k)!} \le \sum_{k=2}^{j}(-1)^k(\varepsilon h_r^M)^k \frac{j^k}{k!} \quad (11)$$

$0<\beta_j<1$이고 $\beta_j \ge \int_0^x \frac{(x-t)^j}{j!}(-1)^{j+1}e^{-x}dt\Big|_{x=-\varepsilon h_n^M}$를 만족하는 β_j를 놓으면

$$\sum_{k=2}^{j}(-1)^k(\varepsilon h_r^M)^k \frac{j!}{k!(j-k)!} \le e^{-\varepsilon h_r^M k}+\beta_j \le 1+\beta_j$$

그러므로

$$
\begin{aligned}
& \sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta\rho(t-j)\mathrm{sgn}(v(t-j)-w_r(t-j)) \\
& \le \delta\sum_{j=0}^{m+1}(1+\beta_j)\rho(t-j)\mathrm{sgn}(v(t-j)-w_r(t-j)) \quad (12) \\
& \le 2\delta\sum_{j=0}^{m+1}\rho(t-j)\mathrm{sgn}(v(t-j)-w_r(t-j))
\end{aligned}
$$

위 값의 기대값은 가정 3의 ii(i.e. $\eta=0$)와 Lemma 1에서

$$E[\,2\delta\sum_{j=0}^{m+1}\rho(t-j)\ \mathrm{sgn}(v(t-j)-w_r(t-j))] = 0$$

따라서

$$E[\sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j\delta\rho(t-j) \cdot sgn(v(t-j)-w_r(t-j))] = 0 \quad (13)$$

위 결론을 통해

$$
\begin{aligned}
& E[\ w_r^* - w_r(t+1)] \\
& \le E[\ e^{-\varepsilon h_r^M(t+1)}(w_r^* - w_r(0)) + O(\varepsilon,\ h_r^M) \\
& \quad - \sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta\rho(t)\mathrm{sgn}(v(t-j)-w_r(t-j)) \\
& \quad + \sum_{i=0}^{m+1}(1-\varepsilon h_r^M)^i \delta'] \\
& \le e^{-\varepsilon h_r^M(t+1)}(w_r^* - w_r(0)) + EO(\varepsilon,\ h_r^M) \\
& \quad - E[\sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta\rho(t)\mathrm{sgn}(v(t-j)-w_r(t-j))] \\
& \quad + E[\sum_{j=0}^{m+1}(1-\varepsilon h_r^M)^j \delta'] \\
& \le e^{-\varepsilon h_{rs}^M(t+1)}(w_r^* - w_r(0)) + O'(\varepsilon,\ h_{rs}^M) + \delta'
\end{aligned}
$$

그러므로 시불변 학습율과 이진 강화함수를 가진 SOFM 알고리즘은 일정시간 이후 w_r^* 근방으로 간다. ■

따라서 제안한 알고리즘은 exponentially stable 성질을 만족한다.

Ⅳ. 시불변 학습율과 이진 강화함수가 결합된 SOFM 알고리즘의 약수렴성(Weak Convergence)

약 수렴 (Weak Convergence)이라 함은 σ-Algebra Ω위에 정의된 측도 μ_n이 다음을 만족할 때 "측도 μ로 약 수렴한다"(Weakly Convergent to measure μ)라고 한다[5].

$$\lim_{n\to\infty}\int f(x)\mu_n(dx) = \int f(x)\mu(dx) \quad (14)$$

따라서 약 수렴성을 일반적으로 증명하기 위해서는 알고리즘을 기술하는 미분 방정식의 Diffusion equation에서 측도 혹은 Basis Process를 중심으로 (14)를 증명하여야 한다[7]. 그런데 확률공간 $(\Omega, \mathcal{F}, P)$에서는 측도가 확률로 주어지므로 $\int_R f\,d\mu_n = E[\,f(X_n)]$ 이다. 따라서 sequence X_n이 약 수렴한다 함은 Converge in Distribution임을 의미한다. 따라서 $X_n \to^d X$ 이고 $X_n \to^d X$는 $P\{X_n \in A\} \to P\{X \in A\}$ *for Borel Set* A이다. 여기서 $\to^d$는 Converge in Distribution임을 나타낸다[7][8]. 본 논문에서는 약 수렴을 증명하기 위해 Assumption 2와 Central Limit Theorem, 그리고 정의 $P\{X_n \in A\} \to P\{X \in A\}$ 를 사용한다.

1. 최적 Weight vector w_r^*와 $w_r(t)$간의 covariance

Assumption 2의 iii에서 입력벡터가 i.i.d 조건을 만족하므로 이진 강화함수 부분은 $w_r(t)$와 $\varepsilon h_{rs}(t)\,(v(t)-w_r(t))$에 대하여 independent 이다. 그러므로 이진 강화함수 부분은 Covariance에 영향을 끼치지 못하므로 시불변 학습율을 가진 SOFM 알고리즘에 대한 분석 만으로 충분하다.

Central Limit Theorem을 사용하기 위해 먼저 $\overline{u_r}(t) = w^* - w_r(t)$의 Covariance를 구하도록한다. Theorem 2에서 시 불변 학습율을 가진 SOFM 알고리즘은 다음의 두 조건을 만족한다.

$$w_r(t+1) = w_r(t) + \varepsilon h_{rs}(t)\,(v(t)-w_r(t)) \quad \ldots ①$$

$$w_r^* = w_r^* + \varepsilon h_{rs}(t)\,(v(t)-w_r^*) + \delta \quad \ldots ②$$

여기에서 $\overline{u_r}(t) = w^* - w_r(t)$로 놓으면 ① ②에서 $\overline{u_r}(t)$의 방정식을 얻는다.

$$\overline{u_r}(t+1) = (1-\varepsilon h_{rs}(t))\,\overline{u_r}(t) + \delta \quad (15)$$

위에서 Derivation을 시간에 대한 함수로 놓으면 $\delta(t)=\delta$와 같고 다음을 얻는다.

$$\delta(t) = -\varepsilon h_{rs}(t)\,(v(t) - w_r^*).$$

그러면 $\delta(t)$의 variance 와 covariance 는 i.i.d 조건에서 다음과 같다.

$$
\begin{aligned}
\mathrm{Var}(\delta(t)) &= E(\delta(t)\cdot\delta(t)) \\
&= (\varepsilon h_{rs}(t))^2\{\|D\|^2 - 2E(v(t)\,w_r^*) + \mathrm{Var}\,w_r^*\} \\
\mathrm{Cov}(\delta(t),\delta(t-k)) &= E(\delta(t)\cdot\delta(t-k)) \\
&= (\varepsilon h_{rs}(t))^2\{0 - 2E(v(t)\,w_r^*) + \mathrm{Var}\,w_r^*\}
\end{aligned} \quad (16)
$$

최적 weight vector w^* 다음을 만족한다.

$$w^* = \arg\min\{\ w(t) \in R^n | \sum_i \|v(i) - w(t)\|^2,\quad \forall w(t)\}$$

따라서 "Strong Law of Large Numbers"에 의해 최적 weight vector w^* 는 다음을 확률 1로 만족한다.

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n}v(t) = w_r^* \quad (17)$$

위 (17)에서 $\delta(t)$의 variance 와 covariance는 다음과 같다.

$$\mathrm{Cov}(\delta(t),\delta(t-k)) = (\varepsilon h_{rs}(t))^2\|D\|^2,$$

$$\mathrm{Var}(\delta(t)) = 2(\varepsilon h_{rs}(t))^2\|D\|^2$$

(15)의 연속형에서

$$
\begin{aligned}
\overline{u_r}(t+1) &= \sum_{k=0}^{t-1}(1-\varepsilon h_{rs}(t-k))^k\delta(t-k) \\
&\quad + (1-\varepsilon h_{rs}(t-k))^t\,\overline{u_r}(t)
\end{aligned} \quad (18)
$$

여기에서 $\overline{u_r}(t+1)$의 variance는 다음과 같이 구해진다.

$$
\begin{aligned}
&E((\overline{u_r}(t+1))^2) \\
&= E[\ \{ \sum_{k=0}^{t-1} \delta(t-k) \prod_{i=0}^{k} (1-\varepsilon h_{rs}(t-i)) \\
&\quad + \prod_{i=0}^{t} (1-\varepsilon h_{rs}(i))\, \overline{u_r}(0) \}^2] \qquad (19) \\
&= E[\ \{ \sum_{k=0}^{t-1} \delta(t-k) \prod_{i=0}^{k} (1-\varepsilon h_{rs}(t-i)) \}^2] \\
&\quad + 2E[\ \prod_{i=0}^{k} (1-\varepsilon h_{rs}(i))\, \overline{u_r}(0) \sum_{k=0}^{t-1} \delta(t-k) \prod_{i=0}^{k} (1-\varepsilon h_{rs}(t-i))] \\
&\quad + E[\ \prod_{i=0}^{t} (1-\varepsilon h_{rs}(i)) \prod_{j=0}^{t} (1-\varepsilon h_{rs}(j)) (\overline{u_r}(0))^2]
\end{aligned}
$$

위 방정식에서 $0<((1-\varepsilon(\sup\ h_{rs}(t),\ \forall t>0))<1$ 조건을 만족할 때 $\overline{u_r}(0)$의 극한값은 다음과 같다.

$$\lim_{k\to\infty} E[\ \prod_{i=0}^{k} (1-\varepsilon h_{rs}(i))\, \overline{u_r}(0)] = 0,$$

$$\lim_{k\to\infty} E[\ \prod_{i=0}^{k} (1-\varepsilon\, h_r^M)\, \overline{u_r}(0)] = 0$$

여기에서 h_r^M는

$$h_r^M = \max{}_r \{\ h_{rs}(t-i) \in \{0, 1\} \mid \sup h_{rs}(t-i),\ \forall t>0\ \}$$

따라서 (19)의 두 번째와 세 번째 항은 지워진다. 시간 t의 극한에서의 $\overline{u_r}(t+1)$의 variance 값이다 그러므로

$$
\begin{aligned}
&\lim_{t\to\infty} E((\overline{u_r}(t+1))^2) = \lim_{t\to\infty} E[\ \{ \sum_{k=0}^{t-1} \delta(t-k) \prod_{i=0}^{k} (1-\varepsilon h_{rs}(t-i) \}^2] \\
&\geq \lim_{t\to\infty} E[\ \{ \sum_{k=0}^{t-1} \delta(t-k)(1-\varepsilon\, h_r^M)^k \}^2] \\
&= \lim_{t\to\infty} E[\ \sum_{n=0}^{t-1} \delta(t-n)(1-\varepsilon\, h_r^M)^n \cdot \sum_{k=0}^{t-1} \delta(t-k)(1-\varepsilon\, h_r^M)^k] \\
&= \lim_{t\to\infty} E[\ \sum_{n=0}^{t-1} \sum_{k=0}^{t-1} \delta(t-n)\delta(t-k)(1-\varepsilon\, h_r^M)^{n+k}] \\
&= \lim_{t\to\infty} \sum_{n=0}^{t-1} \sum_{k=0}^{t-1} \{ \mathrm{Cov}[\ \delta(t-n)\delta(t-k)]\ (1-\chi(n-k)) \\
&\quad + \mathrm{Var}\,\delta(t)\,\chi(n-k) \}\, (1-\varepsilon\, h_r^M)^{n+k} \\
&= \lim_{t\to\infty} \sum_{n=0}^{t-1} \sum_{k=0}^{t-1} \{ (\varepsilon\, h_r^M)^2 \|D\|^2 (1+\chi(n-k)) \}\, (1-\varepsilon\, h_r^M)^{n+k} \\
&= \lim_{t\to\infty} \sum_{n=0}^{t-1} (\varepsilon\, h_r^M)^2 \|D\|^2 \frac{1-(1-\varepsilon\, h_r^M)^t}{1-(1-\varepsilon\, h_r^M)} (1-\varepsilon\, h_r^M)^n \\
&\quad + (\varepsilon\, h_r^M)^2 \|D\|^2 \frac{1-(1-\varepsilon\, h_r^M)^{2t}}{1-(1-\varepsilon\, h_r^M)^2} \qquad (20) \\
&= \lim_{t\to\infty} (\varepsilon\, h_r^M)^2 \|D\|^2 [\ \{ \frac{1-(1-\varepsilon\, h_r^M)^t}{1-(1-\varepsilon\, h_r^M)} \}^2 + \frac{1-(1-\varepsilon\, h_r^M)^{2t}}{1-(1-\varepsilon\, h_r^M)^2}] \\
&= (1 + \frac{\varepsilon\, h_r^M}{2-\varepsilon\, h_r^M}) \|D\|^2
\end{aligned}
$$

여기서 $\chi(x)$ 는 characteristic function 으로서 다음을 만족한다.

$$\chi(x) \begin{cases} x = 0 & \chi(x) = 1 \\ x \neq 1 & \chi(x) = 0 \end{cases}$$

그러므로 $\overline{u_r}(t)$의 variance 는 다음과 같다.

$$\lim_{t\to\infty} E((\overline{u_r}(t+1))^2) \geq (1 + \frac{\varepsilon\, h_{rs}^M}{2-\varepsilon\, h_{rs}^M}) \|D\|^2$$

다음을 만족하는 임의의 작은 양수 μ를 놓자.

$$\varepsilon\, h_r^M - \varepsilon h_{rs}(t-i) < \mu < h_r^M.$$

$\lim_{t\to\infty} E((\overline{u_r}(t+1))^2)$의 상계를 구하기 위하여 μ를 (20)에 대입하면 다음을 얻는다.

$$
\begin{aligned}
&\lim_{t\to\infty} E((\overline{u_r}(t+1))^2) \\
&= \lim_{t\to\infty} E[\ \{ \sum_{k=0}^{t-1} \delta(t-k) \prod_{i=0}^{k} (1-\varepsilon h_{rs}(t-i) \}^2] \\
&\leq \lim_{t\to\infty} E[\ \{ \sum_{k=0}^{t-1} \delta(t-k)(1-\varepsilon\, h_r^M + \mu)^k \}^2] \\
&= \lim_{t\to\infty} \sum_{n=0}^{t-1} \sum_{k=0}^{t-1} \{ (\varepsilon\, h_r^M)^2 \|D\|^2 (1+\chi(n-k)) \}\, (1-\varepsilon\, h_r^M + \mu)^{n+k} \\
&= \lim_{t\to\infty} (\varepsilon\, h_r^M)^2 \|D\|^2 [\ \{ \frac{1-(1-\varepsilon\, h_r^M + \mu)^t}{1-(1-\varepsilon\, h_r^M + \mu)} \}^2 \\
&\quad + \frac{1-(1-\varepsilon\, h_r^M + \mu)^{2t}}{1-(1-\varepsilon\, h_r^M + \mu)^2}] \\
&\leq (\frac{\varepsilon\, h_r^M}{(\varepsilon\, h_r^M - \mu)^2} + \frac{1}{2-\varepsilon\, h_r^M}) \varepsilon\, h_r^M \|D\|^2
\end{aligned}
$$

따라서 $\overline{u_r}(t+1)$의 variance의 극한값은 다음의 범위를 갖는다.

$$(1 + \frac{\varepsilon\, h_r^M}{2-\varepsilon\, h_r^M}) \|D\|^2 \leq \lim_{t\to\infty} E((\overline{u_r}(t+1))^2)$$

$$\leq (\frac{\varepsilon\, h_r^M}{(\varepsilon\, h_r^M - \mu)^2} + \frac{1}{2-\varepsilon\, h_r^M}) \varepsilon\, h_r^M \|D\|^2$$

학습율 ε는 $0<\varepsilon\ll 1$인 값이고 weight vector $w_r(t)$가 최적 weight vector w^*주변에 분포하면(다시말해 $w_r(t) \in B^0(w^*,\ 2\|D\|^2)$) h_r^M은 $t\to\infty$에서 1을 가진다. 따라서 $\overline{u_r}(t)$의 variance의 극한값은 input vector의 variance $\|D\|$에 비례한다.

2. 제안한 알고리즘의 약 수렴성

error value δ' 는 위에서

$$E\|\delta'\|^2 < 2(\varepsilon h(r, t))^2 \|D\|^2$$

만큼 작게 존재함을 유도 하였다. 이 결과를 사용하여 제안한 알고리즘의 약 수렴성을 증명해 본다.

Theorem 3 : 시불변 학습율과 이진 강화함수를 가진 SOFM 알고리즘은 w_r^*에 확률측도로 약 수렴한다.

Proof : $\overline{u_r}(t) = w^* - w_r(t)$ 라 놓으면 Theorem 2 에서

$$
\begin{aligned}
\lim_{t\to\infty} u(t+1) &= \lim_{t\to\infty} w^* - w_r(t+1) \\
&\leq \lim_{t\to\infty} e^{-\varepsilon h_r^M (t+1)} (w_r^* - w_r(0)) + O'(\varepsilon,\ h_{rs}^M) + \delta' \qquad (21)
\end{aligned}
$$

여기에서

$$O'(\varepsilon,\ h_r^M) = -(w^* - w_r(0)) \int_0^{\varepsilon h_r^M} \frac{(\varepsilon\, h_r^M - \tau)^{(t+1)}}{(t+1)!} e^{(-\tau)} d\tau \qquad (22)$$

논의를 간편하게 하기위해 $N \in R^n$, $N = w^* - w_r(0)$를 만족하는 N을 놓자. 가정에서 data vector $v(t)$가 이상적이면서 독립적으로 분포하므로 Assumption 2에서 $P\{\ \overline{u_r}(t) \in K_\varepsilon\ \}$ 는 gaussian분포를 따른다고 볼 수 있다. 따라서

$$
\begin{aligned}
&\lim_{t\to\infty} P\{\ \overline{u_r}(t) \in K_\varepsilon\ \} \\
&= \lim_{t\to\infty} \frac{1}{(2\pi)^{\frac{1}{2}} \|C\|} \int_{K_\varepsilon} \exp - \frac{(A-B+\delta')^2}{2\|C\|^2} du(t)
\end{aligned}
$$

$$
\begin{aligned}
&\text{where } A = E(t, w^* \cdot w_r(t)) \\
&B = N \int_0^{\varepsilon h_r^M} \frac{(\varepsilon\, h_r^M - \tau)^{(t+1)}}{(t+1)!} e^{-\tau} d\tau
\end{aligned} \qquad (23)
$$

여기서 $E(t, w^* \cdot w_r(t)) = e^{-\varepsilon h_r^M (t+1)} (w_r^* - w_r(0))$이며, $\|C\|$는 $u(t) = w^* - w_r(t)$의 covariance로서 input vector $v(t)$의 variance $\|D\|$에 대하여 (20)에 의해 $\|D\| < \|C\|$인 관계를 만족하며 다음을 만족한다.

$$\frac{1}{(2\pi)^{\frac{1}{2}} \|C\|} \int_{-\infty}^{\infty} \exp(-\frac{(x(\overline{u_r}(t)))^2}{2\|C\|^2}) du(t) = 1$$

(23)를 분석하기 위해 $x(\overline{u_r}(t))$를 다음과 같이 놓는다.

$$x(u_r(t)) = E(t, w^* \cdot w_r(t)) - N \int_0^{\varepsilon h_r^M} \frac{(\varepsilon\, h_r^M - \tau)^{(t+1)}}{(t+1)!} e^{-\tau} d\tau \qquad (24)$$

따라서 (23)은 다음과 같다.

$$
\begin{aligned}
&\lim_{t\to\infty} P\{\ \overline{u_r}(t) \in K_\varepsilon\ \} \\
&= \lim_{t\to\infty} \frac{1}{(2\pi)^{\frac{1}{2}} \|C\|} \int_{K_\varepsilon} \exp(\frac{-(x(\overline{u_r}(t)))^2}{2\|C\|^2}) du(t)
\end{aligned} \qquad (25)
$$

(24)의 우측항의 수렴성을 판정하기 위하여 다음을 만족하는 ξ_k 를 놓는다.

$$
\begin{aligned}
&\xi_k = (-1)^k (\varepsilon\, h_r^M)^k \frac{\tau^k}{k!} \\
&\because\ |N \int_0^{\varepsilon h_r^M} \frac{(\varepsilon\, h_r^M - \tau)^{(t+1)}}{(t+1)!} e^{-\tau} d\tau| = |\sum_{k=t+1}^{\infty} (-1)^k (\varepsilon\, h_r^M)^k \frac{\tau^k}{k!}|
\end{aligned} \qquad (26)
$$

따라서 ξ_k의 절대값의 극한값은 L'Hophital의 정리에 의해

다음과 같이 구해진다.

$$\lim_{k\to\infty}\xi_k=\lim_{k\to\infty}\frac{\partial\xi_k}{\partial\tau}=\lim_{k\to\infty}\frac{\partial^k\xi_k}{\partial\tau^k}=\lim_{k\to\infty}(\varepsilon h_r^{\,M})^k=0. \quad (27)$$

이는 $\{\xi_k\}$가 수렴하는sequence임을 나타낸다. ξ_k가 수렴하므로 $(\varepsilon h_r^{\,M})^{t+1}\frac{\tau^{t+1}}{t+1!}\le|\eta|\ll\varepsilon$를 만족하는 임의의 작은 값 $|\eta|$를 잡으면

$$|\eta|\ge|N\int_0^{\varepsilon h_r^M}\frac{(\varepsilon h_r^{\,M}-\tau)^{(t+1)}}{(t+1)!}e^{-\tau}d\tau|$$

따라서 $x(\overline{u_r}(t))$의 극한값은 다음과 같다.

$$\begin{aligned}&\lim_{t\to\infty}x(\overline{u_r}(t))\\&=\lim_{t\to\infty}E(t,\ w^{*}\cdot w_r(t))-N\int_0^{\varepsilon h_r^M}\frac{(\varepsilon h_r^{\,M}-\tau)^{(t+1)}}{(t+1)!}e^{-\tau}d\tau\\&=\lim_{t\to\infty}e^{-\varepsilon h_r^M(t+1)}(w_r^{*}-w_r(0))\\&\quad-N\int_0^{\varepsilon h_r^M}\frac{(\varepsilon h_r^{\,M}-\tau)^{(t+1)}}{(t+1)!}e^{-\tau}d\tau \qquad (28)\\&\le|\eta|\end{aligned}$$

(28)의 결과에서

$$\begin{aligned}&\lim_{t\to\infty}P\{\ \overline{u_r}(t)\in K_\varepsilon\ \} \qquad (29)\\&\ge\lim_{t\to\infty}\frac{1}{(2\pi)^{\frac{1}{2}}\|C\|}e^{-E\frac{\delta'^2}{2\|C\|^2}}\int_{K_\varepsilon}\exp\left(\frac{-(x(\overline{u_r}(t)))^2}{2\|C\|^2}\right)du(t)\end{aligned}$$

여기서 집합 $K_\varepsilon=\{\ x\in R^n|\,d(x,0)<\varepsilon,\ \forall\varepsilon>0\ \}$를 잡고 $\varepsilon=\delta'$로 놓는다. $d(x,0)$는 x와 0간의 measure이다. 또한 indicator function χ를 도입하여 다음과 같이 놓는다

$$\chi(x(\cdot),\eta)=\begin{cases}x(\cdot) & \text{if } x(\cdot)\in B^0(x(\cdot),\eta)\\ 0 & \text{if } x(\cdot)\notin B^0(x(\cdot),\eta)\end{cases}$$

where $B^0(x(\cdot),\eta)=\{\ x(\cdot)\in R|\,d(x(\cdot),0)\le|\eta|\ \}$

그러면 (29)는 다음과 같이 쓸 수 있다.

$$\begin{aligned}&\lim_{t\to\infty}\frac{1}{(2\pi)^{\frac{1}{2}}\|C\|}\int_{K_\varepsilon}\exp\left(\frac{-(x(\overline{u_r}(t)))^2}{2\|C\|^2}\right)du(t) \qquad (30)\\&=\frac{1}{(2\pi)^{\frac{1}{2}}\|C\|}\int_{K_\varepsilon}\lim_{t\to\infty}\exp\left(\frac{-(\chi[\ x(\overline{u_r}(t)),\ \eta\ ]^2}{2\|C\|^2}\right)du(t)\end{aligned}$$

여기에서 $t\to\infty$에 따라 $x(u(t))$의 분포는 (27)과 (28)에 의해 다음을 만족할 정도로 작아진다.

$$\begin{aligned}\{\ x(\cdot)\in R|\,d(x(\cdot),0)\ &\le|\eta|\ \}\cap K_\varepsilon\\&=\{\ x(\cdot)\in R|\,d(x(\cdot),0)\le|\eta|\ \}\\&\because(\varepsilon h_r^{\,M})^{t+1}\frac{\tau^{t+1}}{(t+1)!}\le|\eta|\ll\varepsilon\end{aligned}$$

이것은 거의 모든(almost all) $\overline{u_r}(t)$ 이 집합 $\{\ x(\cdot)\in R|\,d(x(\cdot),0)\le|\eta|\ \}$ 내에 있음을 의미한다.
따라서

$$\frac{1}{(2\pi)^{\frac{1}{2}}\|C\|}\int_{K_\varepsilon}\exp\left(\frac{-(x(\overline{u_r}(t))^2}{2\|C\|^2}\right)du(t)=1 \qquad (31)$$

(29), (31)에서

$$\begin{aligned}&\lim_{t\to\infty}\frac{1}{(2\pi)^{\frac{1}{2}}\|C\|}e^{-E\frac{\delta'^2}{2\|C\|^2}}\int_{K_\varepsilon}\exp\left(\frac{-(x(\overline{u_r}(t))^2}{2\|C\|^2}\right)du(t) \qquad (32)\\&=e^{-E\frac{\delta'^2}{2\|C\|^2}}=1-E\left[\frac{\delta'^2}{2\|C\|^2}+\sum_{n=2}^{\infty}(-1)^n\frac{1}{n!}\left(\frac{\delta'^2}{2\|C\|^2}\right)^n\right]\end{aligned}$$

(29), (32)에서 다음의 결론을 얻을 수 있다.

$$\begin{aligned}\lim_{t\to\infty}P\{\ \overline{u_r}(t)\in K_\varepsilon\ \}\ &\ge1-E\frac{\delta'^2}{2\|C\|^2} \qquad (33)\\&\ge1-\varepsilon^2h_{rs}(t)^2)1-\varepsilon\end{aligned}$$

위 결론은 vector $\overline{u_r}(t)$이 확률측도 0에 약 수렴함을 의미한다(weakly converge to Probability measure with 0). ■

V. 결론

본 논문에서는 시불변 학습율과 이진 강화함수를 가진 SOFM 알고리즘이 Exponentially Stable하고 알고리즘으로 만들어진 Sequence $\{\ w_r(t)\ \}_{t=0}^{K\infty}$가 최적 Weight vector w_r^*에 약 수렴함을 증명 하였다. 따라서 임의의 매우 작은 시 불변 학습율 ε을 잡으면 그것의 분포는 학습율의 자승에 비례하는 만큼 최적치에 가깝게 분포가 결정됨을 알 수 있었다. 이 결과는 SOFM 알고리즘을 하드웨어로 구현 할 시 별도의 곱셈기가 없어도 1 뉴런 Unit를 구성 가능함을 보이는 것이다. 그러나 현재까지의 결과는 제안한 알고리즘에서의 약 수렴 결과치가 국소 최적인지 대역 최적인지는 밝히는데 까지는 이르지 못했다. 그러므로 본 증명에서 명시한 최적치는 국소 최적치로 생각하는 것이 타당하며 앞으로 대역 최적치에, 제안한 알고리즘이 어느정도의 확률로 보다 가깝게 접근 할 수 있는가를 분석하여야 할 것이다.

참고문헌

[1] T. Poggio and F. Girosi, "Networks for approximation and learning", *Proc. IEEE*, vol. 78, no. 9, pp. 1481-1497, Sept. 1990.

[2] Z. P. and Lo B. Bavarian, "On the convergence in topology preserving neural networks" *Biol. Cybern.*, vol. 65, pp. 55-63. 1991.

[3] H. Ritter and K. Schulten, "Convergence properties of Kohonen's topology conserving maps : fluctuations, stability, and dimesion selection", *Biol. Cybern,*. vol. 60, pp. 59 71, 1988.

[4] T. Kohonen, "Self organizing map", *Proc. IEEE*, no. 78, pp. 1464-1480, 1990.

[5] Z. P. Lo, Y. Yu and 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, no. pp. 457-461, 1993.

[7] 공 성곤, "개선된 SOFM 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 and K. Schulten (ED.S.) *Models of Neural Networks Springer Verlag*, 2nd printng, 1992.

[9] V. V. Tolat, "An analysis of Kohonen's self-organizing maps using a system of energy functions", *Biol. Cybern.*, vol. 64, no. pp. 155-164, 1990.

[10] van Kampen, *Stochastic Processes in Physics and Chemistry*, N-H, 1981. Armsterdam

[11] C. Darken and J. Moody "Note on learning rate schedule for stochastic optimization "

[12] A. Gersho, "Adaptive filtering with binary reinforcement", *IEEE Trans. Inform.*, vol. IT-30, March 1984.

[13] L. Botton and P. Gallinari "A framework for the cooperation of learning algorithm", Neuro Information system vol. 2 no. 1 pp. 781-788, 1990.

[14] H. Kushner and 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, no. pp. 467- 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.

[18] 조 성원, 석 진욱, "일정 학습계수와 이진 강화함수를 가진 자기 조직화 형상지도 신경회로망", 전자공학회 논문지, 제 32 권 B편, 제 1 호, pp. 180-188.

[19] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Process I*, Springer-Verlag, 1974.

[20] Sheldon. M. Ross, *Stochastic Process,* International ed., Wiley, 1983.

석 진 욱

1969년 6월 26일생. 1993년 2월 홍익대학교 전기제어 공학과 졸업. 1995년 동 대학원 전기 공학과 졸업(석사). 1996년 6월 현재 홍익대학교 대학원 전기공학과 박사과정. 주 관심 분야는 신경회로망, 확률 프로세스, 비선형 최적화, 미분 위상 등임.

조 성 원

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