Proceedings of the 11th
KACC, October 1996

축합조건의 분석을 통한 Langevine 경쟁 학습 신경회로망의 대역 최소화 근사 해석과 필기체 숫자인식에 관한 연구

A Study of Global Minimization Analysis of Langevine Competitive Learning Neural Network Based on Contraction Condition and Its Application to Recognition for the Handwritten Numeral

°석 진 욱*, 조 성 원*, 최 경 삼*
홍익 대학교 전자 · 전기 공학부(Tel: 02-320-1493; Fax: 02-552-6093; Email: seokjeff@hitel.kol.co.ac.kr)

Abstracts In this paper, we present the global minimization condition by an informal analysis of the Langevine competitive learning neural network. From the viewpoint of the stochastic process, it is important that competitive learning gurantees an optimal solution for pattern recognition. By analysis of the Fokker-Plank equation for the proposed neural network, we show that if an energy function has a special pseodo-convexity, Langevine competitive learning can find the global minima. Experimental results for pattern recognition of handwritten numeral data indicate the superiority of the proposed algorithm.

Keywords Global Minimization, Contraction, Handwritten, Fokker-Plank Equation

1. 서 론

현재까지 신경회로망에 대한 연구는 대부분 특정 적용대상에 대한 직관적인 신경회로망 모델 개발에 중점이 맞추어져 이루어져 왔다. 따라서 범용적으로 쓰이는 몇몇 모델(오류 역전파 모델이나 자기 조직화 형상지도 모델 혹은 홉필드, 유전 알고리즘 등)을 제외하고 신경회로망 모델들의 안정성이나 수렴성 혹은 동역학적 연구는 거의 이루어지지 않은 형편이다. 그러므로 각 신경회로망 모델에 대한 동역학(Dynamics)을 파악하여 각 알고리즘이 어떠한 경우에 안정적인 동작을 하는가에 대한 정보와 각 신경회로망 모델들이 어떠한 경우 국소 수렴성을 피할 수 없는 가를 파악하는 것은 공학적으로 매우 중요한 문제라 하겠다.

본 연구에서는 기존의 경쟁학습 신경회로망보다 실험적으로 학습성능이 우수한 것으로 알려지고 하드웨어 구현에 편리한 Langevine 경쟁학습 신경회로망의 확산성과 대역 최소화 성질을 분석한다[8]. 확률 미분 방정식(SDE)의 관점에서 볼 때, Langevine 경쟁학습 신경회로망은 두가지 방법으로 분석 가능하다. 첫 번째 방법은 학습 방정식의 확률적 요소를 추출해 내어 학습 방정식의 Fokker-Plank 방정식을 유도, 분석하는 방법이며, 또 하나의 방법은 학습 방정식을 미분 다양체위의 상 미분 방정식의 해석 기법들을 사용하여 분석하는 방법이다. 첫 번째 방법은 Transition확률이 준군(Semigroup)을 이루는 특징을 사용, 시간과 위치에 관한 2차 편미분 방정식인 Fokker-Plank 방정식을 유도, 선형 준군동형사상(Linear Semigroup Operator)에 대한 해석을 통해 Weight Vector의 확산성, 안정성, 수렴성 등의 동역학적 특징을 분석하는 방법이며 두 번째 방법은 Langevine 경쟁학습 신경회로망 학습 방정식에 확률적 요소(이진 강화함수)가 있음을 통해 학습 방정식을 확률 미분 방정식으로 놓고 일반적인 상 미분 방정식의 해석 방법을 사용하여 동역학을 파악하는 방법이 있다.

본 논문에서는 먼저 Langevine 경쟁학습 신경회로망 학습 알고리즘의 Fokker-Plank 방정식을 사용한다. Fokker-Plank 방정식은 Markov 준군위에 정의된 2차 편미분 방정식이므로 선형 준군동형사상을 도입하여 Transition 확률의 축합(Contraction)이 어떤 조건 위에 성립하는가를 밝혀 대역 최소성의 조건을 밝힌다. 그리고 필기체 숫자 데이터에 대한 패턴인식실험을 통해 단위 단체(Simplex)내에 기존의 확률적 최적화 알고리즘으로 구성된 경쟁학습 (Simple Competitive Learning : SCL) 신경회로망과 성능 비교를 통해 Langevine 경쟁학습 신경회로망의 우수성을 확인하였다.

2. Langvine 경쟁학습 신경회로망 알고리즘

Langevine 경쟁학습 신경회로망의 알고리즘은 다음과 같다[1].

step 0 : Weight vector초기화

step 1 : Input vector x^{μ}에 대하여 다음 조건을 만족하는 Weight vector index r을 선택

$$r = \arg {}_{r}\min_{w_r(t)} d(v(t), w_r(t))$$

step 2 : Input vector x^{μ}과 Weight vector index r 에 대하여 Directional Derivation h_t 계산

$F(t)$: = Final Epoch-t 사이의 임의의 값을 취하는 확률변수

$P(t)$: = 0에서 t 사이의 임의의 값을 취하는 확률변수

$$\delta(t) = 1 \quad \textit{with probability } \frac{1}{2} P_{\delta}(F(t) > P(t))$$

$$\eta(\lambda, \delta, t, v_t, w_t^r) = \begin{cases} \delta(t) = 1 & -\lambda\delta(t)\, sgn(v(t) - w_r(t)) \\ \delta(t) = 0 & 0 \end{cases}$$

$$\begin{aligned} h_t &= \varepsilon_L \nabla J_t + \eta(\lambda, \delta, t, v_t, w_t^r) \\ &= -\varepsilon_L(v(t) - w_r(t)) - \lambda\delta(t)\, sgn(v(t) - w_r(t)) \end{aligned}$$

step 3 : Weight vector w^r_t 을 다음 식에 의해 갱신

$$w_r^{t+1} = w_r^t + \varepsilon_L(v(t) - w_r^t) + \delta\rho(t)\, sgn(v(t) - w_r^t)$$

step 4 : t를 t+1로 변환

step 5 : 한계 Epoch (Final Epoch)까지 진행되었으면 끝. 그렇지 않으면 Step 1.

여기에서 w^r_t는 시간 t일 때 r번째 Weight vector를 의미하며 ε_L은 시간에 대하여 불변인 일정적용 이득, v_t는 시간 t일 때 입력 Vector, 그리고 $\eta(\lambda, \delta, t, v_t, w^r_t)$는 이진 강화함수 항으로서 다음과 같이 구성된다.

$$\eta(\lambda, \delta, t, v_t, w^r_t) = -\lambda\delta(t) sgn(v_t - w^r_t) \qquad (1)$$

식 (1)에서 λ는 이진 강화함수의 이득이며, $\delta(t)$는 1혹은 0을 반환하는 Random Variable로서 본 논문에서는 Epoch에 종속인 P_{δ}의 확률로 1을 반환하며, $sgn(v_t - w^r_t)$은 다음을 만족하는 함수이다.

$$sgn(v_t - w^r_t) = \begin{cases} 1 & v_t - w^r_t \geq 0 \\ -1 & v_t - w^r_t < 0 \end{cases} \qquad (2)$$

이때 제안한 경쟁학습 신경회로망의 성능지표 J_t는 다음과 같다.

$$J_t = \frac{1}{2}\int_{B_t}\sum_{r}^{W}\| v_t - w^r{}_t\|^2 dP(v_t) \qquad (3)$$

또한 Langevine 경쟁학습 신경 회로망의 학습 알고리즘의 Fokker-Plank 방정식은 다음과 같다[1].

$$\frac{\partial P(w^r{}_t, t)}{\partial t} \qquad (4)$$
$$= \frac{\partial}{\partial w^r{}_t}[\nabla J(w^r{}_t)P(w^r{}_t, t)] + \frac{1}{2}\lambda^2 p_\delta \frac{\partial^2}{\partial^2 w^r{}_t}P(w^r{}_t, t)$$

3. Fokker-Plank 방정식을 사용한 Langevine 경쟁학습 신경회로망의 대역 최소화 근사 해석

유도된 Fokker-Plank 방정식의 해를 통한 고찰은 실제로는 국소적인 특성만을 분석한 것이다. 그러나 경쟁학습 신경회로망에서는 다수의 Weight Vector들에 의해 입력 벡터공간을 다수의 단체(Simplex)들로 분할하는 것이므로 하나의 단체를 덮을 수 있는 Neighborhood을 도입하여, Neighborhood내의 대역 최소점을 찾을 수 있는가를 분석하여야 한다. 이를 위해 먼저 Fokker-Plank방정식에 대한 Infinitestimal Operator $\mathcal{L}_t^\sigma$ 도입하고, $\mathcal{L}_t^\sigma$의 Quasi-Norm을 정의하여 Weight vector w_t^r의 Transition 확률에 대한 방정식 (Langevine 경쟁학습 신경회로망의 Fokker -Plank 방정식) $\partial_t P(w_t^r, t) = \mathcal{L}_t^\sigma P(w_t^r, t)$이 Contraction되어 있는가를 살펴 보아야 한다. 일반적으로 확산 방정식(Diffusion Equation)혹은 Fokker-Plank방정식의 해(Transition 확률)가 어떤 조건아래 Contraction 되어 있고 Transition 확률이 성능지표 $J(w_t^r)$의 미분동상 사상(Diffeomorphism)이면 Contraction 조건을 만족하는 범위내에서 대역 최소점을 찾을 수 있음이 알려져 있다[7].

Definition 1: Fokker-Plank방정식에 대한 Infinitestimal Operator $\mathcal{L}_t^\sigma$를 다음과 같이 정의한다.

$$\mathcal{L}_t^\sigma = -\frac{\partial}{\partial w_t^r}a(w_t^r, v_t) + \frac{1}{2}\sigma(t)\frac{\partial^2}{\partial^2 w_t^r} \qquad (5)$$

$a(w_t^r, v_t) = -\varepsilon\nabla J(w_t^r)$, $\sigma(t) = \lambda^2 p_\delta(t)$, $t > 0$ 이며 Fokker-Plank방정식은

$$\frac{\partial P(w^r{}_t, t)}{\partial t} - \mathcal{L}_t^\sigma P(w_t^r, t) = 0 \qquad (6)$$

Definition 2: Infinitestimal Operator $\mathcal{L}_t^\sigma$는 $P(\cdot) \in L[0,1]$인 임의의 함수 f 에 대하여 다음과 같이 Quasi-Norm을 정의한다.

$$\mathcal{L}_t = \|\mathcal{L}_t^\sigma\| := sup_{t_0 \le \tau \le t}\, \mathcal{L}_t^\sigma P(w_t^r, t) \qquad (7)$$

Infinitestimal Operator $\mathcal{L}_t^\sigma$의 Quasi-Norm $\mathcal{L}_t$는 오직 다음의 Norm 성질만 만족한다.

$$\mathcal{L}_t = 0 \quad \forall P(w_t^r, t) \in L[0,1] \quad sup_{t_0 \le \tau \le t}\, \mathcal{L}_t^\sigma P(w_t^r, t) = 0 \qquad \textbf{(8)}$$

Theorem 1: Langevine 경쟁학습 신경회로망의 Fokker -Plank방정식 $\partial_t P(w_t^r, t) = \mathcal{L}_t^\sigma P(w_t^r, t)$은 Infinitestimal Operator $\mathcal{L}_t^\sigma$의 Quasi-Norm의 시간에 대한 적분이 다음을 만족하면 $L[0,1]$에서 Contraction되어 있다.

$$\exists t > t_0,\ t \in R^+\ s.t.\ \int_{t_0}^{t}\mathcal{L}_s ds \le \log(1 - \frac{\varepsilon}{P(w_{t_0}^r, t_0)}) \qquad (9)$$
$$\forall \varepsilon \in (0, P(w_{t_0}^r, t_0))$$

Proof :

Langevine 경쟁학습 신경회로망의 Fokker- Plank방정식은 다음과 같다.

$$\frac{\partial P(w^r{}_t, t)}{\partial t} = \mathcal{L}_t^\sigma P(w_t^r, t) \le \mathcal{L}_t P(w_t^r, t) \qquad (10)$$

$\mathcal{L}_s$를 다음과 같이 정의하고

$$\mathcal{L}_s = sup_{s-\varepsilon < \kappa < s+\varepsilon}\, \mathcal{L}_s^\sigma P(w_s^r, s) \quad \forall P(w_s^r, s) \in L[0,1] \qquad (11)$$

식 (10)의 양변에 $\exp(-\int_{t_0}^{t}\mathcal{L}_s ds)$을 곱한다.

$$\frac{\partial P(w^r{}_t, t)}{\partial t}\exp(-\int_{t_0}^{t}\mathcal{L}_s ds)$$
$$-\mathcal{L}_t P(w_t^r, t)\exp(-\int_{t_0}^{t}\mathcal{L}_s ds) \le 0$$
$$\frac{\partial}{\partial t}P(w^r{}_t, t)\exp(-\int_{t_0}^{t}\mathcal{L}_s ds) \le 0 \qquad (12)$$

식 (12)를 Lebesgue 적분하면

$$\int_{P(w_{t_0}^r, t_0)}^{P(w^r{}_t, t)\exp(-\int_{t_0}^{t}\mathcal{L}_s ds)} \partial P(w^r{}_t, t)\exp(-\int_{t_0}^{t}\mathcal{L}_s ds) \le 0$$
$$\Rightarrow P(w^r{}_t, t)\exp(-\int_{t_0}^{t}\mathcal{L}_s ds) \le P(w_{t_0}^r, t_0)$$
$$P(w^r{}_t, t) \le P(w_{t_0}^r, t_0)\exp(\int_{t_0}^{t}\mathcal{L}_s ds) \qquad 13)$$

식 (13)의 양변에 초기 Transient 확률 $P(w_{t_0}^r, t_0)$를 빼고 Norm을 취하면 다음의 관계식이 성립한다.

$$P(w^r{}_t, t) - P(w_{t_0}^r, t_0)$$
$$\le P(w_{t_0}^r, t_0)\exp(\int_{t_0}^{t}\mathcal{L}_s ds) - P(w_{t_0}^r, t_0)$$
$$\|P(w^r{}_t, t) - P(w_{t_0}^r, t_0)\| \qquad (14)$$
$$\le \|P(w_{t_0}^r, t_0)[\exp(\int_{t_0}^{t}\mathcal{L}_s ds) - 1]\|$$
$$\le \|P(w_{t_0}^r, t_0)\| \cdot \|1 - \exp(\int_{t_0}^{t}\mathcal{L}_s ds)\|$$

가정에서 $\int_{t_0}^{t}\mathcal{L}_s ds \le \log(1 - \frac{\varepsilon}{P(w_{t_0}^r, t_0)})$ 이므로

$$\|P(w^r{}_t, t) - P(w_{t_0}^r, t_0)\|$$
$$\le \|P(w_{t_0}^r, t_0)\| \cdot \|1 - \exp(\int_{t_0}^{t}\mathcal{L}_s ds)\| \qquad (15)$$
$$\le \|P(w_{t_0}^r, t_0)\| \cdot \|1 - (1 - \frac{\varepsilon}{P(w_{t_0}^r, t_0)})\|$$
$$\le \varepsilon$$

가정에서 ε은 $\varepsilon \in (0, P(w_{t_0}^r, t_0))$를 만족하는 임의의 양수이므로 Langevine 경쟁학습 신경회로망의 Fokker- Plank 방정식 $\partial_t P(w^r{}_t, t) = \mathcal{L}_t^\sigma P(w_t^r, t)$ 은 $L[0,1]$에서 Contraction되어 있다. ■

Lemma 1 :

$$\forall \xi > 0,\ \exists t_\xi > t_0\ s.t.\ \int_{t_0}^{t}\frac{(w_\tau^r - w_{t_0}^r)}{\|w_\tau^r - w_{t_0}^r\|}\eta(\tau, \cdot)P(w_\tau^r, t)d\tau < \xi \qquad (16)$$

proof :

Transition확률 $P(w_\tau^r, t)$는 1 보다 작으므로 다음이 성립한다.

$$\int_{t_0}^{t}\frac{(w_\tau^r - w_{t_0}^r)}{\|w_\tau^r - w_{t_0}^r\|}\eta(\tau, \cdot)P(w_\tau^r, t)d\tau \qquad (17)$$
$$\le \int_{t_0}^{t}\frac{(w_\tau^r - w_{t_0}^r)}{\|w_\tau^r - w_{t_0}^r\|}\eta(\tau, \cdot)d\tau$$

논의를 간편히 하기 위해 $y_\tau = \frac{(w_\tau^r - w_{t_0}^r)}{\|w_\tau^r - w_{t_0}^r\|}$ 이라 놓으면, y_τ의 상한은 1 이다. 따라서

$$\int_{t_0}^{t}\frac{(w_\tau^r - w_{t_0}^r)}{\|w_\tau^r - w_{t_0}^r\|}\eta(\tau, \cdot)d\tau = \int_{t_0}^{t}y_\tau\eta(\tau, \cdot)d\tau \qquad (18)$$

Langevine 경쟁학습 신경회로망의 알고리즘 정의에서

$$\eta(\tau, \cdot) = \begin{cases}\lambda & v_\tau - w_t^r \ge 0\ \&\ \delta(\tau) = 1\\ -\lambda & v_\tau - w_t^r < 0\ \&\ \delta(\tau) = 1\end{cases} \quad 혹은$$

$$\eta(\tau, \cdot) = \begin{cases}\lambda & with\ Probability\ 0.5E\delta(t)\\ -\lambda & with\ Probability\ 0.5E\delta(t)\end{cases}$$

이므로

$$\int_{t_0}^{t} y_\tau \eta(\tau,\cdot)d\tau$$
$$=\lambda\int_{t_0}^{t} y_\tau d\tau P(v_\tau - w_\tau^r \ge 0,\ \delta(\tau)=1) \quad (19)$$
$$-\lambda\int_{t_0}^{t} y_\tau d\tau P(v_\tau - w_\tau^r < 0,\ \delta(\tau)=1)$$

그런데 $P(\delta(\tau)=1)$ 은 $\tau\to\infty$ 에 따라 $P(\delta(\tau)=1)\to 0$ 이므로 "Weak Law of Large Numbers"에 의해[8]

$\exists\ t_\xi > t_0\ s.t.\ P[y_t|\ |y_t|\ge\xi]$
$$=\int_{t_0}^{t} y_\tau d\tau P(v_\tau - w_\tau^r\ge 0,\ \delta(\tau)=1)$$
$$-\int_{t_0}^{t} y_\tau d\tau P(v_\tau - w_\tau^r < 0,\ \delta(\tau)=1) \quad (20)$$
$$=\frac{1}{2}\int_{t_0}^{t} y_\tau d\tau P(\delta(\tau)=1)-\frac{1}{2}\int_{t_0}^{t} y_\tau d\tau P(\delta(\tau)=1)$$
$$=\frac{1}{\lambda}\int_{t_0}^{t} y_\tau \eta(\tau,\cdot)d\tau < \xi$$

따라서 $\xi>0$ 인 임의의 ξ에 대하여

$\int_{t_0}^{t}\frac{(w_\tau^r-w_{t_0}^r)}{\|w_\tau^r-w_{t_0}^r\|}\eta(\tau,\cdot)P(w_\tau^r,t)d\tau<\xi$ 을 만족하는 $t_\xi>t_0$인 t_ξ가 존재한다. ■

Assumption 1 : 성능지표 $J(w_t^r)$는 Lipschitz 조건을 만족한다 i.e.

$$\exists\ K>0\quad s.t.\quad \|J(w_{t_1}^r)-J(w_{t_2}^r)\|\le K\|w_{t_1}^r-w_{t_2}^r\|$$
$$\forall J(w_t^r)\in R,\ \forall\, w_t^r\in R^n$$

Lemma 2: R^+, $\forall\varepsilon\in(0,P(w_{t_0}^r,t_0))$에 대하여 성능지표 $J(w^r_t)$가 Convex함수이고 Assumption 1을 만족할 때 다음을 만족하는 $t>t_0$ 이며 R^+의 한 원소인 t 가 존재한다.

$$\int_{t_0}^{t}\mathcal{L}_s ds \le \log\left(1-\frac{\varepsilon}{P(w_{t_0}^r,t_0)}\right) \quad (21)$$

proof :
위 명제가 거짓이다 즉,

$\forall t>t_0,\quad \int_{t_0}^{t}\mathcal{L}_s ds > \log\left(1-\frac{\varepsilon}{P(w_{t_0}^r,t_0)}\right)$이라 가정하자.

Langevine 경쟁학습 신경회로망의 확률미분방정식(SDE)형식에서

$$\partial w^r_t = -\varepsilon_L\nabla J(w^r_t)dt + \eta(t,\cdot)dt$$
$$w_t^r-w_{t_0}^r = -\int_{t_0}^{t}\varepsilon_L\nabla J(w^r_t) - \eta(t,\cdot)dt \quad (22)$$

양변에 $(w_t^r-w_{t_0}^r)P(w_t^r,t)$을 곱하면

$$\|w_t^r-w_{t_0}^r\|^2P(w_t^r,t) = -\int_{t_0}^{t}\big[\varepsilon_L(w_\tau^r-w_{t_0}^r)\nabla J(w^r_\tau) - (w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau \quad (23)$$

Theorem 1과 Convex조건에 의해 Langevine 경쟁학습에 의해 생성되는 수열 $\{w_\tau^r\}_{t_0}^{t}$에 대하여

$\forall\tau>t_0\quad J(w_{t_0}^r)-J(w_\tau^r)\ \ge\ (w_{t_0}^r-w_\tau^r)\nabla J(w_\tau^r)$ 이므로

$$\|w_t^r-w_{t_0}^r\|^2P(w_t^r,t)$$
$$=\int_{t_0}^{t}\big[\varepsilon_L(w_{t_0}^r-w_\tau^r)\nabla J(w^r_\tau) + (w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau$$
$$\le\int_{t_0}^{t}\big[\varepsilon_L(J(w_{t_0}^r)-J(w_\tau^r)) + (w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau$$
$$\le\int_{t_0}^{t}\big[\varepsilon_L\|J(w_{t_0}^r)-J(w_\tau^r)\| + (w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau \quad (24)$$

또한 Lipschtz 조건에서

$$\int_{t_0}^{t}\big[\varepsilon_L\|J(w_{t_0}^r)-J(w_\tau^r)\|+(w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau$$
$$\le\int_{t_0}^{t}\big[\varepsilon_L K\|w_\tau^r-w_{t_0}^r\|+(w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau \quad (25)$$

그러므로

$$\|w_t^r-w_{t_0}^r\|^2P(w_t^r,t)$$
$$\le\int_{t_0}^{t}\big[\varepsilon_L K\|w_\tau^r-w_{t_0}^r\|+(w_\tau^r-w_{t_0}^r)\eta(\tau,\cdot)\big]P(w_\tau^r,\tau)d\tau$$

따라서

$$\|w_t^r-w_{t_0}^r\|P(w_t^r,t)$$
$$\le\int_{t_0}^{t}\left[\varepsilon_L K+\frac{(w_t^r-w_{t_0}^r)}{\|w_t^r-w_{t_0}^r\|}\eta(\tau,\cdot)\right]P(w_\tau^r,\tau)d\tau \quad (26)$$

식 (15)에서 $P(w^r_t,t)\le P(w_{t_0}^r,t_0)\exp(\int_{t_0}^{t}\mathcal{L}_s ds)$이고 $P(w_{t_0}^r,t_0)\exp(\int_{t_0}^{t}\mathcal{L}_s ds)>0$ 이므로

$$\|w_t^r-w_{t_0}^r\|P(w_\tau^r,t)$$
$$\le\int_{t_0}^{t}\varepsilon_L KP(w_{t_0}^r,t_0)\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau$$
$$+\int_{t_0}^{t}\frac{(w_\tau^r-w_{t_0}^r)}{\|w_\tau^r-w_{t_0}^r\|}\eta(\tau,\cdot)P(w_\tau^r,t)d\tau \quad (27)$$
$$\le\varepsilon_L KP(w_{t_0}^r,t_0)\int_{t_0}^{t}\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau$$
$$+\int_{t_0}^{t}\frac{(w_\tau^r-w_{t_0}^r)}{\|w_\tau^r-w_{t_0}^r\|}\eta(\tau,\cdot)P(w_\tau^r,t)d\tau$$

가정에서 $\forall t>t_0,\ \int_{t_0}^{t}\mathcal{L}_s ds > \log\left(1-\frac{\varepsilon}{P(w_{t_0}^r,t_0)}\right)$이므로

$\int_{t_0}^{t}\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau$ 의 하한은

$$\int_{t_0}^{t}\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau$$
$$>\int_{t_0}^{t}\left(1-\frac{\varepsilon}{P(w_{t_0}^r,t_0)}\right)d\tau=\left(1-\frac{\varepsilon}{P(w_{t_0}^r,t_0)}\right)(t-t_0) \quad (28)$$

이다.Lemma1에서 $\lim_{t\to\infty}\int_{t_0}^{t}\frac{(w_\tau^r-w_{t_0}^r)}{\|w_\tau^r-w_{t_0}^r\|}\eta(\tau,\cdot)P(w_\tau^r,t)d\tau\to 0$ 이므로

$$\|w_t^r-w_{t_0}^r\|P(w_\tau^r,t)$$
$$>\int_{t_0}^{t}\frac{(w_\tau^r-w_{t_0}^r)}{\|w_\tau^r-w_{t_0}^r\|}\eta(\tau,\cdot)P(w_\tau^r,t)d\tau=H(t_\xi)$$
$$>\int_{t_0}^{t_\xi+t_o}\frac{(w_\tau^r-w_{t_0}^r)}{\|w_\tau^r-w_{t_0}^r\|}\eta(\tau,\cdot)P(w_\tau^r,t)d\tau\quad\forall t_o>0$$

되는 t_ξ를 취할 수 있고, 식 (27)은 다음과 같이 쓸 수 있다.

$$\frac{\|w_t^r-w_{t_0}^r\|P(w_\tau^r,t)}{\varepsilon_L KP(w_{t_0}^r,t_0)}\le\int_{t_0}^{t}\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau+\frac{H(t_\xi+t_o)}{\varepsilon_L KP(w_{t_0}^r,t_0)}$$

그런데 $\varepsilon\to P(w_{t_0}^r,t_0)$ 에 따라

$inf\int_{t_0}^{t}\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau\to 0^+\quad\forall t>t_0\ ,\ t\in R^+$ 이므로

$$\frac{\|w_t^r-w_{t_0}^r\|P(w_\tau^r,t)}{\varepsilon_L KP(w_{t_0}^r,t_0)}-\frac{H(t_\xi+t_o)}{\varepsilon_L KP(w_{t_0}^r,t_0)}>\int_{t_0}^{t}\exp\left(\int_{t_0}^{\tau}\mathcal{L}_s ds\right)d\tau \quad (29)$$

되는 t가 존재한다. 이는 가정에 모순. 따라서 다음을 만족하는 $t>t_0$인 t가 존재한다.

$$\int_{t_0}^{t}\mathcal{L}_s ds\le\log\left(1-\frac{\varepsilon}{P(w_{t_0}^r,t_0)}\right).\ ■$$

Corollary 2 : R^+, $\forall\varepsilon\in(0,P(w_{t_0}^r,t_0))$에 대하여 성능지표 $J(w^r_t)$가 다음과 같은 Pseudo-Convex함수일 때

식 (21)을 만족하는 $t > t_0$ 이며 R^+의 한 원소인 t 가 존재한다.

$$\forall \tau > t_0 \quad J(w^r_{t_0}) - J(w^r_\tau) \ge (w^r_{t_0} - w^r_\tau)\nabla J(w^r_\tau) - o(\|w^r_\tau\|) \tag{30}$$

여기서 $o(\|w^r_t\|)$은 $(w^r_t, w^r_t + dw^r)$ 범위에서 (30)을 만족할 수 있는 $o(\|w^r_t\|)$의 절대값의 하한값을 가지며 다음의 두 특성을 만족한다.

$$\int_{t_0}^{t} o(\|w^r_\tau\|)P(w^r_\tau, \tau)d\tau \le 0 \qquad o(\|w^r_t\|) \to 0 \ as \ w^r_t \to w^r_*,$$

w^r_*는 대역 최소점(Globally Optimal point)이다.

proof :

Lemma 2의 증명에서 $o(\|w^r_t\|)$항을 포함하여 정리하면

$$\begin{aligned} &\|w^r_t - w^r_{t_0}\| P(w^r_\tau, t) \\ &\le \varepsilon_L \, KP(w^r_{t_0}, t_0) \int_{t_0}^{t} \exp\left(\int_{t_0}^{\tau} \mathcal{L}_s ds\right) d\tau \\ &\quad + \int_{t_0}^{t} o(\|w^r_\tau\|)P(w^r_\tau, t)d\tau \\ &\quad + \int_{t_0}^{t} \frac{(w^r_\tau - w^r_{t_0})}{\|w^r_\tau - w^r_{t_0}\|}\eta(\tau, \cdot)P(w^r_\tau, t)d\tau \end{aligned} \tag{31}$$

가정에서 $\int_{t_0}^{t} o(\|w^r_\tau\|)P(w^r_\tau, \tau)d\tau \le 0$의 최대값은 0이므로 (31)은 (27)과 동일해 진다. 이후의 증명은 Lemma 2의 증명과 동일하다. ■

Theorem1, Lemma 2, Corollary 2의 결론에서 성능지표 $J(w^r_t)$가 단위 단체(Simplex)에서 Convex이거나,Corollary 2의 제한조건을 만족하는 Pseudo-Convex함수의 경우 단위 단체(Simplex)내의 대역 최소점(Globally Optimal point)을 취할 수 있음을 알 수 있다.

4. 실험결과

실험에 사용한 데이터는 필기체 숫자 데이터로서 홍익대학교에 재학중인 남녀 학생 100명에게 0에서 9까지 크기 방향 위치가 다양하게 변형된 필기체 숫자를 받았다. 이 중 무 작위로 50명의 데이터를 학습용 데이터, 나머지 50명의 데이터를 Test용 데이터로 사용하였다. 이렇게 얻어진 필기체 데이터를 전처리 과정을 거쳐 신경 회로망의 입력으로 인가 하였다. 전처리 과정은 먼저 수집된 원 데이터를 CCD카메라와 Frame Grabber를 통해 128×128 Pixel과 128 Gray Level영상을 구한 후 Thresholding을 통해 경계면을 구한 후 구하여진 경계면을 따라 Fourier Descriptor 계수를 구하여 이를 신경 회로망의 입력 벡터로 가져갔다. 입력 벡터로 사용한 Fourier Descriptor계수는 12개이며 Weight Vector수는 72개이다.

제안한 알고리즘에서 조정해주어야 할 파라메타는 경쟁학습 신경회로망의 경우 식 (32)와 같이 놓았으며 Langevine 경쟁학습의 경우 식 (33)과 같이 놓았다. 초기 Weight vector 선정은 각 클레스 당 같은 개수의 데이터 벡터들로 이루어 지도록 했으며 Weight vector 수는 20개이다. 실험환경은 Pentium-100 IBM-PC에서 Visual C++ 1.5 컴파일러로 행하였으며 MFC함수를 사용, Windows Application으로 프로그램 하였다.

$$\varepsilon(t) = 0.9 \cdot \left(1 - \frac{t}{Number\ of\ Iteration}\right) \tag{32}$$

$$\varepsilon = 2^{-4} = 0.0625 \ , \ \lambda = 2^{-5} = 0.03125 \tag{33}$$

식 (33)에서 ε은 시간에 불변인 학습계수이며 λ는 이진 강화함수항의 이득이다.

실험결과 제안한 알고리즘이 학습 및 테스트 양면에서 기존의 경쟁학습 신경회로망 알고리즘보다 평균적으로 더 나은 성능을 보였다. 이는 성능지표의 Convexity가 떨어질 경우 Corollary 2의 결론을 통해 어느정도의 국소 최소점(Locally Optimal point)은 Langevine 경쟁학습 신경회로망이 극복 할 수 있음을 알 수 있다.

7. 결 론

본 논문에서는 먼저 Langevine 경쟁학습 신경회로망의 학습 알고리즘의 Fokker-Plank 방정식을 사용하여, Fokker-Plank 방정식이 Markov 준군위에 정의된 2차 편미분 방정식임을 이용, 선형 준군동형사상을 도입하여 Transition 확률의 축합(Contraction)이 어떤 조건 위에 성립하는가를 밝혀 대역 최소성의 조건을 밝혔으며. 실험을 통해 해석 결과를 살펴 보고 증명된 해석 조건들의 타당성을 실험을 통해 검증하였다.

본 논문에서 유도된 단위 단체(Simplex)들에 대한 대역 최소화 충분조건들은, 그러나 상당히 강한 조건들로서 $\int_{t_0}^{t} \mathcal{L}_s ds \to -\infty$가 만족 되려면 $t \mapsto \infty$임을 의미한다. 따라서 현재까지 알려진 거의 유일한 대역 최소화 알고리즘인 "Simulated Annealing"의 경우에서와 마찬가지로 유한번의 Epoch로 대역 최소점을 찾아내기는 사실상 어렵다는 것을 의미한다. 앞으로의 과제로는 본 논문에서 논한 분석이 매우 약한 위상에서의 분석임에 따라 좀 더 강한 위상에서, 전통적인 확률 미분 방정식의 해석방법으로 재 해석하여 본 알고리즘이 구체적으로 어떤 Nonconvex 조건까지를 만족 하면서 대역 최소점에 도달 할 수 있는지를 밝히는 것과, Homology 분석을 통해 초기 Weight Vector들의 입력공간 분할에 대한 해석을 하는 것이 요구된다.

Epoch	경쟁학습		Langevine 경쟁학습	
	학습 (%)	테스트 (%)	학습 (%)	테스트 (%)
100	88.44	73.56	86.89	75.11
200	90.00	72.00	90.22	79.56
300	86.89	73.56	90.89	78.22
400	87.56	71.11	89.33	77.78
500	87.78	72.44	91.56	77.11
600	89.33	72.223	90.22	78.22
700	88.89	75.11	90.22	76.00
800	88.89	74.00	91.78	79.56
900	88.89	72.00	90.44	78.22
1000	90.00	76.44	92.00	78.89
평균	88.667	73.244	90.355	77.867

<표 1>

필기체 숫자 인식 데이터에 대한 실험 결과

참 고 문 헌

[1] J. Seok, S. Cho, *"Self-Organizing Feature Map with Binary Reinforcement and Constant Adaptation Gain : For an easier Hardware Implementation"*, Proc. ICONIP'94, vol 2, 966-971 1, 1994.

[2] I.I.Gihman, A.V.Skorohod, *The theory of stochastic process I*, Springer-Verlag, 1974.

[3] H. Ritter, K. Schulten, *"Convergence Properties of Kohonen's Topology Conserving Maps : Fluctuations, Stability, and Dimension Selection"*, *Biological Cybernetics*, vol 60, pp. 59-71, 1988.

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

[5] Yu. Ermoliev, R. J-B Wets (Eds.), *'Numerical Techniques for Stochastic Optimization'*, *Springer-Verlag*, 1980.

[7] T. S. Chiang, C. R. Hwang, S. J. Sheu, *"Diffusion for Global Optimization in R^n"*, *SIAM J. Cont. and Optimization*, vol 25, No 3, pp 737-753, May 1987,

[8] P. Billingsley, *Probability and Measure*, Wiley, 3rd ED., 1994.

[9] 석 진욱, 조 성원, 최경삼, *'Langevine 경쟁학습 신경회로망의 확산성과 대역 최소화 성질의 근사해석'*, Proc 96, 대한 전기학회 하계 학술대회 논문집, vol B, pp 1344-1346, 1996