'94 대한전기학회 추계학술대회 논문집(1994. 11. 18)

일정적응 이득과 이진 강화함수를 갖는 경쟁 학습 신경회로망

°석 진 욱　　조 성 원　　최 경삼
*홍익 대학교 전기제어공학과

Competitive Learning Neural Network with Binary Reinforcement and Constant Adaptation Gain

°Jinwuk Seok　Seongwon Cho　Gyung-sam choi
*Dept. of Electrical and Control Eng. Hong Ik Univ.

Abstract

A modified Kohonen's simple Competitive Learning(SCL) algorithm which has binary reinforcement function and a constant adaptation gain is proposed. In contrast to the time-varing adaptaion gain of the original Kohonen's SCL 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 SCL due to the constant adaptation gain. Since the proposed algorithm does not have the complicated multiplication, it's digital hardware implementation is much easier than one of the original SCL.

I.서 론

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. 따라서 디지탈 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. 본 논문은 신경회로망 알고리즘의 일종인 Simple Competitive Learning(SCL) 알고리즘을 하드웨어 구현에 편리하게하기 위해 학습시간에 따른 단조감소 함수인 적용 이득을 일정한 값으로 놓고 이를 보상하기 위하여 이진 강화 함수를 결합한 알고리즘을 제안하고 패턴인식면에서 기존의 알고리즘과 성능비교를 하였다. 본 논문은 다음의 순서로 구성되어 있다. 먼저 SCL 알고리즘의 기본개념과 성질을 서술하고 이진 강화함수에 대하여 논하였다. 다음으로 본 논문에서 제안한 알고리즘를 설명하고 제안한 알고리즘의 수렴성을 증명하고 8 클래스를 가진 패턴인식데이타에 대하여 기존 알고리즘과 성능을 비교하였다.

II.Simple Competitive Learning (SCL) 신경회로망

1. SCL 신경회로망의 특징

SCL신경회로망은 비지도학습법(Unsupervised Learning)을 이용하는 대표적인 신경회로망으로서 각 뉴런에 대응하는 weight vector들이 해당 데이타의 Cluster중심으로 찾아 들어가게하는 특징을 지닌다. SCL신경회로망의 학습 알고리즘은 다음과 같다.

$$w_r(t+1) = w_r(t) + \varepsilon\ h(r,t)(v(t) - w_r(t)) \qquad (1)$$

여기서 $\varepsilon(t)$는 적용 이득(Adaptive gain), $h(r,t)$는 r번째 Neuron에 해당하는 Weight Vector가 입력벡터 v(t)에 대하여 winning이면 1, 그렇지 않으면 0을 가리키는 Index 함수, $w_r(t)$는 시간 t에서의 r번째 Weight Vector, v(t)는 입력 벡터이며, 특히 $\varepsilon(t)$의 경우 시간에 따른 단조감소 함수로서 Kohonen은 다음과 같은 감소 함수를 제안 하였다

$$\varepsilon(t) = 0.9 \cdot \left(1 - \frac{t}{Number\ of\ Iteration}\right) \qquad (2)$$

2. 일정 적응 이득을 가진 SCL 신경회로망

적용 이득이 (2)식과 같은 경우 SCL알고리즘에서는 Weight Vector가 입력데이타내 임의 Cluster의 중심에 해당되는 최적 Weight Vector로 수렴한다는 것이 여러사람들에 의하여 증면 된바 있다. 반면 적응이득이 일정할 경우, 특히 적응이득이 0에 접근하는 일정한 값일 경우 Weight Vector는 적응이득과 입력데이타에 비례하는 정도의 오차를 가지고 최적 Weight Vector주변에 분포하게 된다. 이를 살펴보면 먼저 SCL의 Energy 함수를 다음과 같이 정의하고 이를 Weight Vector w_r에 대하여 미분하면 다음 식을 얻는다.

$$E(w) = \frac{1}{2}\sum_r h(r,t)(v(t)-w_r(t))^2$$
$$\frac{\partial E}{\partial w_r} = -\varepsilon\ h(r,t)(v(t) - w_r(t)) \qquad (3)$$

따라서 다음 식이 성립한다.

$$[\ v(t+1) - w_r(t+1)] = (1-\varepsilon\ h(r,t))(v(t) - w_r(t)) \quad (4)$$

이것을 전개하면

$$\begin{aligned}[\ v\ (t+1) - w_r(t+1)] &= (1-\varepsilon\ h(r,t))(v(t) - w_r(t)) \\ &= (1-\varepsilon\ h(r,t))^m(v(t-m) - w_r(t-m)) \\ &= \sum_{i=0}^{m}(-1)^m(\varepsilon\ h(r,t))^m \frac{m!}{i!(m-i)!}(v(t-m) - w_r(t-m))\end{aligned}$$

$$\left(\sum_{i=0}^{m}(-1)^{m}(\varepsilon\ h(r,t))^{m}\frac{k^{m}}{i!}(v(t-m)-w_r(t-m))\right.$$

$$;\ k \triangleq m+1$$

$$[\ v(t+1)-w_r(t+1)]$$

$$\begin{aligned}&\le e^{-\varepsilon\, h(r,t)k}(v(t-m)-w_r(t-m))+O(\varepsilon,\ h(r,t)) \\ &\le e^{-\varepsilon\, h(r,t)(t+1)}(v(0)-w_r(0))+O'(\varepsilon,\ h(r,t))\end{aligned} \tag{5}$$

즉 시간이 지남에 따라 입력 벡터와 Weight Vector간의 차이는 적응이득이 일정한 값을 유지하고 있다 하더라도 Exponentially하게 줄어듦을 알 수 있다. 따라서 적응 이득이 일정하다 하더라도 Exponential Stable은 보장됨과 동시에 $t\to\infty$ 에서 Martingale Sequence임을 알 수 있다. 단, 적응이득이 일정한 SCL알고리즘의 경우 Weight Vector는 최적 Weight Vector주변에 적응이득에 비례하는 어떤 오차를 지니면서 분포함을 알 수 있다.

3. 이진 강화 함수

이진강화 함수는 $w(t)$를 보다 평균제곱 오차를 최소화하는 대역 최소점(Global Minimum Point)으로 보내게 하는데 쓰인다. 이진강화함수를 이용한 적응 필터 알고리즘은 다음과 같다.[12]

$$\begin{aligned} c_{k+1} &= c_k - \alpha u_k\, sgn[\ e_k] \\ e_k &= c_k \cdot u_k - d_k \\ u_k &= \varphi(x_k, x_{k-1}, \ldots) \end{aligned} \tag{6}$$

φ : 관찰 과정(Observable Process) x_k 를 데이타 과정 u_k로 보내는 사상

d_k : 요구 과정(Desirable Process)

α : 일정 적응 이득 (Constant Adaptive gain)

여기에서 $u_k = 1$로 놓고 $d_k = c^* \cdot u_k + f_k$로 놓으면 f_k는 데이타 신호가 된다. c_k와 d_k가 동일 공간상에 존재한다면 이것은 확률적 최급 강하법의 한 종류 가 된다.[13] 위 방정식을 고쳐쓰면 다음과 같다.

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

x_k : 데이타 벡터 열

w^* : w의 최적 평형 벡터

n_k : 오차 혹은 잡음 벡터 열

$\rho(t)$: 시간에 따라 1의 발생확률이 감소하는 랜덤변수

$$\rho(t)\to 0 \quad as \quad t\to\infty$$

이진강화함수의 수렴성 판별은 다음식에 의하여 구한다.

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

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

Ⅲ.이진 강화 함수와 일정 적응 이득이 결합된 SCL알고리즘

신경회로망 알고리즘을 위해 위에서 수정한 이진 강화 함수를 SCL알고리즘에 적용하면 다음과 같다. 먼저 SCL알고리즘에서 기대되는 Weight Vector의 변화율은 다음 식으로 주어진다.

$$E(\Delta w_r | w(t)) = \varepsilon(t)\int_{F_r(w)} (v(t)-w(t))\,P(v(t))\,dv \tag{8}$$

식 (8)에서 $F_r(w)$는 Neuron r에 해당하는 Weight Vector가 시간 t에서의 입력 데이타 v(t)에 대하여 선택될 모든 입력 데이타 v(t)공간위의 Receptive Field이다. 일반적인 신경 회로망에서의 학습 방정식은 다음과 같은 최급 강화법을 사용한다.

$$w_r(t+1) = w_r(t) - \frac{\partial E(t)}{\partial w_r(t)} \tag{9}$$

(즉 SCL알고리즘의 경우 (9)식에 (3)식을 대입하면 유도 된다.)
식 (9)에서 편미분항을 다음과 같이 선형 분리하면

$$w_r(t+1) = w_r(t) - \frac{\partial E(t)_1}{\partial w_r(t)_1} - \frac{\partial E(t)_2}{\partial w_r(t)_2} \tag{10}$$

여기서 Index 1은 식 (3)에서 주어지는 일정 적응 이득이 결합된 SCL알고리즘에서의 Weight Vector의 편미분치이고 Index 2는 이진 강화 함수에서의 Weight Vector의 편미분치 이다. 여기서 식 (10)의 Index 2가 붙은 항은 다음과 같으므로

$$-\frac{\partial E(t)_2}{\partial w_r(t)_2} \triangleq E(\Delta w_r(t)_2 | w_r(t)_2) \tag{11}$$

$$w_r(t+1) = w_r(t) - \frac{\partial E(t)_1}{\partial w_r(t)_1} - E(\Delta w_r(t)_2 | w_r(t)_2) \tag{12}$$

식 (12)에서 $E(\Delta w_r(t)_2 | w_r(t)_2)$는

$$\begin{aligned} E(\Delta w_r(t)_2 | w_r(t)_2) &= [\ w_r(t+1) - w_r(t)]\ \ P(v(t)) \\ &= -\rho(t) sgn[\ e_r(t)]\ \ P(v(t)) \end{aligned} \tag{13}$$

따라서 (13)식과 (3)식의 편미분값을 (10)식에 대입하면 (10)식은 다음의 학습 방정식으로 나타난다. 단 여기에서 $P(v(t))$는 $w_r(t)$의 Receptive Field $F_r(w_r(t))$에 해당되는 Vector로 나타날 것인가에 대한 확률이다.

$$\begin{aligned} w_r(t+1) = w_r(t) &+ \varepsilon h(r,t)(v(t)-w(t))P(v(t)) \\ &+ \rho(t) sgn[\ e_r(t)]\ \ P(v(t)) \end{aligned} \tag{14}$$

식 (14)를 해석하면 다음과 같다.

1. 뉴런 r이 승자일 경우

$$\begin{aligned} w_r(t+1) = w_r(t) &+ \varepsilon(t)(v(t)-w_r(t)) \\ &+ \rho(t) sgn(v(t)-w_r(t)) \end{aligned}$$

2. 뉴런 r이 승자가 아닐 경우

$$w_r(t+1) = w_r(t)$$

유도된 알고리즘이 적절하게 수렴 할 것인가는 Weak Convergence이론을 적용하여 증명이 가능하다.

Ⅳ. 실험결과

실험에 사용한 데이타는 미국 Indiana주 West Lafaytte시 근교 지역을 항공 촬영한 원격 탐사 데이타로서 농작물 재배 지역을 촬영한 것이다. 특히 본 데이타는 추정된 확률 분포와 실제 확률분포가 근사하여 분류 결과가 비교적 우수한 특징을 지닌다. 이 데이타는 측정 주파수 범위에 따라 8-밴드를 가지며 0에서 255사이의 정수값을 가진다. 각 클래스는 200개와 375개의 서로다른 패턴으로 구성되어 있으며 학습에는 200개의 데이타로, 학습 결과의 테스트에는 375개의 데이타를 사용했다. 각 클래스의 특징과 실험결과는 다음과 같다.

EPOCH	Original SCL Training	Original SCL Testing	SCL with Reinforcement Training	SCL with Reinforcement Testing
100	93.0	91.567	93.375	92.567
200	92.4375	90.833	93.25	91.733
300	92.5	90.933	93.0625	91.733
400	92.75	91.4	93.0625	91.433
500	92.875	91.5	93.25	92.967
600	92.5625	91.733	93.3125	93.1
700	92.625	92.133	92.6875	92.733
800	92.8125	90.6	92.4375	91.833
900	92.0	90.433	92.75	92.967
1000	92.9375	90.667	93.0625	91.767
Average	92.65	91.17	93.03	92.29

표 1
8 클래스 데이타에 대한 제안한 SCL알고리즘과 일반 SCL알고리즘과의 패턴 분류 성능 비교

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

표 2
8 Class Data의 내용

Original SCL의 경우 적응 이득은

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

이진 강화함수와 일정 적응 이득이 결합된 SCL의 경우 적응이득은 2^{-n} 의 형태로 놓아야 하드웨어 구현에 편리하므로 다음과 같이 놓았다. 랜덤값의 경우 C++ 컴파일러가 제공하는 방법을 택했다.

$$\varepsilon = 2^{-4} = \frac{1}{16} = 0.0625$$
$$\delta = 2^{-5} = \frac{1}{32} = 0.03125$$

IV.결론

본 논문에서는 디지탈 하드웨어 구현에 편리한 이진 강화함수와 일정 적응 이득을 갖는 학습 알고리즘을 제안 하였다.이진 강화함수를 학습 알고리즘에 적용 가능 하도록 변형 하였고 강화함수를 결합함으로서 얻는 이점에 대해서도 분석하였다. 이진 강화함수와 일정 적응 이득이 결합된 SCL을 8 클래스 Data에 대해서 실험한 결과 적응 이득이 충분히 작을 경우 일반 SCL알고리즘보다 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 Dimesion 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 성곤, "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 printng 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"IEEETrans.Pattern Anal.Machine Intell.,vol PAMI-4, no.4,July 1982