Proceedings of KFIS 2001 Fall Conference, 2001. 12. 1

비단조뉴런 DBM 네트워크의 학습 능력에 관한 연구

Learning Ability of Deterministic Boltzmann Machine with Non-Monotonic Neurons

박철영, 이도훈
대구대학교 정보통신공학부
Cheol-Young Park
Dept. of Computer & Communication Engineering, Taegu University
e-mail: cypark@taegu.ac.kr

Abstract

In this paper, We evaluate the learning ability of non-monotonic DBM(Deterministic Boltzmann Machine) network through numerical simulations. The simulation results show that the proposed system has higher performance than monotonic DBM network model. Non-monotonic DBM network also show an interesting result that network itself adjusts the number of hidden layer neurons. DBM network can be realized with fewer components than other neural network models. These results enhance the utilization of non-monotonic neurons in the large scale integration of neuro-chips.

I. 서 론

신경회로망의 연구에서는 단순한 입출력 함수를 갖는 뉴런모델을 이용하여 다양한 네트워크를 구축하여, 패턴인식이나 연상기억 등의 문제에 대해서 많은 연구가 수행되어 왔다. 여기서 활성화 함수로서는 일반적으로 비선형 단조함수가 이용되고 있다. 예를 들면 tanh와 같은 시그모이드(sigmoid)함수이다. 그러나 연상기억문제와 같은 특정분야에 있어서 단조가 아닌 비단조 함수를 활성화 함수로 이용함으로써 기억용량이 증가되는 것이 Morita 등에 의해 보고되었다[1]-[4]. 여기서 이용한 함수는 입력값의 절대값이 어느 한계를 넘으면 출력값이 반전하는 시그모이드 함수로 Morita형 함수로 부른다. 이 결과는 그후 여러 연구자에 의해 수학적으로 해명되고 증명되었으며 뉴런의 활성화 함수에 비단조 함수를 이용하면 보다 고기능의 신경회로망이 실현될 수 있음을 의미한다. 여기서 연상기억문제에서 다루는 신경회로망은 뉴런의 결합강도를 고정한 즉, 기억매몰형으로 학습에 의한 새로운 환경에 적응하는 것을 고려하지 않는다.

현재 학습기능을 갖는 신경회로망에 비단조 함수를 활성화 함수로 사용하는 뉴런을 이용한 연구 예는 거의 없다. 그러나 기억 매몰형의 신경회로망에 있어서 비단조 뉴런의 중요한 효과는 기억용량의 증가이지만 학습 기능을 갖는 신경회로망에 대해서도 학습 성능의 향상이 기대된다. 특히 학습기능을 갖는 신경회로망의 집적화에 있어서 결정론적 볼츠만 머신은 다음과 같은 특징을 갖고 있다.

1. 확률론적 볼츠만머신에 비해 학습시간이 빠르다.
2. 하드웨어화를 고려할 때 확률적 동작이 아니므로 노이즈 소스가 필요 없다.

본 연구에서는 학습기능을 갖는 결정론적 볼츠만 머신에 비단조 뉴런을 이용하여 그 학습성능을 수치 시뮬레이션을 통하여 분석한다. 먼저 네트워크의 은닉층에 비단조뉴런을 적용한 경우와 단조뉴런을 이용한 경우에 대하여 각각 활성화 함수로 시그모이드함수와 end-cut-off 타입의 비단조함수를 사용한 경우에 대하여 성

능을 비교한다. 실제 하드웨어로 구성하는 경우의 실현을 용이하게 하기 위하여 end-cut-off 타입의 비단조 함수는 연속함수가 아닌 구분선형함수를 사용한다. 다음으로는 네트워크의 학습성능을 학습수렴율, 학습안정도, 그리고 학습횟수의 3가지 지표를 이용하여 비교한다. 학습대상으로는 대표적인 비선형 문제인 XOR 학습에 대하여 수행한다.

II. 결정론적 볼츠만머신

볼츠만머신은 학습기능을 갖는 대표적인 신경회로망으로 대칭결합을 갖는 확률론적인 회로망에 학습규칙을 도입한 것으로 Hinton과 Sejnowski에 의해 제안되었다[5]. 특히 결정론적 볼츠만머신[6]은 확률론적 볼츠만머신에 있어서 뉴런의 출력에 평균장 근사를 이용하여 뉴런의 동작을 결정론적으로 하는 네트워크로서 Peterson과 Anderson에 의하여 제안되었다[7]. 평균장 근사에 의해 뉴런의 출력 x_i는

$$x_i = P(S_i=+1)\cdot(+1)+P(S_i=-1)\cdot(-1)$$
$$=\frac{1}{2}\left\{1+\tanh\left(\frac{u_i}{T}\right)\right\}-\frac{1}{2}\left\{1-\tanh\left(\frac{u_i}{T}\right)\right\}$$
$$=\tanh\left(\frac{u_i}{T}\right) \quad (1)$$

로 표현된다. 여기서 뉴런의 막전위 u_i는 다음 식으로 주어진다.

$$u_i = \tau\frac{du_i}{dt}+\sum_j w_{ij}x_j \quad (2)$$

이때 상태 α에 있어서 학습에 의한 하중값 갱신량 Δw_{ij}는 다음 식으로 주어진다.

$$\Delta W_{ij} = \frac{\varepsilon}{T}\sum_{\alpha}\{(x_i^{\alpha}x_j^{\alpha})_{clamped}-(x_i^{\alpha}x_j^{\alpha})_{unclamped}\} \quad (3)$$

여기서 clamped는 입출력 뉴런을 원하는 값으로 고정한 상태로 학습 phase로 부르며 unclamped는 입력 뉴런만을 고정한 상태로서 반학습 phase라 한다. 식 (3)에 의해 하중값의 갱신량 Δw_{ij}는 각 상태에 있어서 학습phase와 반학습 phase의 값 x_ix_j 의 차를 모두 더한 값에 대해서 정수(ε/T) 배이다.

III. 비단조 뉴런의 성능평가

네트워크는 그림 1과 같이 입력층 2뉴런, 은닉층 3뉴런 그리고 출력층 1뉴런으로 구성되는 2-3-1네트워크이다. 출력층 뉴런은 단조 뉴런을 사용하여 입출력의 값은 ±1로 한다. 여기서 입력층 뉴런의 활성화 함수는 $y=x$ 의 단순 선형 뉴런이다.

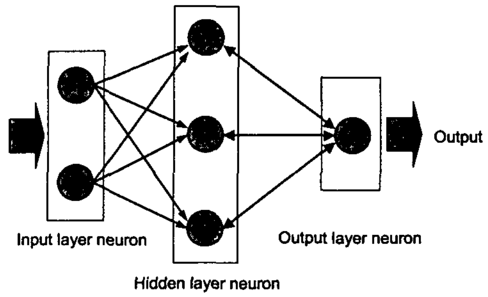


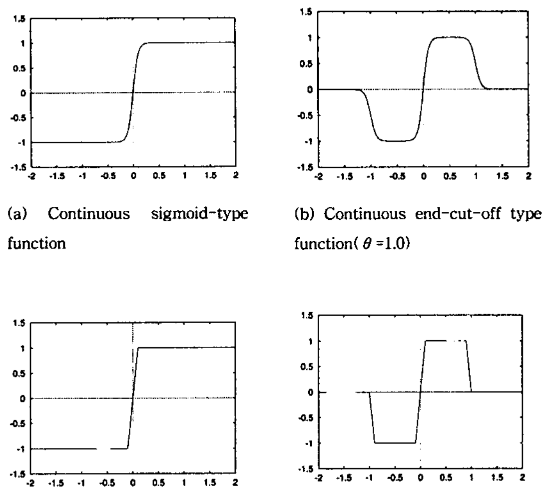
그림 1. DBM 네트워크 모델

시뮬레이션에서 학습 수렴 조건은 표 1에 나타낸 것과 같이.원하는 출력값과 네트워크의 출력값의 오차가 1% 미만이 되는 것으로 하고 최대 학습횟수는 2,000회로 한다.

표 1. 시뮬레이션 조건

온도(T)	0.1(no annealing)
하중값 갱신방법	Batching learning
학습계수(ε)	0.004
초기하중값	[-0.01, 0.01]
최대학습횟수	2000회
목표값	{-1, 1}
샘플수	200

활성화 함수로는 일반적으로 그림 2에 나타낸 것과 같이 막전위의 값이 0을 넘으면 발화하는 시그모이드 단조함수이다. 한편 그림 3은 end-cut-off 타입의 비단조함수를 나타내며 막전위의 절대값 $|u_i|$가 경계값 θ를 넘으면 출력이 0이 되는 함수이다. 이 외에 막전위의 절대값 $|u_i|$가 경계값 θ을 넘으면 출력값이 반전하는 Morita 타입 비단조 함수가 있다. 수치 시뮬레이션에서는 end-cut-off 함수를 이용하였다. 실제로 하드웨어로 구현하는 것을 고려할 때 그림 2의 (a), (b) 와같은 연속함수를 회로로 구현하는 것은 어렵기 때문에 하드웨어화를 위해서는 각각 그림 2의 (c) 및 (d) 처럼 구분선형함수를 이용한다. 이때 함수가 x축과 교차하는 부분의 경사의 절대치는 전부 1/T로 한다. 시뮬레이션에서는 그림 2에 나타낸 4가지 함수를 이용하여 단조 뉴런을 이용한 경우와 비단조 뉴런을 이용한 경우의 성능을 비교하였다.



(a) Continuous sigmoid-type function

(b) Continuous end-cut-off type function(θ =1.0)

(c) Piecewise linear sigmoid-type function

(d)Piecewise linear end-cut-off type function(θ =1.0)

그림 2. 네가지 타입의 활성화 함수

Ⅳ. 시뮬레이션

네트워크의 비선형 분리문제로서 대표적인 XOR문제의 학습을 수행하였다. XOR학습의 수치 시뮬레이션 결과로부터 학습 수렴율, 학습 안정도 및 학습횟수를 각각 그림 3~5에 나타내었다. 학습 수렴률에 관해서는 단조뉴런은 높은 학습수렴률을 나타내고 비단조 뉴런은 경계값 θ를 적당한 값으로 함으로서 동등한 학습수렴율이 얻어진다. 학습안정도에 관해서는 이제까지의 결과와 달리 비단조 뉴런의 경계값 θ를 적당한 값으로 조정함으로서 단조뉴런보다 안정도가 높게 되는 것을 알 수 있다. 실제로 시그모이드 함수를 이용한 단조 뉴런의 학습 안정도가 80%인 것에 대하여 end-cut-off 타입의 함수를 이용한 비단조 뉴런의 경우는 경계값 $\theta=1.3$ 부근에서 96%이다. 학습횟수에 관해서도 비단조뉴런의 함수 θ를 적당한 값으로 함으로써 단조뉴런의 경우와 비교하여 더 빠르게 학습이 수렴하고 있는 것을 알 수 있다. 예를들면 시그모이드 함수를 이용한 단조뉴런의 학습횟수가 143회인 것에 반하여 and-cut-off 타입의 함수를 이용한 비단조뉴런은 경계값 $\theta=1.0$ 부근에서 113회이다. 이상의 결과로부터 XOR 문제에서는 경계값 θ를 적당한 값으로 하면 단조뉴런보다는 비단조 뉴런을 이용한 네트워크가 더 좋은 학습성능을 나타내는 것을 알 수 있다. 또한 구분선형 함수에 관해서는 미분 가능한 연속함수와 비교하여 학습안정도가 조금 높고 학습횟수가 적은 것을 알 수 있다.

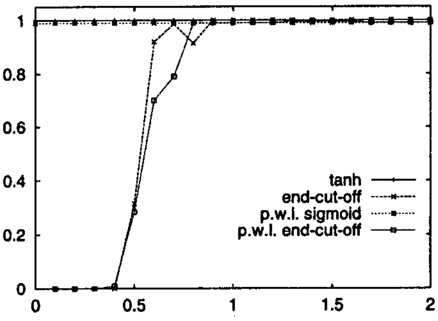


그림 3. XOR 학습에 대한 학습수렴률

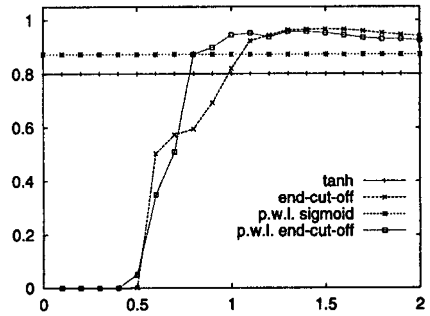


그림 4. XOR 학습에 대한 학습안정도

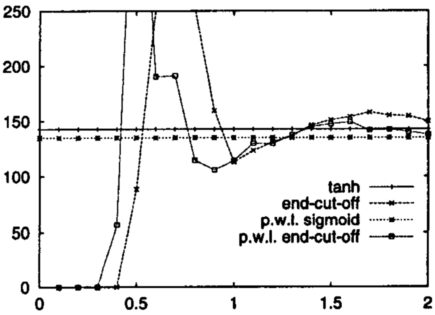


그림 5. XOR 학습에 대한 학습횟수

시뮬레이션 결과에 의해 비단조뉴런을 이용한 결정론적 볼츠만 머신이 단조뉴런을 이용한 경우보다도 꽤 높은 학습성능을 가지는 것을 확인하였다. 여기서 비단조뉴런의 활성화 함수에 사용한 end-cut-off 함수의 입출력 특성에 주목하면 식 (4)에 나타낸 것과 같이 3가지의 함수로 분해되는 것을 알 수 있다.

$$f_{eco}(u) = f_s(u) - \frac{1}{2} f_s(u-\theta) - \frac{1}{2} f_s(u+\theta) \quad (4)$$

여기서 $f_s(u)$는 시그모이드 함수, $f_{eco}(u)$는 end-cut-off 함수, 그리고 θ는 경계값을 나타낸다. 식 (4)에 의해 end-cut-off 함수의 입력응답은 조건은 있지만 3가지 시그모이드함수의 응답과 등가이다. 즉, end-cut-off 함수를 이용한 비단조뉴런에서의 학습성능은 의사적으로는 3배의 개수를 갖는 단조뉴런에서의 학습성능에 상당하는 것으로 생각된다. 따라서 비단조 뉴런을 이용한 네트워크가 높은 학습성능을 갖는 것으로 생각되며, ADD 학습과 같은 단조 뉴런을 이용한 소규모 네트워크에서는 어려운 문제에 대해서도 단조뉴런을 비단조 뉴런으로 대체함으로써 학습이 가능하다. 이 결과는 대규모 네트워크에서 보다 현저하게 확인이 가능할 것으로 생각된다. 또한 막전위 분포에 있어서는 비단조뉴런의 출력이 0이 되는 영역도 이용하여 학습을 진행하고 있는 것을 알 수 있었다. 이것은 다음 층의 뉴런에 전혀 영향을 주지 않으며 해당 뉴런이 전혀 사용되지 않는 것을 나타낸다. 즉 네트워크 자신이 뉴런의 수를 조정하고 있다고 생각할 수 있다.

V. 결론

학습기능을 갖는 대표적인 신경회로망인 결정론적 볼츠만머신의 은닉층에 비단조 뉴런을 적용한 네트워크를 구성하고, 시뮬레이션에 의해 대표적인 비선형문제인 XOR 학습을 수행하였다. 그 결과 단조 뉴런을 이용한 결정론적 볼츠만머신과 비교하여 비단조뉴런을 이용한 것이 학습수렴률, 학습안정도 및 학습횟수에 있어서 모두 성능향상이 되는 것을 확인하였다. 이것은 비단조 함수가 복수의 단조 뉴런의 합성으로 구성되는 것에 기인한다고 생각된다. 또한 네트워크의 막전위 분포를 조사함으로써 end-cut-off 함수를 이용한 비단조 뉴런 특유의 현상인 다음 층의 뉴런에 영향을 주지 않는 뉴런의 출현 즉, 신경회로망에 있어서 뉴런수의 자율적 조정작용을 확인하였다. 이것은 문제에 대한 네트워크의 뉴런수를 명확하게 결정할 수 없는 현재의 상황에 대하여 새로운 돌파구가 될 것으로 기대된다.

VI. 참고문헌

[1] M. Morita, "Memory and Learning of Sequential Patterns by Nonmonotone Neural Networks," Neural Networks, vol.9, pp.1477-1489, 1996.

[2] H. Yanai and S. Amari, " A Theory on a Neural Net with Nonmonotone Neurons," Proc. IEEE Int. Conf. Neural Networks, vol.3, pp.1385-1390, 1993.

[3] T. Fukai, J. Kim, and M. Shiino, " Retrival Properties of Analog Neural Networks and the Nonmonotonicity of Transfer Functions," Neural Networks, vol.8, pp.391-404, 1995.

[4] M. Nakagawa, "Chaos Associative Memory with a Periodic Association Function," J. Phys. Soc. Japan, vol.67, pp.2281-2293, 1998.

[5] G. E. Hinton,and T. J. Sejnowski, "Learning and relearning in Boltzmann machines", Parallel distributed processing, vol. 1, pp. 282-317, 1986.

[6] G. E. Hinton, "Deterministic Boltzmann Learning Performs Steepest Descent in Weight Space," Neural Computation., vol.1, pp.143-150, 1987.

[7] C. Peterson and J. R. Anderson, "A Mean Field Theory Learning Algorithm for Neural Networks," Complex Systems, vol.1, pp.995-1019, 1987.