

비단조 뉴런 모델을 이용한 결정론적 볼츠만머신

강형원, 박철영

대구대학교 정보통신공학부

전화 (053)850-6631 / FAX (053)850-6619

Deterministic Boltzmann Machine Based on Nonmonotonic Neuron Model

Hyoung-won Kang, Cheol-Young Park

Dept. of Computer & Communication Engineering, Daegu University

e-mail: bultae@iis.daegu.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.

서론

신경회로망의 연구에서는 단순한 입출력 함수를 갖는 뉴런모델을 이용하여 네트워크를 다양하게 구축하여, 패턴인식이나 연상기억 등의 문제에 대해 많은 연구가 수행되어 왔다[1]-[3]. 여기서 입출력 함수로서는 비선형 단조의 시그모이드(sigmoid)형 함수가 주로 이용되고 있다. 그러나 연상기억 문제와 같은 특정분야에 있어서 단조가 아닌 비단조 함수를 입출력 함수로 이용함으로써 기억용량이 증가되는 것이 모리타 등에 의해 보고되었다[4]. 여기서 이용한 함수는 입력값의 절댓값이 어느 한계를 넘으면 출력값이 반전하는 시그모이드 함수로 Morita형 함수로 부른다. 이 결과는 그 후 여러 연구자들에 의해 수학적으로 해명되고 증명되었으며 뉴런의 입출력 함수에 비단조 함수를 이용하면 보다 고기능의 신경회로망이 실현될 수 있음을 의미한다[5][6]. 여기서 연상기억문제에서 다루는 신경회로망

은 뉴런의 결합강도를 고정한 즉, 기억 매물형으로 학습에 의한 새로운 환경에 적응하는 것을 고려하지 않는다. 학습기능을 갖는 신경회로망에도 비단조 함수를 활성화 함수로 사용하는 뉴런을 이용해 기억 매물형의 신경회로망에 있어서 비단조 뉴런이 단조 뉴런 신경회로망에 대해서도 학습 성능의 향상되었다[7].

본 연구에서는 학습기능을 갖는 결정론적 볼츠만 머신에 비단조 뉴런을 이용하여 학습 성능을 수치 시뮬레이션을 통하여 분석한다. 네트워크의 은닉층에 비단조 활성화 함수로 end-cut-off 타입의 비단조 함수를 사용한 경우와 백프로퍼게이션 학습에 대하여 성능을 비교한다. 실제 하드웨어로 구성하는 경우의 실현을 용이하기 하기 위하여 end-cut-off 타입의 비단조 함수는 연속함수가 아닌 구분선형 함수를 사용한다. 네트워크의 학습성능을 학습수렴률, 학습횟수를 이용하여 비교한다. 학습 대상으로는 XOR문제에 대하여 수행한다. 4-4-1네트워크에서 N-Parity 문제에 대해서 활성화 함수로 비단조 함수와 단조 함수를 사용한 DBM을 비교하였다. 마지막으로 CMOS 회로를 사용하여 DBM 네트워크에 필요한 비단조 뉴런 회로를 설계하고 SPICE 시뮬레이션을 통하여 그 유용성을 입증한다.

본론

볼츠만머신은 학습기능을 갖는 대표적인 신경회로망으로 대칭결합을 갖는 확률론적인 회로망에 학습규칙을 도입한 것으로 Hinton과 Sejnowski에 의해 제안되었다[8]. 특히 결정론적 볼츠만머신[9]은 확률론적 볼츠만머신에 있어서 뉴런의 출력이 평균장 근사를 이용하여 뉴런의 동작을 결정론적으로 하는 네트워크로서 Peterson과 Anderson에 의하여 제안되었다[10]. 평균장 근사에 의해 뉴런의 출력 x_i 는

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

로 표현된다. 여기서 뉴런의 막전위 u_i 는 다음 식으로 주어진다.

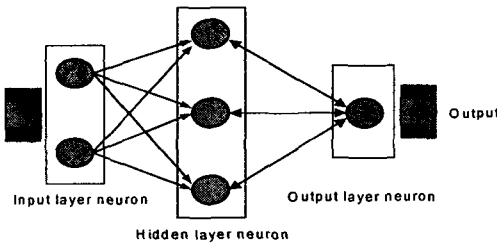
$$u_i = \tau \frac{du_i}{dt} + \sum_j w_{ij} x_j \quad (2)$$

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

$$\Delta W_{ij} = \frac{\epsilon}{T} \sum_q \{ (x_i^a x_j^q)_{clamped} - (x_i^a x_j^q)_{unclamped} \} \quad (3)$$

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

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



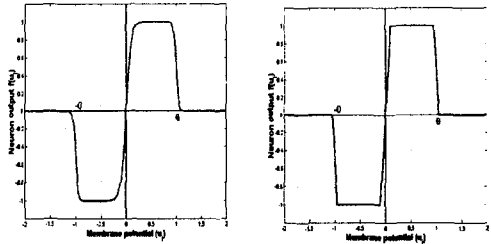
<그림 1> DBM 네트워크 모델

<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

시뮬레이션에서 학습 수렴 조건은 <표 1>에 나타난 것과 같이 원하는 출력값과 네트워크의 출력값의 오차가 1% 미만인 되는 것으로 하고 최대 학습횟수는 2,000회로 한다. 활성화 함수로는 일반적으로 <그림

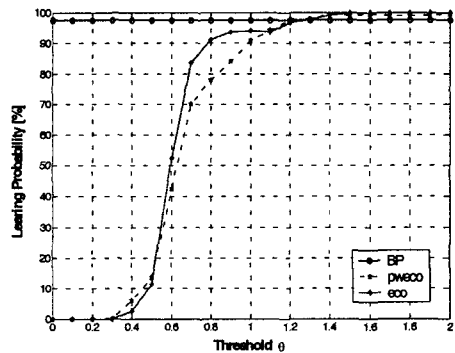
2>의 (a)는 end-cut-off 타입의 비단조함수를 나타내며 막전위의 절대값 $|u_i|$ 가 경계값 θ 를 넘으면 출력이 0이 되는 함수이다. 이 외에 막전위의 절대값 $|u_i|$ 가 경계값 θ 를 넘으면 출력값이 반전하는 Morita 타입 비단조 함수가 있다.



(a) Continuous end-cut-off type function($\theta=1.0$) (b) Piecewise linear end-cut-off type function($\theta=1.0$)

<그림 2> 2가지 타입의 활성화 함수

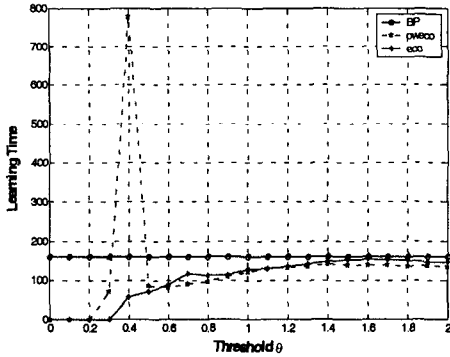
수치 시뮬레이션에서는 end-cut-off 함수를 이용하였다. 실제로 하드웨어로 구현하는 것을 고려할 때 <그림 2>의 (a)와 같은 연속함수를 회로로 구현하는 것은 어렵기 때문에 하드웨어화를 위해서는 각각 <그림 2>의 (b)처럼 구분선형 함수를 이용한다. 이때 함수가 x축과 교차하는 부분의 경사의 절대치는 전부 $1/T$ 로 한다. 시뮬레이션에서는 <그림 2>에 나타난 2가지 함수를 이용하여 백프로퍼게이션 학습을 이용한 네트워크와 비단조 뉴런을 이용한 결정론적 볼츠만머신의 성능을 비교하였다.



<그림 3> XOR문제에 대한 결과(학습 수렴율)

<그림 3>은 신경 회로망에서 비선형 문제 해결을 위해 가장 많이 사용되고 있는 XOR학습의 학습 수렴율을 나타낸 것이다. 학습 수렴율에 관해서 백프로퍼게이션은 98%의 높은 학습 수렴율을 나타내고 있고, 비단조 뉴런은 경계값 θ 를 적당한 값으로 설정함으로써 동등한 학습 수렴율을 얻어진다. 또한 θ 값이 1.4을 넘으면 비단조 뉴런의 경우가 백프로퍼게이션보다 높

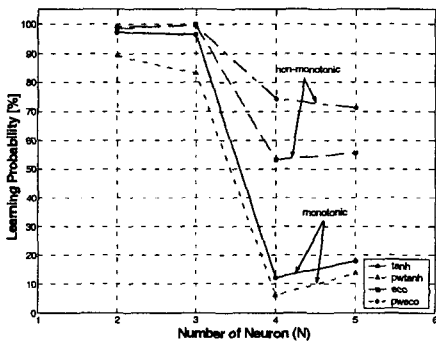
은 학습 수렴율을 나타내고 있다.



<그림 4> XOR문제에 대한 결과(학습횟수)

<그림 4>는 학습횟수를 나타낸 것으로 백프로퍼게이션은 경우 학습횟수가 평균 161회인 것에 반해 비단조 뉴런은 θ 값이 0에서 0.5까지 불안하긴 하지만 θ 값이 1.4일때는 비단조 뉴런의 학습은 142회이다. 이상의 결과로부터 XOR 문제에서는 경계값 θ 를 적당한 값으로 하면 비단조 뉴런을 이용한 네트워크가 백프로퍼게이션보다 비슷하거나 오히려 더 좋은 학습 성능을 나타내는 것을 알 수 있다.

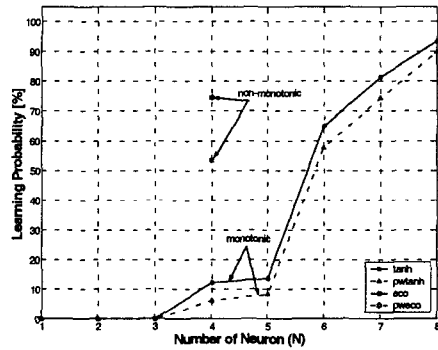
XOR는 입력이 두개인 특수한 경우의 패리티 문제라 할 수 있다. 따라서, 4-패리티 문제에서와 같은 비선형 수치 시뮬레이션에서 단조함수를 활성화 함수로 사용한 DBM과 비단조 함수를 활성화 함수로 사용한 DBM 사이의 학습 성능이 어떠한 차이를 두는지 비교하기 위하여 N-Parity 문제에 대한 시뮬레이션을 하였다



<그림 5> N-N-1 네트워크의 N-패리티 문제에 대한 시뮬레이션 결과

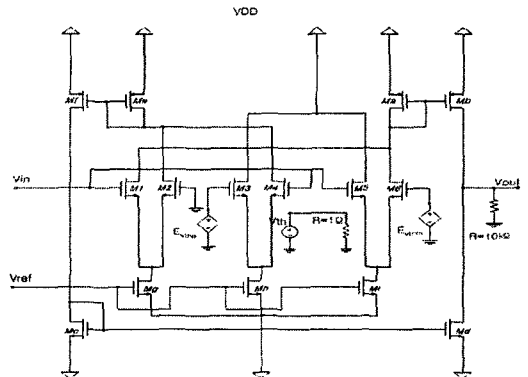
<그림 5>은 N-N-1 네트워크에서 N의 갯수 즉, 입력층과 은닉층을 같이 증가시킬 때의 학습 성능에 대한 평가를 해 보았다. 우선, N의 갯수가 2개와 3개일 때에는 학습 성능 면에서 큰 차이가 없었으나 N의 갯수

가 4개 이상으로 증가할 때 활성화 함수로 단조함수를 사용한 경우 낮은 학습 수렴율을 나타내는 것을 보였다. 이에 반해 비단조 함수를 사용한 경우 N의 갯수를 4개, 5개로 증가시켜도 학습 수렴율은 단조 함수에 비해 큰 차이가 없는 것으로 나타났다.



<그림 6> 4-X-1 단조 네트워크에서의 은닉층 노드수 변화에 대한 시뮬레이션 결과

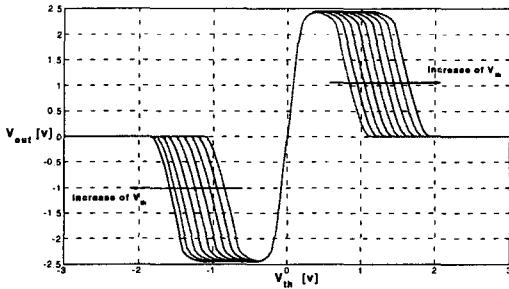
<그림 6>에서는 4-X-1의 단조 네트워크의 은닉층 수를 8개까지 증가시켜가면서 4-4-1 비단조 네트워크의 학습 수렴률과 동일해지는 은닉층 노드수를 찾아 보았다. 그림에서 보는 바와 같이 연속선형함수에서는 단조함수의 은닉층 노드수가 6개 일때 비슷한 수치를 나타내었으며, 하드웨어 구현을 위해 사용한 구분선형 단조함수의 경우는 은닉층의 노드수가 7개일때 비슷한 수치를 나타내었다. 그리고, 은닉층 노드수가 증가할수록 학습 수렴률이 높아지는 것을 알 수 있다.



<그림 7> End-Cut-Off 함수의 하드웨어 구현

<그림 7>는 본 연구에서 제시한 <그림 2> (a)의 end-cut-off형 함수를 전류미러(current mirror)를 활용하여 회로를 구성한 것이다. <그림 8>에는 그 시뮬레이션 결과를 나타내었다[11],[12]. <그림 7>에서 V_{in} 는 입력 전압의 값이며, V_{out} 은 end-cut-off형 함수의 출력을 나타낸다. 경계값 θ 에 해당하는 V_{th} 전압은

VCVS(Voltage Controlled Voltage Source)회로를 사용하여 조정하였다. <그림 7>의 회로도에는 3개의 시그모이드 함수를 만드는 전류미러 회로를 이용한 복합적인 합성으로 end-cut-off 함수를 구현해낼 수가 있으며, 수치 시뮬레이션시 변경하여 실행한 경계값 θ 는 SPICE 시뮬레이션의 파라미터 값 V_{th} 에 따라 자유로이 조정할 수 있다는 것을 알 수 있다.



<그림 8> SPICE 시뮬레이션 결과

<그림 8>에 나타난 시뮬레이션 결과처럼 CMOS 회로를 사용하여 본 논문에서 제시한 End-Cut-Off 회로를 구현할 수가 있으며, 시뮬레이션 파라미터 값과 V_{th} 값을 적당히 변화시킴으로써 DBM 네트워크에 필요한 비단조 뉴런에 관한 함수를 얻을 수 있다는 것을 알 수 있다.

결론

본 연구에서는 학습기능을 갖는 대표적 신경회로망인 결정론적 볼츠만머신의 은닉층에 비단조 뉴런을 적용한 네트워크를 구성을 제안하고, 학습 대상으로 XOR 문제에 관한 시뮬레이션을 수행하였다. 그 결과 백프로퍼게이션 학습과 비교하여 end-cut-off 형을 이용한 비단조 뉴런은 경계값 θ 를 적절하게 조정함으로써 학습 수렴율은 높고 학습횟수는 적은 것을 알 수 있었다. N-Parity 문제에 대한 학습 성능의 평가 결과로부터 비단조 함수를 활성화 함수로 사용한 DBM이 단조함수를 활성화 함수로 사용한 DBM보다 학습수렴율이 우수한 것을 알 수 있었다. 이러한 결과로 볼 때 비단조 함수의 사용이 더욱 효과적인 하드웨어 구현이 가능하리라 생각된다. 향후 이들 결과에 근거하여 실시간 처리를 요구하는 제어나 고도 정보처리에 응용하는 방법에 관한 연구가 필요하다.

참고문헌

[1] S. Nara, P. Davis, and H. Totsuji, "Memory search using complex dynamics in a recurrent neural network model", *Neural Networks*, vol.6, pp.963-973, 1993.
 [2] Y. Hayashi, "Oscillatory neural network and

learning of continuously transformed patterns", *Neural Networks*, vol.7, pp.219-231, 1994.

[3] C. Y. Park, Y. Hayakawa, K. Nakajima and Y. Sawada, "Limit cycles of one-dimensional neural networks with the cyclic connection matrix", *ICEICE Trans. on Fundamentals*, vol.E79-A, no.6, pp.752-757, 1996.

[4] M. Morita, "Associative Memory with Nonmonotone Dynamics", *Neural Networks*, no.6, pp.115-126, 1993.

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

[6] 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.

[7] C. Y. Park, D. H. Lee, "Learning Ability of Deterministic Boltzmann Machine with Non-Monotonic Neurons", *KFIS*, Vol.11, No.2 pp.275-278.

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

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

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

[11] Tetsuya Asai, H. Yokotsuka, T. Fukai "A Mos Circuit for a Nonmonotonic Neural Network with Excellent Retrieval capabilities", *IEEE Trans. on Neural Networks*, vol.7, no.1, 1996.

[12] C. A. Mead, 1987, *Analog VLSI and Neural systems*, Computer Science.