FPGA를 이용한 비단조 DBM 네트워크 설계

강형원, 박철영
대구대학교 전자정보공학부

Design of Non-monotonic Deterministic Bolzmann Machine Using FPGA

Hyung-won, Kang, Cheol-yung, Park
Dept. of Computer & Communication Engineering, Daegu University.

요 약

본 연구에서는 학습기능을 갖는 결정론적 볼츠만 머신에 비단조 뉴런을 이용하여 학습 성능을 수치 시뮬레이션을 통하여 분석한다. 먼저 네트워크의 은닉층에 비단조 및 단조뉴런을 이용한 경우에 대하여 각각 활성화 함수로 시그모이드 함수와 end-cut-off 타입의 비단조함수를 사용한 경우에 대하여 성능을 비교한다. 또한, VHDL을 이용해 설계한 DBM 네트워크에 시그모이드 함수와 end-cut-off 타입의 비단조함수를 사용한 경우에 대하여 시뮬레이션을 통해 수치 시뮬레이션과 성능이 같은지 비교하고 그 유용성을 입증한다.

1. 서 론

신경회로망의 연구에서는 단순한 입출력 함수를 갖는 뉴런모델을 이용하여 네트워크를 다양하게 구축하여, 패턴인식이나 연상기억 등의 문제에 대해 많은 연구가 수행되어 왔다[1]-[3]. 여기에서 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 절대값이 어느 한계를 넘으면 출력값이 반전하는 시그모이드 함수로 Morita 형 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]-[7]. 여기서 연상기억문제에서 다루는 신경회로망은 뉴런의 결합강도를 고정한 즉, 기억 매몰형으로 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 증가

이지만 학습 기능을 갖는 신경회로망에 대해서도 학습 성능의 향상이 기대된다.

본 연구에서는 학습기능을 갖는 결정론적 볼츠만 머신에 비단조 뉴런을 이용하여 학습 성능을 수치 시뮬레이션을 통하여 분석한다. 먼저 네트워크의 은닉층에 비단조 및 단조뉴런을 이용한 경우에 대하여 각각 활성화 함수로 시그모이드 함수와 end-cut-off 타입의 비단조함수를 사용한 경우에 대하여 성능을 비교한다. 네트워크의 학습 성능을 학습수렴율, 학습안정도, 그리고 학습 횟수의 3 가지 지표를 이용하여 비교한다. 또한, VHDL 을 이용해 설계한 DBM 네트워크에 시그모이드 함수와 end-cut-off 타입의 비단조함수를 사용한 경우에 대하여 시뮬레이션을 통해 학습수렴율, 학습안정도 및 학습 횟수를 이용하여 수치 시뮬레이션과 성능이 같은지 비교하고 그 유용성을 입증한다.

2. 본 론

볼츠만머신은 학습기능을 갖는 대표적인 신경회로망으로 대칭결합을 갖는 확률론적인 회로망에 학습 규칙을 도입한 것으로 Hinton 과 Sejnowski 에 의해 제안되었다[8]. 특히 결정론적 볼츠만머신[9]은 확률론적 볼츠만머신에 있어서 뉴런의 출력에 평균장 근사를 이용하여 뉴런의 동작을 결정론적으로 하는 네트워크로서 Peterson 과 Anderson 에 의하여 제안되었다[10]. 평균장 근사에 의해 뉴런의 출력 x_i 는

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

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

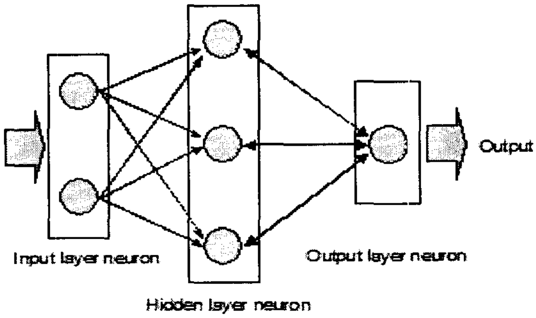
$$u_i = \tau\frac{du_i}{dt}+\sum_j w_{ij}x_j \qquad (2)$$

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

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

여기서 clamped 는 입출력 뉴런을 원하는 값으로 고정한 상태로 학습 phase 로 부르며 unclamped 는 입력 뉴런만을 고정한 상태로서 반학습 phase 라 한다. 식 (3)에 의해 하중값의 갱신량 ΔW_{ij} 는 각 상태에 있어서 학습 phase 와 반학습 phase 의 값 $x_i x_j$ 의 차를 모두 더한 값에 대해서 정수(ε/T) 배이다.

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



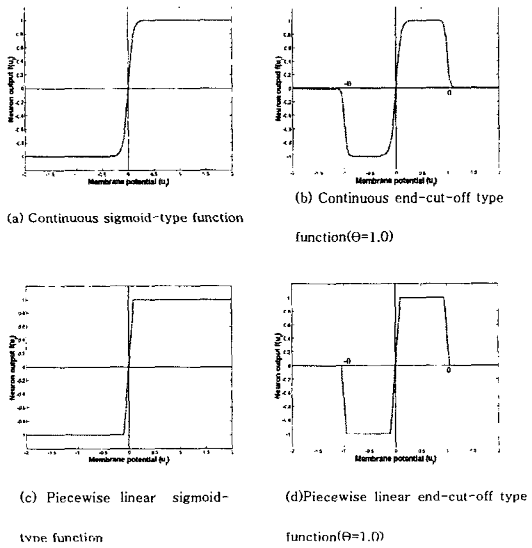
<그림 1> DBM 네트워크 모델

<표 1> 시뮬레이션 조건

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

시뮬레이션에서 학습 수렴 조건은 표 1에 나타낸 것과 같이 원하는 출력값과 네트워크의 출력값의 오차가 1% 미만이 되는 것으로 하고 최대 학습횟수는 2,000회로 한다. 활성화 함수로는 일반적으로 그림 2의 (a)에 나타낸 것과 같이 막전위의 값이 0을 넘

으면 발화하는 시그모이드 단조함수이다. 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 구분선형 함수를 이용한다. 이때 함수가 x축과 교차하는 부분의 경사의 절대치는 전부 $\frac{1}{T}$로 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

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 경계값를 적당한 값으로 함으로서 동등한 학습수렴율이 얻어진다. 학습안정도에 관해서는 이제까지의 결과와 달리 비단조 뉴런의 경계값를 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 θ 를 적당한 값으로 함으로써 단조뉴런의 경우와 비교하여 더 빠르게 학습이 수렴하고 있는 것을 알 수 있다. 예를 들면 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 경계값를 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

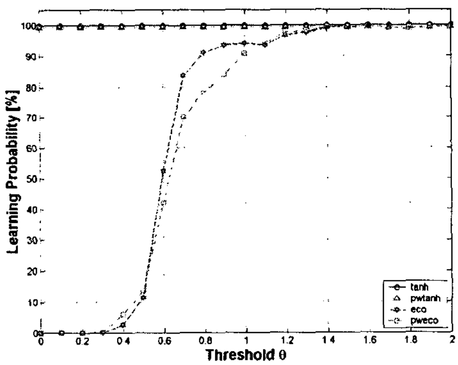


그림 3. XOR 학습에 대한 학습 수렴율(수치)

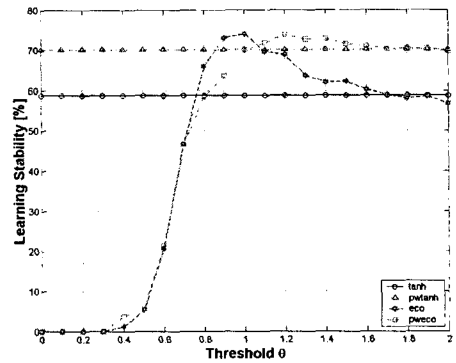


그림 4. XOR 학습에 대한 학습안정도(수치)

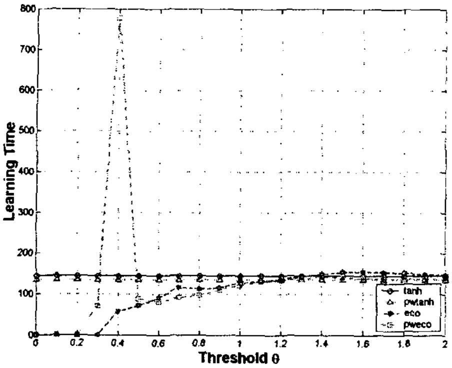


그림 5. XOR 학습에 대한 학습 횟수(수치)

위의 수치 시뮬레이션과 동일한 조건하에서 VHDL로 설계된 DBM네트워크의 시뮬레이션을 시행 하였다. 시뮬레이션 대상으로는 구분선형 시그모이드 함수와 구분선형 end-cut-off함수를 사용하여 학습수렴율과 학습 횟수에 대하여 실험 하였다. 그림 6은 학습 수렴율을 나타낸 것이다. 비단조 함수를 사용했을 때 수치 시뮬레이션과 같이 경계값 $\theta=1.3$ 부근에서 100%이다.

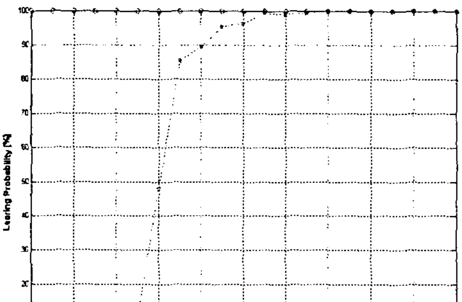


그림 6. XOR 학습에 대한 학습 수렴율(H/W)

그림 7은 학습 횟수를 나타내는 것이다. 단조함수를 사용 했을 때는 학습 횟수가 145회인 것에 반해 비단조 함수를 사용하면 경계값 $\theta\doteq1.0$ 부근에서 124회이다.

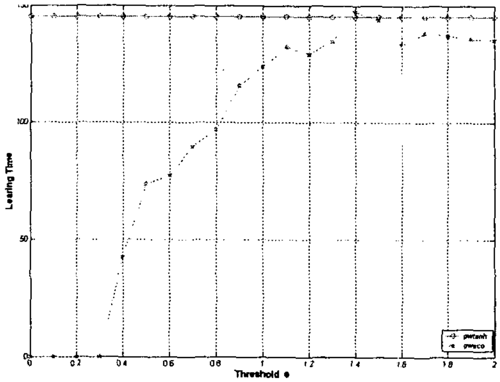


그림 7. XOR 학습에 대한 학습 횟수(H/W)

시뮬레이션 결과에서 보듯이 수치 시뮬레이션과 VHDL로 설계된 DBM네트워크의 시뮬레이션 결과가 약간의 차는 있지만 비슷하다는 것을 알 수 있다. 경계값 θ를 적당히 조절을 하면 비단조 함수를 사용한 경우가 단조함수를 사용한 경우보다 성능이 같거나 좋은 것을 알 수 있다.

3. 결 론

본 연구에서는 학습기능을 갖는 대표적 신경회로망인 결정론적 볼츠만머신의 은닉층에 비단조 뉴런을 적용한 네트워크를 구성을 제안하고, 학습 대상으로 XOR 문제에 관한 시뮬레이션을 수행하였다. 그 결과 단조 뉴런을 이용한 결정론적 볼츠만머신과 비교하여 end-cut-off 형을 이용한 비단조 뉴런은 경계값을 적절하게 조정함으로써 학습 수렴율이 향상되는 것을 확인하였다. 이것은 비단조 함수가 복수의 단조 뉴런의 합성으로 구성되며 이에 기인하여 좋은 성능이 나타난다고 생각된다. 또한 제시한 DBM 모델에 대한 하드웨어를 VHDL 로 구현하고 이를 시뮬레이션을 수행하였을 때도 수치 시뮬레이션과 비슷한 결과가 나오는 것을 알 수 있다. 이 결과로부터 비단조 함수의 사용이 동일한 완성도를 유지하는 더욱 효과적인 하드웨어 구현이 가능하리라 생각된다. 영상신호나 음성데이터 처리와 같은 실시간 처리에 활용하기 위해서는 실시간 동작이 가능한 하드웨어의 구현이 필수

적이다. 이를 이용한 실시간 연산이 가능하면 여러 분야에서 응용 가능성이 커질 것으로 기대된다.

4. 참 고 문 헌

[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] Simon Haykin, 1999, *Neural Networks,* 2nd, Prentice Hall.

[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.