양자화결합을 갖는 결정론적 볼츠만 머신 학습 알고리듬

Learning Algorithm for Deterministic Boltzmann Machine with Quantized Connections

박 철 영

대구대학교 정보통신공학부

Cheol-Young Park

Department of Computer & Communication Engineering, Taegu University

E-mail: cypark@biho.taegu.ac.kr

요약

본 논문에서는 기존의 결정론적 볼츠만 머신의 학습알고리듬을 수정하여 양자화결합을 갖는 볼츠만 머신에도 적용할 수 있는 알고리듬을 제안한다. 제안한 알고리듬은 2-입력 XOR문제와 3-입력 패리티문제에 적용하여 성능을 분석하였다. 그 결과 하중이 대폭적으로 양자화된 네트워크도 학습이 가능하다는 것과 은닉 뉴런수를 증가시키면 한정된 하중값의 범위로 유지할 수 있다는 것을 보여주었다. 또한 1회에 갱신하는 하중의 갯수 m_s를 제어함으로써 학습계수를 제어하는 효과가 얻어지는 것을 확인하였다.

Ⅰ. 서 론

신경회로망의 학습방법으로는 여러 가지 방법이 제안되어 있다. 대표적인 것으로는 Hebb rule[1], 백프로퍼게이션 학습[2] 및 볼츠만머신 학습[3] 등을 들 수 있다. 이들 학습방법의 구체적인 실현방법으로 현시점에서는 네트워크 구성상의 용이함 때문에 범용계산기를 이용하여 적절한 하중값을 계산하는 방법이 일반적이다. 그러나 네트워크의 규모가 증가하면 이 방법은 적합하지 않다.

또한 학습방법은 일반적으로 기본적인 아날로그 하중값을 갖는 신경회로망을 전제로 하지만 실현 할 수 있는 회로망은 유한한 분해능을 갖는다. 따라서 반도체회로 또는 광을 이용한 병렬계산기에 의한 신경회로망의 실현은 그 실현 용이성 때문에 하중값을 양자화한 모델이 제안되어 있으며 학습(하중값 결정법)에 대한 연구도 이루어지고 있다[4]-[6]. 그러나 이것들은 단순히 아날로그 하중값을 양자화하는 것으로, 일반적인 학습법이 적용되지 않으며 네트워크가 갖고 있는 능력의 저하도 피할 수 없다.

본 연구에서는 양자화결합 신경회로망에 적용 가능하면서 하드웨어화가 용이한 학습알고리듬을 확립한다. 구체적으로는 양자화결합을 갖는 결정론적 볼츠만 머신(deterministic Boltzmann machine)에 적용이 가능한 학습방법을 제안하고 그 성능을 평가한다. 여기서 하중값으로는 {±1, 0}의 세 가지 값을 이용하여 종래의 결정론적 볼츠만 머신의 학습 알고리듬을 양자화 결합 신경회로망에 적용 가능한 형태로 개량하여 하드웨어화가 용이한 적응적 학습알고리듬을 실현한다.

Ⅱ. 볼츠만머신 학습

환경 S가 특정 입출력 패턴을 어떤 확률로 생성한다고 가정하고 볼츠만머신의 입력과 출력 뉴런의 상태를 각각 환경 S의 입출력 패턴으로 고정시킨다. 여기서 은닉뉴런은 자유롭게 동작시키면 뉴런 i와 뉴런 j가 평형상태에서 동시에

1을 출력하는 확률은

$$p_{ij}=\sum_{\alpha,\beta}p(\alpha,\beta)E[x_ix_j|\alpha,\beta] \qquad (1)$$

로 표현된다. 여기서 $p(\alpha,\beta)$는 환경 S가 입출력 패턴 (α,β)를 생성하는 확률이고 $E[x_ix_j|\alpha,\beta]$는 볼츠만머신의 입력과 출력뉴런의 상태가 각각 α와 β일 때 x_ix_j의 기대값이다. 이때 하중값은

$$\Delta w_{ij}=\varepsilon p_{ij} \qquad (2)$$

이며 ε은 학습계수이다. 이와 같은 동작을 CLAMP phase라 부른다. 다음으로 입력 뉴런만을 고정하고 출력뉴런과 은닉뉴런은 모두 자유롭게 동작시킬 경우 뉴런 i와 j가 평형상태에서 동시에 1을 출력하는 확률은

$$p'_{ij}=\sum_{\alpha}p(\alpha)E[x_ix_j|\alpha] \qquad (3)$$

로 나타난다. 이때 하중값을 식(4)와 같이 변화시키는 동작을 UNCLAMP phase라 부른다.

$$\Delta w_{ij}=-\varepsilon p'_{ij} \qquad (4)$$

이상과 같이 CLAMP phase와 UNCLAMP phase를 반복함으로써 볼츠만머신은 어떤 입력이 주어졌을 때 환경 S와 같은 확률분포로 그 출력패턴을 생성하게 된다. 이것이 볼츠만머신 학습이다. 여기서 CLAMP phase에 있어서 입력뉴런이 상태 α을 가질 때 출력뉴런이 상태 β를 가질 조건확률을 $p(\beta|\alpha)$ 그리고 UNCLAMP phase에 있어서는 $p'(\beta|\alpha)$로 두면 학습은 식(5)의 함수 $G(w)$의 최급강하법과 등가가 된다

$$G(w)=\sum_{\alpha,\beta}p(\alpha,\beta)\log\frac{p(\beta|\alpha)}{p'(\beta|\alpha)} \qquad (5)$$

$$\frac{G}{w_{ij}}=-\frac{1}{T}(p_{ij}-p'_{ij}) \qquad (6)$$

여기서 $G(w)$는 Kullback-divergence로 $p_{ij}=p'_{ij}$일 때만 $G(w)=0$이고 그 외에는 $G(w)>0$이다. 또한 이상의 확률론적인 값을 평균장 정리를 이용해서 아날로그 값으로 취급하는 것을 결정론적 볼츠만머신(deterministic Boltzmann machine)(이하 DBM이라 약칭함)이라 부른다. DBM은 계산의 용이함 등으로 많이 사용되며 본 논문에서도 이를 사용한다.

Ⅲ. 양자화결합 신경회로망에 적용 가능한 학습방법

양자화결합 신경회로망의 학습은 일반적으로 어려우며 양자화된 정도가 클수록 현저하다. 일반적으로 최급강하법에 근거한 알고리듬이 성공하는 조건의 하나는 학습계수가 충분히 작아야 한다는 것이다. 학습계수가 작을수록 평가함수의 연속적인 기울기를 정확하게 따라가면서 하중값은 갱신되고 하중공간에 있어서 최소점(혹은 극소점)에 수렴하지만, 학습계수가 크고 1회의 갱신량이 큰 경우는 극소점으로의 수렴이 곤란하게 된다. 양자화 결합신경회로망의 학습이 곤란한 이유도 이와 같이 생각할 수 있다.

이상의 검토 결과 하중공간에 있어서 1회의 변화량이 작도록 억제함으로써 양자화 결합신경회로망의 학습성능이 향상되는 것으로 생각할 수 있어서 1회의 갱신에서 변화시키는 하중값의 갯수를 제어하는 학습을 제안한다. 이것은 최급강하법에 근거한 알고리듬에 일반적으로 적용할 수 있는 가능성이 있지만 하드웨어로 실현할 때의 장점 때문에 DBM에 적용하였다. 제안하는 알고리듬은 아래와 같다.

$$\Delta w'_{ij}=(V_iV_j)_c-(V_iV_j)_{uc} \qquad (7)$$

$$\Delta w''_{ij}=\begin{cases}+1 & (w'_{ij}>\varepsilon_p)\\ 0 & (|w'_{ij}|\le\varepsilon_p)\\ -1 & (w'_{ij}<-\varepsilon_p)\end{cases} \qquad (8)$$

$$\Delta w_{ij}=\begin{cases}\Delta w''_{ij} & (a_s=1)\\ 0 & (a_s=0)\end{cases} \qquad (9)$$

여기서 V_i와 V_j는 DBM에 있어서 각각 뉴런 i와 j의 출력으로 $[-1,1]$의 범위를 기지며 첨자 c와 uc는 각각 CLAMP와 UNCLAMP phase를 나타낸다. 식(9)의 a_s는 확률 p_s로 $a_s=1$을 가지며 확률 $1-p_s$로 $a_s=0$를 갖는다. 또한 식(9)는 전체 시냅스의 개수가 n_s일 때 「1회의 갱신에서 임의로 선택한 $m_s(=p_sn_s)$개의 하중만 갱신한다」와 같은 의미이다. 볼츠만머신을 채용한 이유는 학습의 국소성 때문이며 이것은 대규모 하드웨어를 실현하는 경우에 유리하다.

Ⅳ. 시뮬레이션

제안한 학습알고리듬을 검증하기 위하여 컴퓨터시뮬레이션을 수행하였다. 네트워크는 바이어스뉴런을 포함하는 3층의 DBM 네트워크를 이용하였다. 하중은 $\{+1, 0, -1\}$의 3가지 값으로 된 양자화결합으로 식(9)의 부분은 확률 p_s가 아니라, 갱신되는 하중의 갯수 m_s로 제어된다. 또한 뉴런의 모델로서 운동방정식은 연속시간의 미분방정식을 이용하고 그 활성화 함수로는

hyperbolic-tangent를 이용하였다.

학습방식으로서 on-line 방식을 사용한다. on-line 방식의 양자화결합 네트워크의 경우 CLAMP와 UNCLAMP의 두 가지 phase를 한 조로 하여 1회의 하중갱신을 함으로써 학습수렴을 촉진한다. 학습수렴의 판단으로는 자승평균오차(root mean square error)가 1%이하인 것을 정답으로 하며, 정답률 99% 이상인 상태가 학습이 계속되어도 일정시간(10스텝)유지되는 경우에 학습이 성공하여 수렴하였다고 판단한다.

먼저 2입력 XOR 문제에 제안된 알고리듬을 적용하였다. 초기상태로 뉴런의 막전위에 해당하는 변수(식(11)의 u_i)를 $[-T,\ T]$ 범위의 균일 난수로 설정함으로써 다양한 초기상태에 대해 살펴본다. 입력, 출력 그리고 바이어스 뉴런의 수를 각각 2, 1, 1로 하고 갱신하는 하중의 갯수 $m_s=1$일 때 은닉뉴런의 수에 대한 학습수렴률을 그림 2에 나타낸다. 일반 DBM네트워크의 경우 은닉뉴런의 수가 2개일 때 거의 100%에 가까운 학습수렴율을 나타내고 있다. 같은 조건으로 양자화결합 네트워크에서는 은닉뉴런의 수가 4개 이상에서 동등한 수렴율이 달성되는 것을 알 수 있다. 이 결과는 하중이 대폭적으로 양자화된 네트워크도 학습이 가능하다는 것과 또한 은닉 뉴런수를 증가시키면 한정된 하중값의 범위로 유지할 수 있는 것을 나타내고 있다.

볼츠만머신에 simulated annealing의 기법을 도입한 경우에 대해서 다음에 설명한다. 온도 T는 단순히 변수 u_i의 초기 설정시 균일난수의 범위를 $[-\frac{1}{10}T,\ +\frac{1}{10}T]$로 충분히 작게 하여 2단계의 온도의 simulated annealing을 행한다. 그 결과 양자화결합 네트워크, 일반 DBM의 모두 그림 3에 나타낸 것과 같이 학습수렴율이 개선되는 것을 알 수 있다.

다음은 3 입력 패리티문제에 대하여 알고리듬을 적용하였다. simulated annealing을 적용하지 않은 경우의 학습수렴률을 그림 4에 그리고 적용한 경우의 수렴률을 그림 5에 각각 나타내었다. 그래프에서 2입력 XOR 문제의 경우와 유사하다고 할 수 있다. 마지막으로 simulated annealing이 있는 조건으로 입력층-은닉층-출력층의 뉴런수를 각각 3-4-1로 하고 바이어스뉴런의 수를 1로 한 네트워크에 3입력 패리티문제를 학습시키는 경우에 학습수렴까지 걸리는 시간을 검토하였다. 그림 6은 100%의 수렴률을 달성하는 조건으로 1회에 갱신하는 하중의 갯수(m_s)에 대한 학습수렴까지 걸리는 평균 스텝수를 나타낸 것이다. 그림 7에 m_s의 최적값(=6)과 $m_s=1$인 경우의 양자화결합네트워크 및 일반 DBM학습에 대한 스텝수의 분포를 각각 나타낸다. 평균값으로 비교하면 $m_s=1$의 경우는 일반 DBM의 약 4.6배 그리고 $m_s=6$인 경우는 일반 DBM의 약 1.3배의 스텝수에서 학습이 수렴되고 있다. 이것은 일반 최급강하법을 이용한 학습에서 학습계수를 변화시킨 경우의 동작과 정성적으로 일치한다. 이는 파라미터 m_s(혹은 p_s)를 제어함으로써 학습계수를 제어하는 효과가 얻어지는 것을 나타낸다.

Ⅴ. 결론

본 논문에서는 기존의 결정론적 볼츠만 머신의 학습알고리듬을 수정하여 양자화결합을 갖는 볼츠만 머신에도 적용할 수 있는 알고리듬을 제안하였다. 제안한 알고리듬을 2-입력 XOR문제와 3-입력 패리티문제에 적용하여 성능을 분석하였다. 그 결과 하중이 대폭적으로 양자화된 네트워크도 학습이 가능하다는 것과 은닉 뉴런수를 증가시키면 한정된 하중값의 범위로 유지할 수 있는 것을 보여주었다. 또한, 1회에 갱신하는 하중의 갯수 m_s를 제어함으로써 학습계수를 제어하는 효과가 얻어지는 것을 확인하였다.

Ⅵ. 참고문헌

[1] Hebb, "The Oroanization of behavior", Wiley, 1949.

[2] D. E. Rumeljart, G. E. Hinton, and R. J. Williams, "Learning representations by back propagating errors", Nature, vol. 323, pp. 533-536, 1986.

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

[4] V. Michel, S. Bruno, M. V. Andre, and G. A. J. Paul, "Neural Networks for

High-Storage Capacity Content-Addressable Memory: VLSI Circuit and Learning Algorithm", IEEE Journal of Solid-State Circuits, vol. 24, no. 3, pp. 562-569, 1989.
[5] P. Chung, C. Tsai, and Y. Sun, "Characteristics of Hebbian-Type Associate Memories with Quantized Interconnections", IEEE Trans. Circuits Sys., vol. 41, no. 2, pp. 168-171, 1994.
[6] H. Mizutani, "A weighting method for the Hopfield neural networks which have discrete weights", IEICE Trans. vol. J78-A, no. 9, pp.1111-1118, 1995.

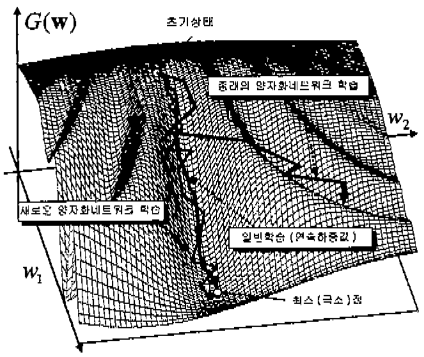

그림 1. 하중공간에서의 최급강하법의 개념도

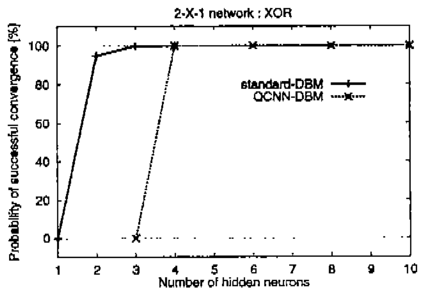

그림 2. 2-입력 XOR 문제의 학습수렴율 (annealing 미적용)

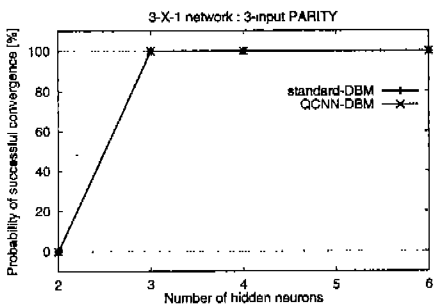

그림 3. 2-입력 XOR 문제의 학습수렴율 (annealing 적용)

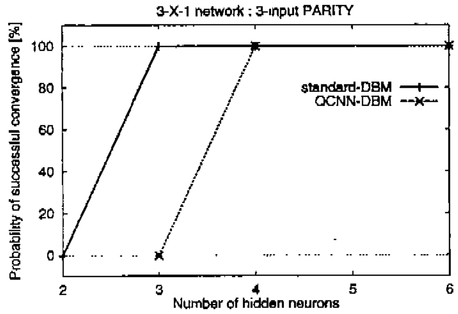

그림 4. 3-입력 Parity 문제의 학습수렴율 (annealing 적용)

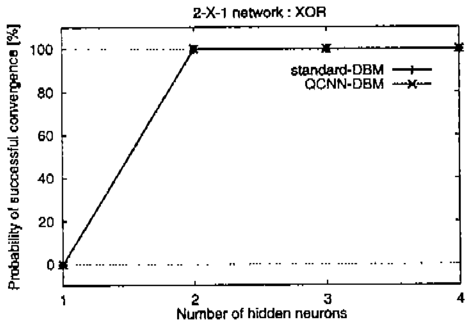

그림 5. 3-입력 Parity 문제의 학습수렴율 (annealing 미적용)

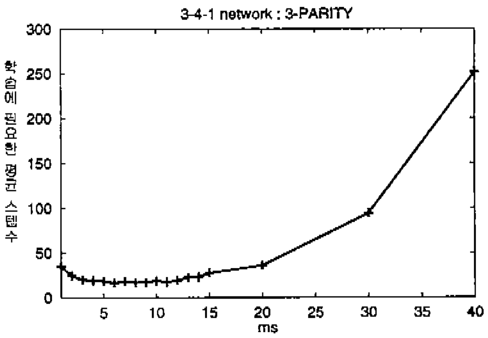

그림 6. m_s에 대한 학습수렴에 걸리는 평균스텝수

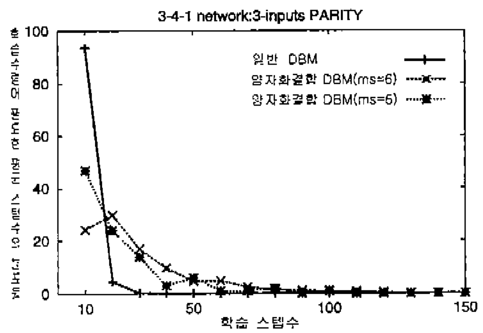

그림 7. 학습 스텝수의 분포