# 양자화 결합 네트워크를 위한 수정된 결정론적 볼츠만머신 학습 알고리즘†
# (A Modified Deterministic Boltzmann Machine Learning Algorithm for Networks with Quantized Connection)

박 철 영*
(Cheol-Young Park)

**요 약** 본 논문에서는 기존의 결정론적 볼츠만 머신의 학습알고리즘을 수정하여 양자화결합을 갖는 결정론적 볼츠만 머신에도 적용할 수 있는 알고리즘을 제안하였다. 제안한 알고리즘을 2-입력 XOR 문제와 3-입력 패리티 문제에 적용하여 성능을 분석하였다. 그 결과 하중이 대폭적으로 양자화된 네트워크에 대해서도 학습이 가능하다는 것과 은닉층 뉴런의 수를 증가시키면 한정된 하중값의 범위로 유지할 수 있는 것을 보여준다. 또한 1회에 갱신하는 하중의 갯수를 제어함으로써 학습계수를 제어하는 효과가 얻어지는 것을 확인하였다.

***Abstract*** From the view point of VLSI implementation, a new learning algorithm suited for network with quantized connection weights is desired. This paper presents a new learning algorithm for the DBM(deterministic Boltzmann machine) network with quantized connection weight. The performance of proposed algorithm is tested with the 2-input XOR problem and the 3-input parity problem through computer simulations. The simulation results show that our algorithm is efficient for quantized connection neural networks.

## 1. 서 론

뉴로 컴퓨팅은 독자적인 아키텍처를 유효하게 활용하기 위하여 전용하드웨어가 필수적으로 요구된다. 기존의 노이만형 계산기가 계산량의 폭주를 발생하는 것과 같은 복잡한 문제에 대해서 신경회로망의 우위성이 보다 명확하게 나타나기 위하여 보다 많은 수의 소자를 집적화하는 것이 요구되며 특히 VLSI에 의한 고집적화가 기대되고 있다. 실용적인 응용에 필요한 회로망의 규모(뉴런의 수)는 명확하게는 알려져 있지 않으나, 예를 들면 인간의 뇌는 약 $10^{10}$개의 신경세포(뉴런)가 존재한다고 한다. 반도체기술의 진보에 의해 신경회로망의 집적화는 많은 성과가 얻어졌으나 아직 충분하다고는 말할 수 없다. 일반적으로 시냅스의 수는 뉴런 수의 자승에 비례하기 때문에 시냅스가 차지하는 면적이 지배적이며 이것이 고집적화를 곤란하게 하는 최대의 요인이다. 따라서 시냅스를 작게 실현하는 것이 고집적화에서 가장 중요하다. 시냅스가 담당하는 기능은 결합하중을 위한 승산과 하중치를 기억하는 메모리이다. 이들 기능을 소규모로 실현하기 위하여 아날로그 승산기나 아날로그 메모리 등이 이용되는 경우가 많다. 또한 하중치를 양자화 하여 정밀도(비트의 수)를 감소시켜 회로를 간략하게 하는 것도 대단히 유효한 방법이다. 그러나 신경회로망은 원래 무한의 분해능을 갖는 연속치를 취급하는 것으로 모델링되어 있어서 하드웨어화 에서 유한한 분해능이 문제가 된다. 실용적인 학습에는 하중치에 10비트 이상의 분해능이 필요하다고 알려져 있으며[1],[2] 하중의 양자화나 아날로그회로에 의한 접근에서도 이 분해능을 만족하기 위해서는 시냅스 회로는 많은 면적을 필요로 한다. 이것이 신경회로망의 고집적화를 방해

* 대구대학교 정보통신공학부
† 이 논문은 2001년도 대구대학교 학술 연구비 지원에 의한 논문임

하는 하나의 원인이다.

신경회로망의 학습방법으로서는 여러 가지 방법이 제안되어있다. 대표적인 것으로는 Hebb 학습, 백프로퍼게이션, 그리고 볼츠만 머신의 학습알고리듬 등이 있다. Hebb 학습[3]은 신경세포의 발화에 근거한 시냅스 가소성의 가설로부터 얻어진 것으로 「동시에 발화하는 신경세포사이의 결합하중은 증대된다」라는 원리에 근거한다. 백프로퍼게이션 학습방법[4]과, 볼츠만머신의 학습알고리즘[5]은 각각 오차함수와 Kullback-divergence의 최급강하법에 근거한다. 구체적인 학습의 실현방법으로 현시점에서는 이들 학습방법의 구체적인 실현방법으로 네트워크 구성상의 용이함 때문에 범용계산기를 이용하여 적절한 하중치를 계산하는 방법이 일반적이다. 그러나 신경회로망의 규모가 증대함에 따라 학습에 필요한 계산량도 증대하기 때문에 범용계산기에 의한 학습은 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 모델[6],[7] 제안되어 있으며 학습(하중값 결정법)에 대한 연구도 이루어지고 있다[8]-[10]. 그러나 이것들은 단순히 아날로그 하중값을 양자화하는 것으로, 일반적인 학습방법이 적용되지 않으며 네트워크가 갖고 있는 능력의 저하도 피할 수 없다. 따라서 네트워크뿐만 아니라 학습기능도 같은 하드웨어상에서 수행할 필요가 있다. 동일 집적회로 상에 네트워크와 학습기능을 실현하는 방식은 on-chip 학습이라 부른다.

실용적인 신경회로망의 응용에는 대규모 네트워크가 수반되어야하므로 전용하드웨어에 의한 고집적화와 학습기능의 하드웨어화(on-chip 학습)이 필수요건이다. 그러나 이 두 가지 요건은 상반된 요구를 갖기 때문에 대규모 시스템의 실현에 지장을 초래한다.

본 연구에서는 기존의 결정론적 볼츠만 머신의 학습알고리즘을 수정하여 양자화결합을 갖는 결정론적 볼츠만 머신에도 적용할 수 있는 알고리즘을 제안한다. 그리고 제안하는 알고리즘을 XOR 문제와 패리티문제에 적용하여 그 성능을 평가한다. 여기서 하중값으로는 {+1, 0, −1}의 세 가지 값으로 양자화된 것을 사용한다.

## 2. 결정론적 볼츠만머신 학습

환경 $S$가 특정 입출력 패턴을 어떤 확률로 생성한다고 가정하고 볼츠만머신의 입력과 출력뉴런의 상태를 각각 환경 $S$의 입출력 패턴으로 고정시킨다. 여기서 은닉뉴런은 자유롭게 동작시키면 뉴런 $i$와 뉴런 $j$가 평형상태에서 동시에 1을 출력하는 확률은

$$p_{ij} = \sum_{\alpha,\beta} p(\alpha,\beta) E[x_i x_j \mid \alpha,\beta] \tag{1}$$

로 표현된다. 여기서 $p(\alpha,\beta)$는 환경 $S$가 입출력 패턴 $(\alpha,\beta)$를 생성하는 확률이고 $E[x_i x_j \mid \alpha,\beta]$는 볼츠만머신의 입력과 출력뉴런의 상태가 각각 $\alpha$와 $\beta$일 때 $x_i x_j$의 기대값이다. 이때 하중값은

$$\Delta w_{ij} = \varepsilon p_{ij} \tag{2}$$

이며 $\varepsilon$은 학습계수이다. 이와 같은 동작을 CLAMP phase라 부른다. 다음으로 입력 뉴런만을 고정하고 출력뉴런과 은닉뉴런은 모두 자유롭게 동작시킬 경우 뉴런 $i$와 $j$ 가 평형상태에서 동시에 1을 출력하는 확률은

$$p'_{ij} = \sum_{\alpha} p(\alpha) E[x_i x_j \mid \alpha] \tag{3}$$

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

$$\Delta w_{ij} = -\varepsilon p'_{ij} \tag{4}$$

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

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

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

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 정리[11]를 이용해서 아날로그

값으로 취급하는 것을 결정론적 볼츠만머신(deterministic Boltzmann machine: 이하 DBM이라 약칭함)이라 부른다[12]. 학습기능을 갖는 신경회로망의 집적화에 있어서는 결정론적 볼츠만 머신 학습이 백프로퍼게이션 학습에 비해 회로 구성이 간단하고 학습시간이 빠르므로 고집적 대규모 신경회로망의 구현에 유리하다.

## 3. 양자화결합 네트워크의 학습방법

양자화된 하중치를 갖는 신경회로망의 학습은 일반적으로 어려우며 양자화된 정도가 클수록 현저하다. 일반적으로 최급강하법에 근거한 알고리즘이 성공하는 조건의 하나는 학습계수가 충분히 작아야 한다는 것이다. 그림 1에 나타낸 것과 같이 학습계수가 작을수록 평가함수의 연속적인 기울기를 정확하게 따라가면서 하중값은 갱신되고 하중공간에 있어서 최소점(혹은 극소점)에 수렴한다. 그러나 학습계수가 크고 1회의 갱신량이 큰 경우는 극소점으로의 수렴이 곤란하게 된다. 양자화 결합 신경회로망의 학습이 곤란한 이유도 이와 같이 생각할 수 있다.

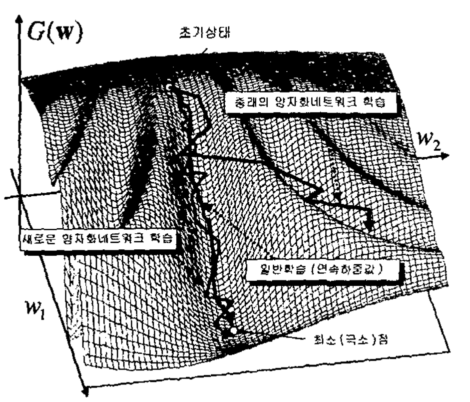


그림 1. 하중공간에서의 최급강하법의 개념도

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 알고리즘은 아래와 같다.

$$\Delta w'_{ij} = (V_i V_j)_c - (V_i V_j)_{uc} \tag{7}$$

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

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

여기서 $V_i$와 $V_j$는 DBM에 있어서 각각 뉴런 $i$ 와 $j$ 의 출력으로 $[-1, 1]$의 범위를 가지며 첨자 c와 uc는 각각 CLAMP와 UNCLAMP phase를 나타낸다. 식 (9)의 $a_s$는 확률 $p_s$로 $a_s=1$을 가지며 확률 $1-p_s$로 $a_s=0$를 갖는다. 또한 식 (9)는 전체 시냅스의 개수가 $n_s$ 일 때 「1회의 갱신에서 임의로 선택한 $m_s(=p_s n_s)$개의 하중만 갱신한다」 라는 의미이다. 볼츠만머신을 채용한 이유는 학습의 국소성 때문이며 이것은 대규모 하드웨어를 실현하는 경우에 유리하다.

## 4. 시뮬레이션

제안한 학습 알고리즘을 검증하기 위하여 컴퓨터를 이용한 수치 시뮬레이션을 수행하였다. 네트워크는 그림 2와 같이 바이어스 뉴런을 포함하는 3층의 DBM 네트워크를 이용하였다.

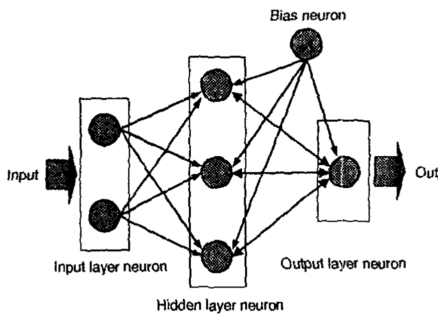


그림 2. 3층의 결정론적 볼츠만머신의
네트워크 모델의 예 (2-3-1 네트워크)

하중은 $\{+1, 0, -1\}$의 세 가지 값을 갖는 양자화결합으로 식 (9)의 부분은 확률 $p_s$가 아니라, 갱신되는 하중의

갯수 $m_s$로 제어된다. 또한 뉴런의 모델로서 운동방정식은 식 (10)과 같은 연속시간의 미분방정식을 이용하고 그 활성화 함수로는 식 (11)과 같은 hyperbolic-tangent를 이용하였다.

$$\tau\frac{du_i}{dt} = -u_i(t) + \sum_{j=1}^{n} w_{ij}V_j \quad (10)$$

$$V_i = f(u_i) = \tanh(u_i/T) \quad (11)$$

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(10 스텝)유지되는 경우에 학습이 성공하여 수렴하였다고 판단한다.

먼저 대표적인 비선형 문제인 XOR 문제에 제안된 알고리즘을 적용하였다. 초기상태로는 뉴런의 막전위에 해당하는 식 (11)의 $u_i$ 를 [−T, T] 범위의 균일 난수로 설정함으로써 다양한 초기상태에 대해 네트워크의 성능을 살펴본다. 입력, 출력 그리고 바이어스 뉴런의 수를 각각 2, 1, 1로 하고 갱신하는 하중의 갯수 $m_s=1$일 때 은닉뉴런의 수에 대한 학습 수렴률을 그림 3에 나타낸다. 여기서는 simulated annealing을 적용하지 않았다.

일반 DBM 네트워크의 경우 은닉층 뉴런의 수가 2개일 때 거의 100%에 가까운 학습 수렴율을 나타내고 있다. 같은 조건으로 양자화결합 네트워크에서는 은닉층 뉴런의 수가 4개 이상에서 동등한 수렴율이 달성되는 것을 알 수 있다. 이 결과는 하중이 대폭적으로 양자화된 네트워크도 학습이 가능하다는 것을 나타낸다. 또한 은닉층의 뉴런 수를 증가시키면 한정된 하중값의 범위로 유지할 수 있는 것을 나타내고 있다.

다음으로 DBM에 simulated annealing의 기법을 도입한 경우의 알고리즘의 성능에 대하여 설명한다. 막전위 $u_i$의 초기값 설정시 균일 온도 $T$ 는 균일 난수의 범위를 $[-\frac{1}{10}T, +\frac{1}{10}T]$ 로 충분히 작게 하여 2 단계 온도의 simulated annealing을 행한다. 그 결과 양자화결합 네트워크와 일반 DBM의 모두 그림 4에 나타낸 것과 같이 학습 수렴율이 개선되는 것을 알 수 있다.

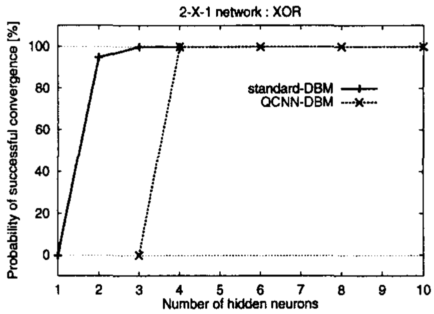


그림 3. 2-입력 XOR 문제에 대한 학습수렴율 (simulated annealing 미적용)

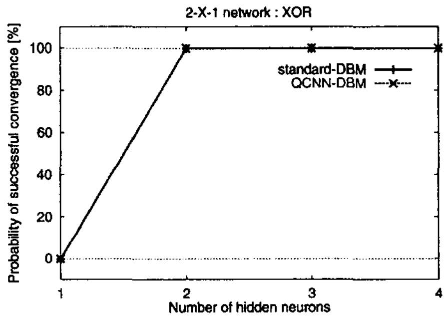


그림 4. 2-입력 XOR 문제에 대한 학습수렴율 (simulated annealing 적용)

다음은 3 입력 패리티문제(parity problem)에 대하여 제안한 알고리듬을 적용하였다. simulated annealing을 적용하지 않은 경우의 학습수렴률을 그림 5에 그리고 적용한 경우의 수렴률을 그림 6에 각각 나타내었다. 그림으로부터 2입력 XOR 문제의 경우와 유사한 결과가 나오는 것을 알 수 있다.

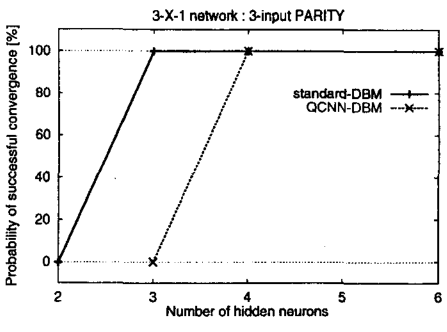


그림 5. 3-입력 패리티 문제에 대한 학습수렴율 (simulated annealing 미적용)

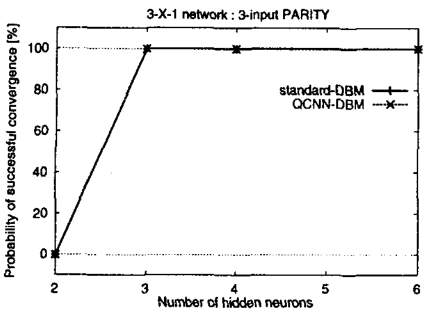


그림 6. 3-입력 패리티 문제에 대한 학습수렴율 (simulated annealing 적용)

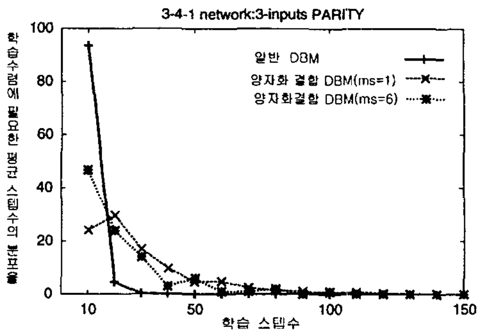


그림 8. 학습수렴에 소요되는 스텝수의 분포

마지막으로 simulated annealing이 있는 조건 아래에서 입력층-은닉층-출력층의 뉴런 수를 각각 3-4-1로 하고 바이어스 뉴런의 수를 1로 한 네트워크에 3입력 패리티문제를 학습시키는 경우에 학습수렴까지 걸리는 시간을 검토하였다. 그림 7은 100%의 수렴률을 달성하는 조건으로 1회에 갱신하는 하중의 갯수( $m_s$)에 대한 학습수렴까지 걸리는 평균 스텝수를 나타낸 것이다. $m_s$가 어떤 값 이하에서 학습이 수렴하는데 걸리는 시간이 최소가 되는 것을 알 수 있다. 그림 8에 $m_s$의 최적값 $m_s=6$과 $m_s=1$인 경우의 양자화결합 네트워크 및 일반 DBM 학습에 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 일치하며 파라미터 $m_s$(혹은 $p_s$)를 제어함으로써 학습계수를 제어하는 효과가 얻어지는 것을 나타낸다.

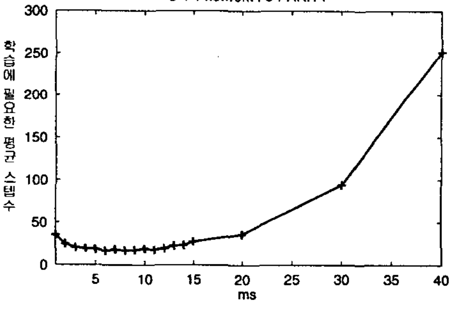


그림 7. 1회에 갱신하는 하중의 갯수 $m_s$의 변화에 따른 학습수렴에 걸리는 평균스텝 수

## 5. 결 론

본 논문에서는 기존의 결정론적 볼츠만 머신의 학습 알고리즘을 수정하여 양자화결합을 갖는 볼츠만 머신에도 적용할 수 있는 알고리즘을 제안하였다. 제안한 알고리즘을 2-입력 XOR 문제와 3-입력 패리티 문제에 적용하여 성능을 분석하였다. 그 결과 하중이 대폭적으로 양자화된 네트워크에 대해서도 학습이 가능하다는 것과 은닉 뉴런수를 증가시키면 한정된 하중값의 범위로 유지할 수 있는 것을 보여주었다. 또한 1회에 갱신하는 하중의 갯수를 제어함으로써 학습계수를 제어하는 효과가 얻어지는 것을 확인하였다. 본 논문의 연구 결과는 대규모 신경회로망 구성과 학습에 유효하며 신경회로망의 응용가능성을 높인다. 또한 신경회로망 연구자체를 용이하게 할 것으로 기대된다.

## 참 고 문 헌

[1] P. W. Hollis, J. S. Harper, and J. J. Paulos, "The Effect of Precision Constraints in a Backpropagation Learning Network," Neural Computation, vol. 2., pp. 363-373, 1990.

[2] T. Morie, O. Fujita, and Y. Amemiya, "Analog VLSI implementation of adaptive algorithms by an Extended Hebbian synapse circuit," ICEICE Trans. on Electron, vol. E75-C, no. 2, pp. 303-311, 1992.

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

[4] D. E. Rumeljart, G. E. Hinton, and R. J. Williams, "Learning representations by back

propagating errors," Nature, vol.323, pp.533-536, 1986.

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

[6] 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.

[7] C. Y. Park, K, Nakajima, "Analog CMOS Implementation of Quantized Interconnection Neural Networks for Memorizing Limit Cycles," IEICE Trans. Fundamentals, vol. E82-A, no.6, pp. 952-957, 1999.

[8] 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.

[9] 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.

[10] C. Y. Park, K, Nakajima, "Majority Algorithm: A Formation for Neural Networks with the Quantized Connection Weights," ICEICE Trans. on Fundamentals, vol .E82-A, no. 6, pp. 1059-1065, 2000.

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

[12] G. E. Hinton, "Deterministic Boltzmann Learning Performs Steepest Descent in Weight Space," Neural Computation., vol. 1, pp. 143-150, 1989.

**박 철 영** (Cheol-Young Park)

1984년 경북대학교 전자공학과 졸업(공학사)
1986년 경북대학교 전자공학과 졸업(공학석사)
1997년 일본 동북대학 정보과학 연구과 졸업(공학박사)
1985년～1990년 (주)금성사 중앙연구소 근무
1991년～1993년 (주)금성사 영상미디어연구소 근무
1993년～1997년 LG전자 멀티미디어연구소 근무 (책임연구원)
1997년～현재 대구대학교 정보통신공학부 조교수
<주관심분야> 신경회로망을 이용한 지능정보처리, 신경회로망 VLSI 설계 및 신개념 지능형 소자 개발