
신경회로망을 사용한 N 비트 패리티 판별

최재승*

*신라대학교

N bit Parity Discrimination using Perceptron Neural Network

Jae-seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

본 논문에서는 오차역전파 알고리즘을 사용한 3층 구조의 퍼셉트론형 신경회로망으로 네트워크의 학습을 실시하여, N비트의 패리티판별에 필요한 최소의 중간유닛수의 해석에 관한 연구이다. 따라서 본 논문은 제안한 퍼셉트론형 신경회로망의 중간 유닛의 수를 변화시켜 N 비트의 패리티 판별 실험을 실시하였다. 본 시스템은 패리티 판별의 실험을 통하여 N 비트 패리티 판별이 가능하다는 것을 실험으로 확인한다.

키워드

Parity discrimination, back propagation algorithm, hidden unit number, neural network

I. 서 론

최근의 신경회로망의 연구에 있어서 패리티 판별회로는 제안된 이론을 뒷받침하는 회로의 연구로서 자주 사용되어지고 있다[1, 2, 3]. 이것은 다른 어려운 문제가 하나의 부분 문제로서 이러한 것들을 많이 포함하고 있기 때문이다. 이 경우에, N비트 패리티 판별에는 중간층수는 활성 함수로서 비선형 함수를 사용한 경우, 최저 N비트가 필요하다라는 암묵의 법칙이 있다고 할 수 있다[4]. 이 원인은 Rumelhart 등에 의해서 쓰여진 저서 "PDP(Parallel Distributed Processing) 모델[5]"에 의한 한 가지 설이 믿어지기 때문이다.

본 논문에서는 실험에 의한 N비트 패리티 판별에는 중간 유닛의 수가 N개보다도 적은 네트워크를 사용함으로써 문제해결을 실시하고 있지만, 일반적으로는 중간유닛수가 최소 $(N+1)/2$ 개로 패리티 판별이 가능하다. 따라서 본 논문에서는 최

소 유닛수로서 패리티 판별 문제를 쉽게 해결하기 위하여 네트워크의 초기값의 부여방법을 제안하여, 최소 유닛수에 대한 패리티 판별 신경회로망의 각 파라미터의 최적해에의 조건을 사용하여 실험을 실시한다.

II. 오차역전파 알고리즘 및 패리티 판별

오차역전파 알고리즘은 1986년 Rumelhart 등이 논문을 발표한 것이 단서가 되어 다층 퍼셉트론의 가중치 학습 알고리즘이 광범위하게 연구자의 사이에서 논의되었다. 오차역전파 알고리즘은 일반성이 높은 뛰어난 알고리즘이며 많은 문제해결에 응용되고 있다. 또한, 오차역전파 알고리즘은 극치탐색수법이며, 평가함수가 극소치를 많이 포함하는 경우에 이것들에 떨어지기 때문에 대역적인 최소값에 수속하지 않을 가능성이 있다. 이 문

제를 피하는 결정적인 수법은 현재는 알려져 있지 않지만, 복수의 초기값으로부터 극소탐색을 시험하는 방법, 중간층의 유닛수를 증가시키는 방법, 확률적으로 가끔 에너지를 증대하는 방향으로 의도적으로 가중치 수정을 실시하는 등의 방법이 일반적으로 시험되고 있다.

패리티 판별은 입력 패턴의 패리티를 조사하는 문제이며, 입력 패턴이 기수개의 1을 포함한 경우에만 1을 출력하는 문제이다. 이 문제는 거의 비슷한 입력에 대해서 다른 출력을 나타내지 않으면 안되기 때문에, 네트워크에 있어서는 어려운 문제이다 [3, 4]. 또한 XOR 문제는 입력 유닛의 수가 2개의 패리티 문제이란 것을 말할 수 있다. Rumelhart [6]등의 실험에 의하면 입력 유닛으로부터 출력 유닛에의 직접적인 결합이 없는 계층형의 네트워크를 사용하는 경우, 입력의 크기 N의 패리티 문제를 해결하기 위해서는, 적어도 N개의 은닉 유닛이 필요하다고 알려져 있다.

III. 패리티 판별 실험 결과 및 고찰

본 장에서는 오차역전과 알고리즘을 사용하여 패리티 판별을 위한 신경회로망의 시뮬레이션의 결과 및 N비트의 패리티 판별에 필요한 최소 중간 유닛수의 도출, 초기값을 조작하는 것에 의한 학습상황이 어떻게 변화하는 것 등을 나타낸다.

3층 퍼셉트론형 신경회로망의 중간 유닛의 수를 변화시켜 2비트에서 5비트까지의 패리티 판별을 실시하였다. 학습에 사용한 파라미터는 η (학습계수): 0.2, α (관성계수): 0.9, 수속판정오차: 1.0×10^{-4} , 최대학습횟수: 30,000회이다. 또한 하나의 네트워크에 대하여 1,000회의 시행을 실시하여 이 결과를 표 1에 정리하였다. 또한 표 중에 있어서 "(분리가)"는 학습횟수가 30,000회에 도달한 것으로 평가함수의 값이 수속판정오차보다도 크지만, 출력이 0.5보다 작은 경우에는 0, 큰 경우에는 1로써 보면 목표 패턴과 일치하는 것도 수속횟수에 포함시킨다. 표 1로부터 중간 유닛을 증가시킬 수록 수속횟수는 많게 되며 평균학습횟수는 적게 되는 것을 알 수 있다. 또한 평균학습횟수에 대한 표준편차가 적게 되어가므로 보다 네트워크의 안정성이 증가되는 것을 알 수 있다. 결론적으로 3, 4, 5 비트의 입력에 대해서, 수속횟수

는 적지만 N비트의 입력에 대해서 N개 이하의 중간 유닛수로도 패리티를 판별하는 것이 가능하다는 것을 알 수 있다.

표 1. 중간유닛수와 수속횟수 및 평균학습횟수와 의 관계

네트워크	수속횟수 (분리가)	평균 학습횟수	학습횟수의 표준편차
3-2-1	818(853)	22422.6	3919.1
3-3-1	995(1,000)	2975.9	2021.9
3-4-1	1,000	2061.9	564.6
3-5-1	1,000	1777.8	420.4
4-2-1	48(141)	23439.0	3941.1
4-3-1	429(495)	6866.4	8011.2
4-4-1	763(799)	4774.0	6018.5
4-5-1	910(916)	3383.4	4079.9
5-2-1	1(33)	27953.0	0.0
5-3-1	269(327)	20423.9	5107.7
5-4-1	535(557)	9221.6	8002.2
5-5-1	751(758)	5271.1	5979.6

표 1의 결과로부터 2비트 패리티와 3비트 패리티가 최소 2개의 중간 유닛으로 판별 가능하다는 사실을 이용하면, N비트 패리티를 판별하기 위해서는 최소 $(N+1)/2$ (소숫점 이하는 반올림)개의 중간 유닛이 필요하다는 것을 알 수 있다.

다음으로 표 1의 5-3-1 네트워크에 대해서 고찰한다. 입력이 5비트의 경우, x_i 를 i 번째의 입력 유닛의 값, i 번째의 입력 유닛으로부터 j 번째의 중간 유닛에 향하는 결합의 가중치 계수를 w_{ji} 로 두면 다음 식과 같이 된다.

$$w_{j1} = w_j; (w_j \text{는 정수}) \dots\dots (1)$$

부호가 음수가 되는 $w_{jl} (2 \leq l \leq 5)$ 에 대해서는, 입력비트가 반전한 것으로 생각하여 $w_{l1} = -w_j$ 로 두고 이 l 번째의 입력 유닛의 값을 $(1-x_l)$ 로 한다. 이와 같이 두면 j 번째의 중간유닛의 활성화 값 $f_{aj}(x_1, x_2, x_3, x_4, x_5)$ 는 다음 식과 같다.

$$f_{aj}(x_1, x_2, x_3, x_4, x_5) = w_j(x_1 + x_2 + x_3 + x_4 + x_5) - 4w_j \dots\dots\dots (2)$$

여기에서 $X_a = (x_1 + x_2 + x_3 + x_4 + x_5); (0 \leq X_a \leq 5)$ 로 두면, 중간 유닛의 출력 f_{aj} 는 다음 식과 같이 입력 비트의 1의 수 X_a 의 함수로 나타낼 수 있다.

$$f_{aj}(X_a) = w_j(X_a - 4) \dots\dots\dots (3)$$

j 번째의 중간유닛의 문턱치를 θ_{aj} , j 번째의 중간 유닛으로부터 출력유닛에의 가중치계수를 w_{oj} , 출력유닛의 문턱치를 θ_{ak} 로 하면, 출력층의 활성화 값 $g_a(X_a)$ 를 다음 식과 같이 구할 수 있다.

$$g_a(X_a) = \sum_{j=1}^3 \left(w_{oj} \times \frac{1}{1 + \exp[-f_{aj}(X_a) + \theta_{aj}]} \right) + \theta_{ak} \dots (4)$$

따라서 출력 유닛의 출력 $z_a(X_a)$ 는 다음 식과 같이 나타낼 수 있다.

$$z_a(X_a) = \frac{1}{1 + \exp[-g_a(x_a)]} \dots \dots \dots (5)$$

이러한 관계를 그림 1과 그림 2에 나타낸다. 그림 2에서 X_a 가 우수인 경우 $z_a(X_a)$ 는 0에 가까워지고, X_a 가 기수일 때에는 1에 근접해지고 있다. 이것은 패리티 판별이 가능하다는 것을 나타낸다.

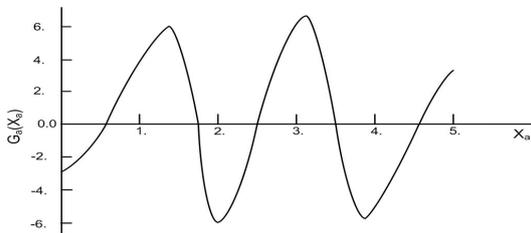


그림 1. 입력 유닛의 1의 비트수와 출력 유닛의 활성화치와의 관계

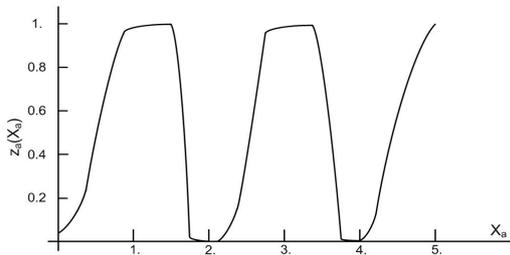


그림 2. 입력유닛의 1의 비트수와 출력과의 관계

다음으로는 입력층에서 중간층까지의 가중치 계수를 조작하는 방법으로 네트워크를 학습시켰을 때와 일반적인 방법으로 학습시켰을 때의 초기값의 조작에 대한 결과를 비교한다. 본 논문에서는 초기치의 조작에는 다음의 2종류의 방법을 실험하였다.

1. 전절과 동일한 방법. 즉, $w_{ij} = w_j$ ($1 \leq i \leq 5, 1 \leq j \leq 3$)로 하고, w_j 는 -1부터 1까지의 랜덤값을 취한다.
2. 입력 유닛의 비트의 반전도 고려한다. 위와 동일한 방법으로 초기치를 설정한 후, 이 가중치 계

수 행렬의 열 벡터 w_i 를 랜덤하게 선택하여 각 요소에 -1을 곱한다.

단, 초기치를 조작하는 이외의 방법은 일반적인 학습방법과 동일하게 실험을 하였다. 학습에 사용한 파라미터 (η, α), 수속판정오차, 최대 학습횟수는 III.1절과 동일하게 적용하여 실험을 하였다. 표 2와 같이 3-2-1 네트워크에 대하여 비교하였다.

표 2. 3-2-1 네트워크에 대한 일반적인 방법과 초기치의 조작방법 1, 2와의 비교

실험방법	수속횟수	평균 학습횟수	학습횟수의 표준편차
일반적인 방법	818	22422.6	3919.1
1의 방법	743	23709.1	3615.0
2의 방법	921	22413.5	3864.4

표 2의 결과로부터 2의 방법이 가장 양호하다는 것을 알 수 있다. 1의 방법은 일반적인 방법보다 수속 횟수가 적게 되어 있지만, "(분리가)"가 상당히 많이 있기 때문에 비교적 양호한 결과라고 말할 수 있다. 표 3은 5-3-1 네트워크에 대하여 비교하였다.

표 3. 5-3-1 네트워크에 대한 일반적인 방법과 초기치의 조작방법 1, 2와의 비교

실험방법	수속횟수 (분리가)	평균 학습횟수	학습횟수의 표준편차
일반적인 방법	1(33)	27953.0	0.0
1의 방법	4(31)	24704.0	3281.7
2의 방법	1(67)	29391.0	0.0

표 3을 "(분리가)"의 수로 평가하면, 2의 방법이 가장 양호한 결과를 내고 있다고 말할 수 있다. 이 3비트 패리티 판별과 5비트 패리티 판별에 대한 비교로부터 초기치의 조작을 하는 쪽이 수속하기 쉬운 결과가 나왔다. 이 결과로부터, 초기치의 조작이 각 파라미터를 초기치의 단계에서 해집합에 근접하는 위치에 이동한다고 생각한다면, 이 해집합과 입력층에서 중간층으로 향하는 결합의 가중치 계수가 갖추어진 것의 집합은 서로 밀접한 관계가 있다고 생각되어진다. 또한 이러한 것은 전절에서 기술한 최적해의 조건을 증명하는

것이라고 생각한다.

IV. 결론

본 논문에서는 오차역전과 알고리즘을 사용한 3층 퍼셉트론형의 신경회로망의 시뮬레이션을 실시하였다. 실험 결과로부터 N비트의 패리티 판별에 필요한 최소 중간 유닛수를 도출하여 최소 중간 유닛수에서의 최적해의 조건 등에 대해서 고찰하였다. 또한, 초기치를 조작하는 것에 의하여 학습상황이 어떻게 변화하는가를 나타냈다. 이상, 본 실험으로부터 구해진 결과를 정리하면 다음과 같다.

1. 중간 유닛수를 증가시키면 수축횟수가 많게 되어, 평균학습 횟수와 표준편차가 적게 된다.
2. N비트 패리티는 최소 $(N+1)/2$ 개의 중간 유닛수로 판별 가능하다.
3. "최소중간 유닛수로 패리티를 판별할 때, 평가함수를 최소로 하는 각 파라미터의 최적해는 각 입력 유닛으로부터 어떤 중간 유닛에 향하는 결합의 가중치 계수가 갖추어지는 것이다"라는 가능성이 높다.

향후의 연구과제로서는 평가함수를 변경하는 등의 개선책 및 엄밀한 최적해의 조건의 해결 등을 검토할 필요가 있다.

참 고 문 헌

- [1] Ooyen A. V. and Nienhuis B. "Improving the convergence of the back-propagation algorithm," Neural Networks 5, 3, pp. 465-471, 1992.
- [2] Romaniuk S. G. and Hall L. O., "Divide and conquer Neural Network," Neural Networks 6, 8, pp. 1105-1116, 1993.
- [3] J. M. Minor, "Parity with two layer feedforward nets," Neural Networks 6, 5, 1993.
- [4] Stork D. G. and Allen J. D., "How to solve the N-bit parity problem with two hidden units," Neural Networks, 5, 6, pp. 923-926, 1992.
- [5] Rumelhart D. E. and McClelland J. L. (Eds),

"Parallel Distributed Processing : Explorations in the Microstructure of Cognition," Chap. 8, MIT Press, London, pp. 334-335, 1986.

- [6] D. Rumelhart, G. Hinton and R. Williams, "Learning representations by back-propagation errors," Nature323, pp. 533-536, 1986.