

은닉노드 목표 값을 가진 2개 층 신경망의 분리학습 알고리즘

(A Separate Learning Algorithm of Two-Layered Networks
with Target Values of Hidden Nodes)

최 범 기[†] 이 주 홍^{††} 박 태 수[†]
(Bum-Ghi Choi) (Ju-Hong Lee) (Tae-Su Park)

요 약 역전파 학습 방법은 속도가 느리고, 지역 최소점이나 고원에 빠져 수렴에 실패하는 경우가 많다고 알려져 있다. 이제까지 알려진 역전파의 대체 방법들은 수렴 속도와 변수에 따른 수렴의 안정성 사이에서 불균형이라는 대가를 치루고 있다. 기존의 전통적인 역전파에서 발생하는 위와 같은 문제점 중, 특히 지역 최소점을 탈피하는 기능을 추가하여 적은 저장 공간으로 안정성이 보장되면서도 빠른 수렴속도를 유지하는 알고리즘을 제안한다. 이 방법은 전체 신경망을 은닉층-출력층(hidden to output)을 의미하는 상위연결(upper connections)과 입력층-은닉층(input to hidden)을 의미하는 하위연결(lower connections) 2개로 분리하여 번갈아 훈련을 시키는 분리 학습방법을 적용한다. 본 논문에서 제안하는 알고리즘은 다양한 classification 문제에 적용한 실험 결과에서 보듯이 전통적인 역전파 및 기타 개선된 알고리즘에 비해 계산량이 적고, 성능이 매우 좋으며 높은 신뢰성을 보장한다.

키워드 : 역전파, 분리학습, 은닉노드, 지역 최소점

Abstract The Backpropagation learning algorithm is known to have slow and false convergence aroused from plateau and local minima. Many substitutes for backpropagation announced so far appear to pay some trade-off for convergence speed and stability of convergence according to parameters. Here, a new algorithm is proposed, which avoids some of those problems associated with the conventional backpropagation problems, especially with local minima, and gives relatively stable and fast convergence with low storage requirement. This is the separate learning algorithm in which the upper connections, hidden-to-output, and the lower connections, input-to-hidden, separately trained. This algorithm requires less computational work than the conventional backpropagation and other improved algorithms. It is shown in various classification problems to be relatively reliable on the overall performance.

Key words : backpropagation, separate learning, hidden-nodes, local minima

1. 서론

인공신경망(Artificial neural network)은 망의 구성 방식에 따라 여러 가지로 분류되는데 가장 활발히 연구되고 유용한 구성은 층으로 이루어진 전방향 위상 또는 LFF로 표기되는 구성 방식이 있다. 본 논문에서 LFF

뉴론 모델은 이진 임계치 단위로써 뉴론의 단순한 모델로부터 제안된 McCulloch 와 Pitts[1]의 방법을 기반으로 하였다. 1960년대에 Frank와 Rosenblatt[2] 연구팀은 퍼셉트론이라 불리는 네트워크에 대해 집중적으로 연구하였다. 여기서 퍼셉트론이란 한 층과 그 다음 층이 전방향으로 연결된 층으로 구성된 네트워크를 의미한다. 하지만 Minsky와 Papert[3]는 그들의 저서 'Perceptrons'에서 Rosenblatt의 단층 퍼셉트론은 선형 분리의 단순한 예인 XOR 문제도 해결할 수 없다고 지적하였다. 20년 정도 후에 Rumelhartt, Hinton, Williams[4], 그리고 Parker[5]에 의해 BP알고리즘이라 알려진 역전파가 개발되었다. 비록 BP알고리즘이 네트워크에서 임의의 모든 문제를 학습할 수는 없었지만, 단순한 단층

· 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 육성 지원사업의 연구결과로 수행되었음

† 학생회원 : 인하대학교 컴퓨터정보공학과
neural@inha.ac.kr

taesu@datamining.inha.ac.kr

†† 정 회원 : 인하대학교 컴퓨터정보공학과 교수
juhong@inha.ac.kr

논문접수 : 2005년 10월 19일

심사완료 : 2006년 11월 7일

퍼셉트론으로는 해결할 수 없는 XOR문제와 같은 많은 문제를 해결할 수 있었다. 현재에는 BP알고리즘과 그와 관련된 응용분야에 많은 연구가 집중되고 있다. 본 논문에서는 다층 네트워크보다 단지 2계층 네트워크로 관심분야의 제한을 둔다. 그 이유는 본 논문에서 서술하는 모든 주제의 경우 2계층으로서 설명이 충분하고 쉽게 여러개의 층을 가진 신경망으로 확장할 수 있기 때문이다.

그림 1은 2개 층 네트워크에 관한 설명이다. 수식 표기에 관한 내용은 x_i 는 i 번째 입력을 나타내고, w_{ih^*} 는 하위연결 i 번째 입력에서 h 번째 은닉노드로 가는 가중치를, z_h 는 h 번째 은닉노드의 출력 값을, w_{hj} 는 h 번째 은닉노드에서 j 번째 출력노드로 가는 가중치를, y_j 는 j 번째 출력노드의 출력 값을 나타낸다.

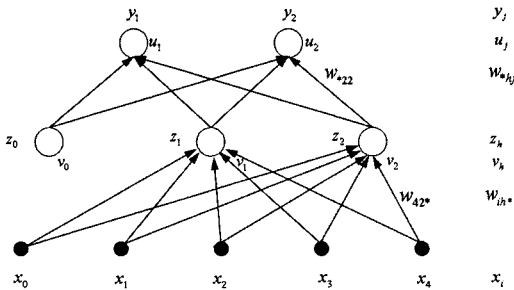


그림 1 2개 층 네트워크 구성도 및 개념도

본 논문에서의 오차 기준과 비용 함수는 다음과 같이 표기한다.

$$E[W] = \frac{1}{2} \sum_j [d_j - y_j]^2$$

$$y_j = S(u_j)$$

$$u_j = \sum_h w_{hj} z_h$$

$$z_h = S(v_h)$$

$$v_h = \sum_i w_{ih^*} x_i$$

여기서 d_j 는 j 번째 출력 노드에 대한 목표 값을 나타낸다. S 는 활성화함수(Activation function)를 나타낸다.

은닉층에서 출력층의 연결에 대한 기울기 가장 규칙은 다음과 같이 표기한다.

$$\Delta w_{hj} = -\eta \frac{\partial E}{\partial w_{hj}} = \eta (d_j - y_j) S'(u_j) z_h$$

입력층에서 은닉층 연결은 w_{ih^*} 에 의해서 편미분된다.

$$\Delta w_{ih^*} = -\eta \frac{\partial E}{\partial w_{ih^*}} = \eta \sum_j \{ (d_j - y_j) S'(u_j) w_{jh^*} \} S'(v_h) x_i$$

BP 알고리즘은 지역 최소점이나 고원 문제와 같은 수렴 문제를 내포하고 있다고 알려져 있다. 고원은 매우

느리게 수렴되는 가중치 벡터의 지역이고, 지역 최소점의 경우에는 모든 방향의 기울기가 0과 같아진다. 만약 훈련 프로세스가 지역 최소점에 빠지면, 가중치 벡터의 업데이트 프로세스는 정지하게 된다. 여기서 우리는 다음과 같은 문제를 제기한다. 'BP 알고리즘과 다른 방법들이 비용 함수의 전역 최소점에 도달함으로써 수렴을 보장할 수 있는가? 그리고, 고원에서 이른 포화상태로부터 수렴의 속도를 높일 수 있는가?'이다. 이 두 타입의 문제들은 잘못된 가중치전달 및 잘못된 인자의 선택으로 인한 것일 수 있다. 네트워크에 대한 초기 가중치에 대한 BP의 민감도는 이미 밝혀졌고 카오스[6]와 관련하여 과민성의 종류에 대하여 수학적으로 그 증거가 판명되었다. 특히 초기 가중치는 네트워크에서 훈련에 대해 수렴하기 위하여 매우 중요하다. 일반적으로 초기 가중치의 임의적인 집합은 네트워크에서 훈련에 대한 수렴을 보장하지 못한다. 지금까지 많은 방법들이 제안되어 왔고, 그 방법들은 두개의 클래스로 범주화시켰다. 그 중 한 연구 분야가 학습율과 모멘텀의 동적 변동에 관한 것이다[7-9]. 다른 방법들은 향상된 활성화함수(activation function)를 선택하거나 새로운 가중치 갱신 규칙으로 오차를 평가하는 것이다. Quickpro[10]와 탄력 있는 전파(RPROP)[11]는 빠른 수렴율을 제공하지만, 전역 최소점으로서의 수렴을 보장하지 못한다. 유전자 알고리즘[12], simulated annealing을 탄력있는 전파와 결합한 SARPRO[13]와 임의로 가중치를 변경(stochastic perturbation)하는 가중치 진화(weight evolution)[14]같은 방법들이 지역 최소점 탈피 방법으로 제시되었지만 많은 저장 공간을 요구하거나 무한히 많은 반복 횟수, 반복당 계산량의 증가 등의 단점들이 있다. 이제까지 알려진 역전파의 대체 방법들은 지역 최소점 같은 문제를 해결하기 위하여 많은 저장 공간과 계산량을 요구하거나 또는 수렴 속도는 빠르나 지역 최소점을 해결하지 못하는 부류로 나누어진다. 즉 적은 저장 공간을 이용하여 빠른 수렴을 유지하는 동시에 지역 최소점에 빠지지 않는 방법은 제시되지 않았다.

본 논문에서는 임의의 초기 가중치에 대해 더욱 유연하고, 매개변수의 넓은 범위에 대해서도 빠른 수렴을 유지하면서도, 지역 최소점을 극복할 수 있는 새로운 알고리즘을 제안한다. 새로운 알고리즘은 은닉 노드의 목표 값을 이용하여 분리 학습하는 알고리즘이다. 이 알고리즘은 출력 노드의 현재 오차로부터 은닉 노드의 목표 값을 추정하고, 입력 층에서 은닉 층, 은닉 층에서 출력 층의 두 연결 부분으로 나뉘서 훈련한다. 마지막으로 다양한 classification 문제들에 대하여 실험한다. BP알고리즘과 본 논문에서 제안하는 분리된 학습알고리즘을 반복적인 훈련에 대하여 비교하고, 분리된 학습알고리즘

이 네트워크의 전체 연결의 한 부분만을 학습하기 때문에 계산량이 훨씬 더 적을 뿐 아니라 제시된 실험들에서 경험하게 되는 지역 최소점들은 BP 및 기타 개선된 알고리즘들은 탈피하지 못하지만 제시된 알고리즘은 훌륭히 극복한다는 것을 보임으로써 우수성을 입증한다.

2. 분리 학습 방법

본 장에서는 보다 빠르고 안전하게 지역 최소점과 고원을 탈출하는 방법에 대하여 소개하도록 하겠다.

2.1 기본 개념

제한된 알고리즘에서 학습 프로세스는 두 부분으로 나뉜다. 첫 번째 부분에서 하위 연결을 고정하고 상위 연결만 학습된다. 지역 최소점이나 고원으로 인해 학습이 느려질 경우마다 학습은 강제적으로 정지하게 되고, 이번에는 반대로 상위연결을 고정하고 하위연결만 학습하게 된다. 이 프로세스는 출력 값이 목표 값으로부터 허용 오차 범위 내에 포함 될 때까지 반복하게 된다. 여기서 교통 체증 시 우회에 대한 비유는 학습 방법의 원리를 명확하게 한다. 2개의 층으로 이루어진 네트워크 학습에서 본 논문의 목표는 비용함수의 지역 최소점의 잘못된 수렴을 회피하여 전역 최소점에 빠르고 안정적으로 도달하는 방법을 찾는 것이다. BP알고리즘에서는 모든 가중치의 일괄 갱신으로 전역 최소점에 가장 짧은 경로를 주는 것처럼 보인다. 만약 비용 함수가 지역 최소점이나 고원을 포함하고 있지 않다면, BP알고리즘은 가장 좋은 탐색 방법이다. 하지만 BP알고리즘은 지역 최소점이나 고원과 같은 너무 많은 장애 요소가 존재한다. 장애 요소를 우회하는 것이 분리 학습방법의 기본 원리이다.

본 논문에서 제안하는 알고리즘은 그림 2에서 굵은 실선으로 표시하였고 BP는 가는 곡선으로 표시하였다.

좌표의 가로 방향은 하위 연결에 대한 학습의 진행을 의미하고 세로 방향은 상위 연결에 대한 학습의 진행을 의미한다. 그림에서 보면 BP가 시작점에서 전역 최소점까지의 진행시간이 거리상 짧아 보이지만 BP는 막히는 좁은 길로 그냥 가는 것이고 분리 학습은 마치 막히지 않는 큰길로 돌아가는 듯한 효과를 낼 수 있다. 우선 네트워크의 상위 연결 부분만을 학습하다가 지역 최소점 등의 장애 요소를 만났을 때 상위 연결을 동결시키고 이번에는 하위 연결 부분을 학습함으로써 장애 요소를 우회할 수 있다.

위에 언급한 장점에도 불구하고, 본 논문에서 제안된 방법은 은닉 노드의 목표 값에 대한 적절한 근사치가 없으면 사용할 수 없다. 그림 2에서 은닉 노드의 목표 값은 점선으로 표현하였다. 이제 목표 값을 추정하는 방법에 대하여 설명하도록 하겠다.

2.2 은닉 노드의 목표 값에 대한 근사치

은닉 노드의 목표 값이 얼마나 정확한지에 따라서 분리학습의 성공을 이끌어 낼 수 있다. 목표 값의 정확도에 따라 전역 최소 점으로의 학습 확률이 좌우 된다. 여기서 은닉 노드의 목표 값은 가능한 한 최소의 오차로 출력 추정치를 전파하는 은닉 노드의 값을 의미한다.

비용함수에서 변수가 은닉 노드의 값들이 된다면 이것에 대한 목표 값은 서로 다른 여러 방법에서 조사되어 왔다[16-20]. 하나의 출력 노드는 여러개의 은닉노드들을 입력변수로 하는 다변수 함수로서 표현된다. 은닉 노드들로부터 출력 노드로의 함수는 그림 3에 표현하였다. 은닉 노드의 목표 값 γ_j 를 얻기 위해서 역함수를 구해 정확한 목표 값을 구할 수는 없지만 수치해석에서 사용하는 뉴턴의 방법(Newton's Method)에 의해 근사치는 유도할 수 있다.

그림 3의 d_j 는 j번째 출력노드의 목표 값, u_j 는 j번째

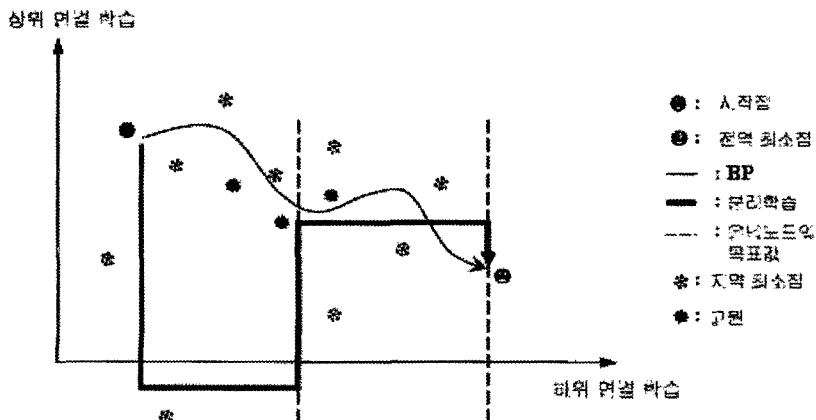
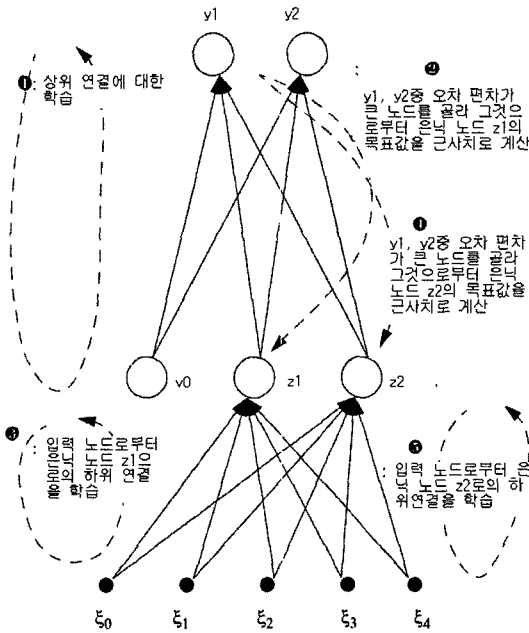


그림 2 분리 학습의 우회기법

의 다음과 같은 비용 함수를 만든다.

$$E[W] = \frac{1}{2}(\gamma_h - z_h)^2$$

3. 2의 비용함수에 대해 2과 같은 기준으로 훈련이 느려지거나 수렴할 때 까지, 입력 노드들로부터 선택된 은닉 노드와의 하위연결을 훈련시킨다.
4. 1로 되돌아간다.



학습 순서는 ①→②→③→④→⑤→⑥→⑦→⑧
이 과정은 한계 오차내로 수렴이 될 때까지 계속된다.

그림 4 분리 학습 알고리즘의 학습 절차

2.4 분리 학습의 지역 최소점 탈출에 대한 논리적 근거

분리 학습이 지역 최소점을 탈출할 수 있을지에 대한 논리적 근거는 다음과 같다. 지역 최소점이라 하면 비용 함수에 대한 모든 가중치의 편미분 값이 0이 되는 가중치 백터공간의 한 점을 의미한다. 즉 이 때 학습과 관련된 가중치 갱신 법칙에서 가중치의 변경량은 모두 0이 되어 학습이 중단되게 된다. 하지만 분리 학습에서는 그러한 상황을 만나게 되면 바로 하나의 은닉 노드의 목표 값을 설정하여 그 은닉 노드 하나에 대한 비용 함수를 최소화 하는 방향으로 학습을 진행한다. 즉 전체 비용 함수의 대한 가중치의 편미분 값이 0이라 할지라도 선택된 은닉 노드와 목표 값으로부터 창출된 비용 함수는 가중치의 편미분 값을 0으로 만들지 않을 가능성이 크다. 왜냐하면 선택된 은닉 노드로부터의 하위 연결은 전체 비용 함수와는 다른 비용 함수를 가지며 그 비용 함수

는 연결의 단순성으로 인해 지역 최소점이 발생할 가능성이 낮아지기 때문이다. 따라서 학습 과정의 완전 중지라는 극한 상황을 피할 수 있을 것으로 기대된다. 이는 제시된 실험을 통하여 BP의 경우 학습이 아무리 진행되더라도 어느 값 이하로 비용 함수의 값이 떨어지지 않는, 지역 최소점으로 의심되는 상황에서도 분리 학습은 성공적으로 수렴하는 것을 볼 수 있다.

분리 학습 방법은 가중치를 갱신하는 규칙과 네트워크 구조를 변경시키지 않고, 은닉 노드의 목표 값을 주어 학습과정을 네트워크의 위 아래로 분리하여 학습하는 과정에서 지역 최소점과 같은 함정을 만날 때마다 교대로 학습하는 방법으로 함정을 우회하는 학습 전략을 구사한다. 그러나 지역 최소 점으로부터의 탈출이 전역 최소 점으로의 수렴을 보장하는 것은 아니다. 그러한 탈출이 오히려 전역 최소점으로부터 멀어지는 상황을 맞게 될 수도 있다. 이것은 은닉 노드의 목표 값이 잘 못 설정되었을 때 발생할 것이다. 은닉 노드의 목표 값은 정확한 해가 아닌 근사치이기 때문이다. 실사 은닉 노드의 정확한 목표 값이 설정되었다 하더라도 해당 하위 연결의 훈련이 수렴된다는 보장은 하지 못한다. 그러한 단점에도 불구하고 BP보다는 높은 확률로 전역 최소점에 도달할 것이라는 것은 위의 논거에서 충분하다. 또한 계산상의 이점은 분리 학습의 1회 반복 당 시간이 BP 알고리즘보다 적게 걸린다는 사실이다. 다른 장점은 위에서 언급되었던 향상된 알고리즘들과의 결합하여 보다 개선된 성능을 기대할 수 있다. 왜냐하면, 분리 학습 방법은 단지 네트워크를 분리하여 학습하는 전략적 방향 전환임으로 각 분리된 네트워크 부분에는 위에서 언급한 어떠한 개선된 알고리즘도 적용될 수 있기 때문이다.

3 실험 및 평가

3.1 실험 환경

본 논문의 실험은 512MB RAM에 AMD XP 2600+ 2.0GB CPU의 컴퓨터에서 수행되었다. 실험 데이터로서 다음의 3가지 종류의 데이터를 사용하였다. ①합성데이터 ②Parity Data ③Iris Data. 합성데이터는 다음과 같은 방법으로 생성하였다. 입력 벡터 X, 입력노드의 수 n, 확률변수 α 값을 입력받는다. 각각의 입력 패턴 같은 -1부터 1사이의 임의의 값을 가지고, 입력패턴의 수는 10 ~ 20개를 사용하였고, 클래스의 수는 3 ~ 10개, 확률변수(α)는 입력노드의 수에 따라 3.0이상의 값을 할당하여 클래스 간에 겹치는 영역이 비교적 크도록 데이터를 생성하였다.

- C_i : 가장 가까운 클래스 μ_k^i : C_i 의 center μ^i 의
- C_j : 두 번째로 가까운 클래스 k번째 차원의 값

$$d_i = \|X - C_i\| = \left(\sum_{k=1}^n (x_k - \mu_k^i)^2 \right)^{\frac{1}{2}}$$

$$t = \frac{d_j - d_i}{d_j + d_i} \quad d_j > d_i, 0 \leq t \leq 1$$

$$P_\alpha(C_i|X) = \frac{1}{1 + e^{-\alpha t}}, P_\alpha(C_j|X) = 1 - P_\alpha(C_i|X)$$

각각의 입력벡터에 대해 모든 클래스 중심 벡터들과의 거리(d)를 계산하여 가장 가까운 클래스와 두 번째로 가까운 클래스를 찾은 후에 주어진 확률 값에 의해 소속될 클래스를 결정하였다. 실험은 은닉노드의 수를 2개에서 50개까지 증가시키면서 각각의 은닉 노드에 대하여 20번씩 반복 수행하였다.

실험은 분리학습 방법과 BP, QuickPRO, SARPRO에 대하여, 은닉노드의 증가에 따른 수렴율, 수행시간, 평균 오차를 비교, 평가하였다. 각 실험마다 약 60초의 제한 시간과 0.01의 수렴 오차 한계를 설정해서 제한 시간 내에 한계치 보다 적은 오차를 보이는 것만 성공적인 수렴율에 포함시켰다.

3.2 실험 결과

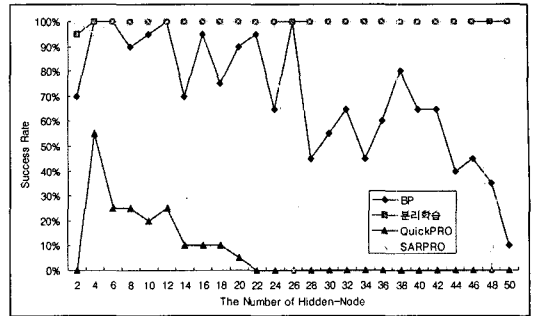
3.2.1 합성 데이터 실험결과

그림 5는 합성 데이터에 대하여 은닉노드의 수를 2개에서 50개까지 증가시켰을 때, BP, QuickPRO, SARPRO 알고리즘과 분리학습의 성능을 비교한 것이다. 수렴이 안 되었을 경우 수행 시간은 20000, 오차는 0.015로 설정하였다. BP알고리즘은 은닉 노드의 수가 증가할수록 수렴율이 낮아지며, 수행시간은 점차 증가한다. 이는 필요 이상의 많은 은닉노드의 수로 인해 네트워크의 복잡도가 올라가 많은 지역 최소점을 생성할 가능성이 높기 때문이다. QuickPRO의 경우 은닉 노드의 수가 20개 미만일 경우에만 수렴이 보장됐고, 그 이후에는 수렴하지 못하였다. 또한 수행시간이나, 평균 오차의 경우 다른 방법들에 비해 상당한 높은 것을 볼 수 있다. SARPRO 알고리즘은 수렴율, 수행 시간, 평균 오차의 모든 경우에 대하여 전반적으로 좋은 성능을 보이고 있으나 일부 조건에서 수렴하지 못했다. 반면 제안된 방법은 은닉노드의 증가에 상관없이 높은 수렴율을 보이고 있어 지역 최소점등의 문제에 있어 비교적 안전하다고 말할 수 있다. 또한, 다른 방법들에 비해 가장 빠른 수행 시간을 보이고 있다.

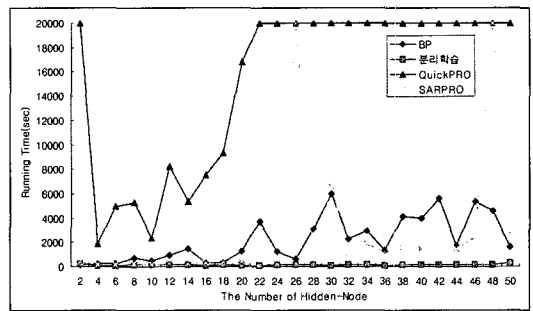
3.2.2 Parity Data 실험결과

Parity 문제는 -1과 1로 구성된 순차에서 1의 개수가 홀수일 때 1로 구분하는 문제이다. 이 문제는 지역 최소점 탈출에 관한 대표적인 문제이다. 본 논문에서는 3-3-1, 4-6-1, 5-7-1, 6-9-1, 7-11-1, 8-16-1의 parity 문제를 이용하여 제안하는 방법의 성능을 평가하였다.

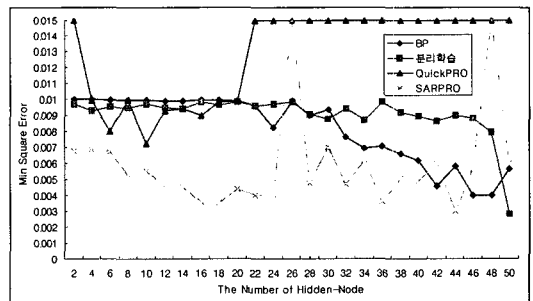
표 1은 각 비트당 수렴율과 수행시간, 최소 평균 오차



(a) 수렴율



(b) 수행 시간



(c) 평균 오차

그림 5 은닉노드 수의 증가에 따른 성능 비교

결과를 나타낸 것이다. 3비트의 경우 모든 방법들이 매번 수행마다 수렴이 되었다. 특히 본 논문에서 제안하는 분리학습 방법과 SARPRO는 매우 빠른 시간에 수행이 되어 거의 0초에 가까운 정도의 수행시간을 보이고 있다.

비트가 증가할수록 QuickPRO 방법은 수렴율이 낮아지고, 수행시간 또한 다른 알고리즘에 비해 크게 증가한 것을 볼 수 있다. BP는 8비트 parity 문제에서 수렴율이 낮아졌다. 그에 반해 SARPRO와 분리학습은 100%의 수렴율을 보이고 있다. 특히 분리학습의 경우 비트가 증가할수록, 즉 은닉 노드의 수가 증가할수록 다른 알고리즘에 비하여 수행시간 면에서 월등한 성능을 보이고 있다. 7, 8비트의 경우 SARPRO에 비해 100분의 1정도

표 1 Parity 문제에 대한 실험 결과

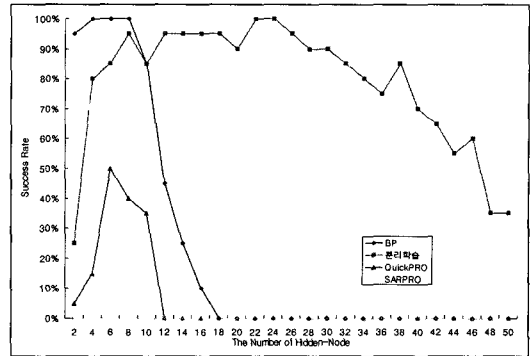
3bit			
	수렴율	수행시간	평균 오차
BP	100%	3.15	0.0489146
분리학습	100%	0	0.0487066
QuickPRO	100%	3.1	0.0489045
SARPRO	100%	0	0.0156022
4bit			
	수렴율	수행시간	평균 오차
BP	100%	3.1	0.0959723
분리학습	100%	24.2	0.0081847
QuickPRO	75%	1289.6	0.0558098
SARPRO	100%	7.05	0.0410337
5bit			
	수렴율	수행시간	평균 오차
BP	100%	8.6	0.0884058
분리학습	100%	1.6	0.0993306
QuickPRO	65%	2530.07	0.0684415
SARPRO	100%	18.75	0.0278664
6bit			
	수렴율	수행시간	평균 오차
BP	100%	65.55	0.086861
분리학습	100%	4.75	0.0907167
QuickPRO	55%	3538.09	0.0580891
SARPRO	100%	68.75	0.0577752
7bit			
	수렴율	수행시간	평균 오차
BP	100%	36.05	0.0647135
분리학습	100%	0.8	0.0983707
QuickPRO	50%	1453.3	0.0558013
SARPRO	100%	129.8	0.04354
8bit			
	수렴율	수행시간	평균 오차
BP	80%	763.6	0.043603
분리학습	100%	5.45	0.0647438
QuickPRO	0%	0	0
SARPRO	100%	351.6	0.064169

의 빠른 수행 시간을 보이고 있다. 이는 네트워크의 복잡도가 높아짐에 따라 다른 알고리즘들은 성능이 저하되는 반면에 분리학습방법은 유연하게 대처할 수 있음을 보인다.

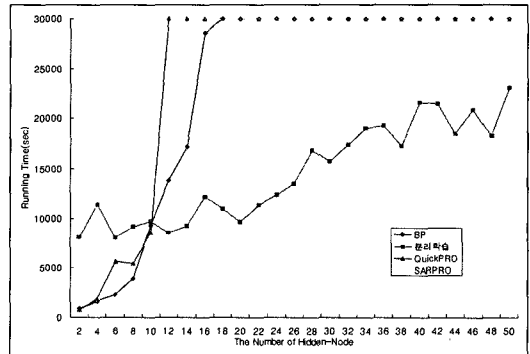
3.2.3 Iris Data 실험결과

Iris Data는 꽃받침의 길이(sepal length), 꽃받침의 두께(sepal width), 꽃잎의 길이(petal length), 꽃잎의 두께(petal width)의 4개 변수로 구성되며, 데이터의 수는 클래스의 수에 따라 50개씩 150개이다. 클래스는 붓꽃의 3가지 종류(Setosa, Versicolor, Vignica)이다.

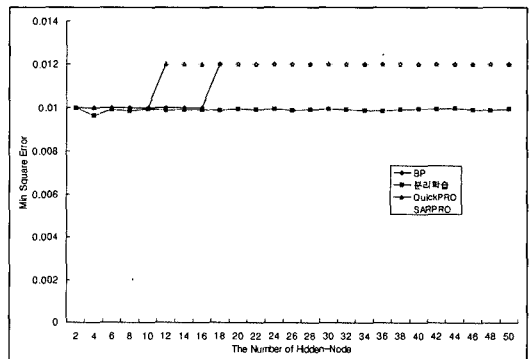
그림 6은 Iris data에 대하여 은닉 노드의 수를 2개에



(a) 수렴율



(b) 수행 시간



(c) 평균 오차

그림 6 은닉노드 수의 증가에 따른 성능 비교

서 50개까지 증가시켰을 때 BP, QuickPRO, SARPRO 알고리즘과 분리학습의 성능을 비교한 것이다. 수렴이 되지 않았을 경우 수행시간은 30000, 평균 오차는 0.012로 설정하였다. BP 알고리즘은 은닉 노드의 수가 12개 미만일 경우 매우 좋은 수렴율을 보이지만, 은닉 노드의 수가 증가할수록 수렴율이 낮아지는 것을 볼 수 있고, 수행 시간과 평균 오차의 값 또한 은닉 노드의 수가 증가할수록 큰 값을 가지는 것을 볼 수 있다. QuickPRO

알고리즘 역시 은닉 노드의 수가 적을 경우 수렴을 하지만, 은닉 노드의 수가 증가할수록 수렴하지 못하고 지역 최소점에 빠지는 것을 볼 수 있다. SARPRO 같은 경우는 전혀 수렴을 하지 못하였다. 반면에 본 논문에서 제안하는 방법은 다른 방법들에 비하여 수렴율이 높고, 수행 시간에 있어 은닉 노드의 수가 증가함에 따라 수행 시간이 증가하지만, 다른 방법들에 비해 그 시간이 매우 짧으며, 평균 오차는 은닉 노드의 증가에 관계없이 일정한 값을 가지는 것을 볼 수 있다.

전체적으로 실험 데이터와 실제데이터에서 분리학습이 은닉노드의 증가에 상관없이 수렴율과 수행시간, 평균 오차에 대하여 기존의 방법들보다 월등한 성능을 보인다.

4. 결론

지금까지 LFF에서의 분리 학습 알고리즘에 대하여 기술하였고, 임의의 classification 문제의 실험을 통하여 모멘텀을 가지는 BP, Quickpro, SARPRO 등의 알고리즘들과 성능을 비교한 결과 논리적으로 근거를 제시했듯이 지역 최소점등의 문제를 탈피하면서도 빠르고 효율적으로 수렴하는 알고리즘이라는 것이 실험결과를 통해 확인되었다. 하지만 개선되어야 측면들도 존재한다. 상위나 하위의 한 연결의 훈련에서 다른 연결의 훈련으로 언제 전환하느냐 하는 전환 시점에 관한 세련된 방법이 필요하다. 또 하나의 개선되어야 할 문제는 지역 최소 점으로부터의 탈출이 곧 전역 최소 점으로의 수렴을 보장하는 것은 아니다 라는 것이다. 이것은 은닉 노드의 목표 값에 대해 단지 근사치를 유도했기 때문이다. 보다 정확한 목표 값의 유도는 후속 연구로서 적당한 주제이다.

참 고 문 헌

- [1] McCulloch, W.S., Pitts, W., "A logical Calculus of Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5, 115-133, 1962.
 - [2] Rosenblatt, F., "Principle of Neurodynamics," New York: Spartan, 1962.
 - [3] Minsky, M.L and Papert, S.A., "Perceptrons," Cambridge: MIT Press, 1969.
 - [4] Rumelhart, D.E., G.E. Hinton, and Williams, R.J., "Learning Internal Representations by Error propagation," In *Parallel Distributed Processing*, vol. 1, chap8, 1986.
 - [5] Parker, D.B., "Learning Logic," Technical Report TR-47, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, MA, 1985.
 - [6] Kolen, J. F. and Pollack, J. B., "Back Propagation is Sensitive to Initial Conditions," *Complex System* 4, 269-280, 1990.
 - [7] Jacobs, R. A., "Increased Rates of Convergence Through Learning Rate Adaptation," *Neural Networks* 1, 293-280, 1988.
 - [8] Vogl, T. P., J.K. Magis, A.K. Rigler, W.T. Zink, and D.L. Alkon., "Accelerating the Convergence of the Back-Propagation Method," *Biological Cybernetics* 59, 257-263, 1988.
 - [9] Allred, L. G., Kelly, G. E., "Supervised learning techniques for backpropagation networks," In *Proc. of IJCNN*, vol. 1, 702-709, 1990.
 - [10] Fahlman, S. E., "Fast learning variations on backpropagation: An empirical study," in *Proc. Connectionist Models Summer School*, 1989.
 - [11] Riedmiller, M. and Braun, H., "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in *Proc. Int. Conf. Neural Networks*, vol. 1, 586-591, 1993.
 - [12] Montana D. J., Davis L., "Training feedforward neural networks using genetic algorithms," in *Proc. Int. Joint Conf. Artificial Intelligence*, Detroit, 762-767, 1989.
 - [13] Nicholas K. Treadgold and Tamas D. Gedeon., "Simulaed Annealing and Weight Decay in Adaptive Learning: The SARPRO Algorithm," *IEEE Trans. On Neural Networks*, vol. 9, pp. 662-668, 1998.
 - [14] S. C. Ng and S. H. Leung, "On Solving the Local Minima Problem of Adaptive Learning by Deterministic Weight Evolutionary Algorithm," *Proc. of Congress in Evolutionary Computation(CEC2001)*, Seoul, Korea, May 27-20, 2001, vol. I, 251-255, 2001.
 - [15] Watrous, R. L., "Learning algorithms for connectionist network: applied gradient methods
 - [16] Touretzky, D. S., "San Mateo," Morgan Kaufmann, 1989.
 - [17] Grossman, T., "The CHAIR Algorithm for Feed Forward Networks with Binary Weights," In *Advances Neural Information Processing Systems II*, 1989.
 - [18] Krogh, A., G.I. Thorbergerson, and J.A. Hertz., "A Cost Function for Internal Representations," In *Advances in Neural Information Processing Systems II*, 1989.
 - [19] Saad, D. and E. Marom., "Learning by Choice of Internal Representations-An Energy Minimization Approach," *Complex Systems* 4, 107-118, 1990.
 - [20] Saad, D. and E. Marom., "Training Feedforward Nets with Binary Weighted via a Modified CHIR Algorithm," *ComplexSystems* 4, 573-586, 1990.
- of nonlinear optimization," in *Proc. Ist. Int. Conf. Neural Networks*, vol. II, 619-628, 1987.

최 범 기

정보과학회논문지 : 소프트웨어 및 응용
제 33 권 제 7 호 참조

이 주 흥

정보과학회논문지 : 소프트웨어 및 응용
제 33 권 제 7 호 참조



박 태 수

2004년 공주대학교 정보통신공학부(학사). 2006년 인하대학교 컴퓨터정보공학과(석사). 현재 인하대학교 컴퓨터정보공학과 박사과정. 관심분야는 데이터 마이닝, 알고리즘, 신경망