한국경영과학회/대한산업공학회 2003 춘계공동학술대회
2003년 5월 16일-17일 한동대학교(포항)

다층 퍼셉트론의 새로운 두 단계 학습 알고리즘

최형준, 이재욱
(우편번호 : 790-784) 경상북도 포항시 남구 효자동 산31번지
포항공과대학교 산업공학과

이메일 : chj@postech.ac.kr
전화 : 054-279-8258, 팩스 : 054-279-2870

New Two Phases Training Algorithm for Multilayer Perceptrons

Hyoungjoon Choi, Jaewook Lee
Department of Industrial Engineering, Pohang University of Science and Technology,
Pohang, Kyungbuk 790-784, Korea.

Email : chj@postech.ac.kr
Telephone : 054-279-8258, Fax : 054-279-2870

한국경영과학회/대한산업공학회 2003 춘계공동학술대회
2003년 5월 16일-17일 한동대학교(포항)

다층 퍼셉트론의 새로운 두 단계 학습 알고리즘

New Two Phases Training Algorithm for Multilayer Perceptrons

Ab stract

본 논문에서는 다층 퍼셉트론의 학습을 위한 새로운 두 단계 학습방법을 제안하였다. 첫 번째 단계는 국소최적해로 빨리 수렴하기 위해 Levenberg-Marquardt 알고리즘을 이용한 국소 탐색 단계이다. 두 번째 단계는 첫 번째 단계에서 찾은 국소최적해가 원하는 수준에 미치지 못할 경우 새로운 국소최적해로 벗어나기 위한 선형탐색을 기반의 터널링 단계이다. 이 방법은 연결가중치 공간에서 전역최적해를 빠르게 찾을 수 있는 새로운 학습 방법을 제공한다. 4가지 벤치마크 문제에 기존의 다층 퍼셉트론의 학습 알고리즘과 비교 실험을 통해, 제안된 알고리즘이 빠른 수렴 속도와 낮은 오차값을 가짐을 알 수 있었다.

Keywords: 다층 퍼셉트론, 신경망, 학습 속도 향상, Levenberg-Marquardt, Line-search based tunneling

1. 연구배경

Rumerhart의 역전파(error back-propagation) 알고리즘이 제안된 이후로 다층 퍼셉트론(multilayer percep-trons) 신경망은 패턴인식, 로보틱스와 자동화, 금융공학, 최적화등 다양한 분야에서 어려운 문제들을 해결하는데 유용하게 쓰이고 있다.(Rumerhart, 1986) 역전파 알고리즘은 경사도(gradient)를 계산하는데 효과적이며, 최대하강법(steepest gradient descent)을 이용하여 연결가중치(weights)를 변화시켜 나간다는 점에서 비교적 간단하며 적용이 용이하다. 그러나 오차함수가 연결가중치에 대해서 평평한 경우 최적해로 수렴이 느리고, 기울기 방향을 잘못 선택했을때 국소최적해(local minimum)에 수렴해 빠져나오기가 힘들다는 문제점이 있다.(Jacobs, 1988)

역전파 알고리즘의 수렴속도를 빠르게 하기 위해서 다양한 방법들이 제안되었다. 가장 일반적인 방법으로 학습률 적응(learning rate aaptation) 방법이 있다. (Jacobs 1988), (Magoulas, 1997) 이론적으로 알고리즘의 수렴을 보장하기 위해서는 학습률을 충분히 작게 할 필요가 있다. 그러나 학습률을 너무 작게 할 경우 네트웍 학습의 수렴 속도가 느려지게 된다. 따라서 적절한 학습률 적응은 어려운 문제이다. 수렴 속도를 빠르게 하기 위한 또다른 방법으로 모멘텀(momentum)을 이용하는 방법이 있다. 모멘텀은 곡률(curvature) 정보를 이용하여 곡률이 큰 평평한 곳에서는 학습률을 크게하고, 곡률이 작은 곳에서는 학습률을 작게 해서 비교적 안정적으로 수렴 속도를 빠르게 한다.(Fahlman, 1989)

기존 연구 결과 중에 수렴 속도를 빠르게 하는 방법 중으로 오차 곡선의 이차 도함수 정보를 이용하는 방법이 수렴 속도에 큰 향상을 보였다. 학습과정에서 이차 도함수의 정보를 이용한 방법으로는 켤레경사도법(conjugate gradient)(Charalambous, 1992), 뉴튼방법(Newton's mthod)(Batti, 1992), (Becker 1989), Boyden-Fletcher-Golfarb-Shanno(BFGS)와 같은 할선법(secant methods)(Batti, 1990) 등이 있다. 이러한 방법들은 모두 비제약 최적화 방법으로 귀결된다. 이러한 비제약 최적화 방법들 중에서는 Levenberg-Marquardt 알고리즘이 가장 효율적인 알고리즘으로 알려져 있다.(More, 1977) 본 논문에서는 네트웍의 학습 알고리즘으로 기존의 역전파 알고리즘 대신 Levenberg-Marquardt 역전파 알고리즘을 적용하였다. Levenberg-Marquardt 알고리즘은 국소최적해까지 수렴속도는 빠르나 역시 국소최적해에서 빠져 전역최적해

(global minimum)로 수렴하지 못한다는 문제점을 가지고 있다. 또한 고차 도함수의 정보를 이용할 경우 기존의 역전파 알고리즘에 비해 국소최적해로 빠지는 경우 빈번하다.(Haykin 1999) 따라서 Levenberg- Marquardt 알고리즘을 이용하여, 수렴속도를 빠르게 하기 위해서는 국소최적해로 부터 벗어나 전역최적해로 이동할 수 있는 방법이 추가되어야 한다.

국소최적해를 벗어나는 기존의 연구들은 국소최적해에 빠졌을 경우, 초기 연결가중치를 다르게 주어 학습을 시키거나, 은닉층 뉴런의 개수를 변화시키거나, 연결가중치나 학습 데이터에 잡음을 추가하거나, 유전자 알고리즘을 이용하는 방법들이 이용되었다. 또한 최근에는 동적터널링 기법을 이용하여 국소최적해를 벗어나는 방법이 제안되었다.(Roychowdhury, 1999) 그러나 이 논문에서 제안한 방법은 기존의 역전파 알고리즘을 이용해서 수렴 속도의 큰 향상을 보이지 못하였다. 또한 동적 터널링의 방향을 적절히 제시하지 못해 실제적으로 구현이 어렵고, 터널링 과정에 이용되는 Lipschitz 조건의 위반을 이용하므로 결과가 안정적이지 못하다. 본 논문에서는 국소최적해를 벗어나는 방법으로 선형탐색(line search) 기반의 터널링 기법을 제안하였다.

본 논문의 2절에서는 다층 퍼셉트론의 기본적인 구조와 역전파 알고리즘의 대해서 살펴 볼 것이다. 3절에서는 새로운 학습 알고리즘과 계산 방법에 대해서 소개하고, 4절에서 새로운 학습방법을 기존의 알고리즘들과 실험을 통해서 비교, 검토할 것이다. 끝으로 5절에서 실험결과를 바탕으로 결론을 맺도록 하겠다.

2. 다층 퍼셉트론의 구조와 개요

2.1 다층 퍼셉트론의 기본구조

기본적으로 다층 퍼셉트론의 학습은 신경망의 각 층에서 정보의 흐름 방향에 따라 전방향 흐름과 역방향 흐름 두가지로 구성된다. 전방향 흐름은 입력 데이터가 입력층으로부터 네트웍의 각 뉴런에서 적용되어 마지막 출력층에서 출력값으로 나오게 되는 과정이다. 이러한 전방향 흐름동안은 신경망의 연결가중치들은 고정된 값을 갖는다. 전방향을 통해 나온 출력값을 기대 결과값과 비교해 오차를 계산한다. 이 오차 값이 역방향을 통해서 출력층으로부터 입력층까지 전달된다. 이 역방향 흐름에서는 연

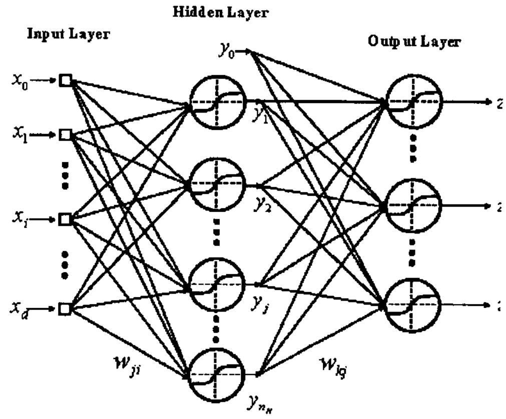

그림 1. 3개의 층을 가진 다층 퍼셉트론

결가중치들은 네트웍의 출력값이 기대 결과값과 같아지도록 변화된다. 이러한 방법의 학습 알고리즘을 역전파 알고리즘이라 한다.

2.2 역전파 알고리즘

그림1.는 3개의 층을 가진 다층 퍼셉트론을 보여주고 있다. 은닉층의 j 번째 뉴런 에서는 입력값과 연결가중치 내적값이 활성화함수(activation function)를 통과해서 y_j 가 구해진다.

$$net_j = \sum_{i=0}^{d} x_i w_{ji} = \mathbf{w}_j^T \mathbf{x} \qquad (1)$$

$$y_j = f(net_j) \qquad (2)$$

출력층의 k 번째 뉴런에서는 y_j 와 연결가중치의 내적값이 활성화 함수를 통과해서 z_k 가 구해진다.

$$net_k = \sum_{j=0}^{n_H} w_{kj} y_j = \mathbf{w}_k^T \mathbf{y} \qquad (3)$$

$$z_k = f(net_k) \qquad (4)$$

일반적으로 출력값(z_k)값과 기대 결과값(t_k) 차의 제곱합을 오차함수로 정의한다. 학습 데이터의 개수가 Q개라고 했을때 오차함수 ,,, 、는

$$J(\mathbf{w}) \equiv \frac{1}{2}\sum_{k=1}^{c}(t_k - z_k)^2 = \frac{1}{2}\|\mathbf{t} - \mathbf{z}\|^2 \qquad (5)$$

이다. 역전파 알고리즘은 오차함수를 줄이기 위해서 각

연결가중치를 변화시키는 방법이다. 이때 출력층에서부터 입력층으로 오차값이 역전파 해나가면서 각 층의 연결 가중치 값을 변화시켜 나간다. 각 층에서의 연결가중치 변화량은 다음과 같다.

- 은닉층과 출력층 사이 연결가중치의 변화량

$$\Delta w_{kj} = \eta (t_k - z_k) f'(net_k) y_j \tag{6}$$

- 입력층과 은닉층 사이 연결가중치의 변화량

$$\Delta w_{ji} = \eta \left[\sum_{k=1}^{c} w_{kj} \delta_k \right] f'(net_j) x_i \tag{7}$$

$$\delta_k = (t_k - z_k) f'(net_k) \tag{8}$$

3. 다층 퍼셉트론의 새로운 학습 알고리즘

본 논문에서 제안한 알고리즘은 두가지 단계를 갖는다. 첫 번째 단계는 신경망의 학습속도를 빠르게 하기 위해서 Levenberg-Marquardt 역전파 알고리즘을 이용해 빠르게 국소최적해를 찾는다. 첫 번째 단계에서 찾은 국소최적해가 원하는 오차 값보다 작으면 학습을 끝낸다. 그러나 오차값보다 클 경우, 선형탐색 기반의 터널링 기법을 통해서 새로운 초기 연결가중치를 찾아 첫 번째 단계를 수행한다. 터널링 과정이 실패한 경우 새로운 초기 연결가중치로 학습을 시작한다. 이러한 방법을 원하는 오차 값보다 작은 값이 나올때까지 반복적으로 수행한다.

3.1 Levenberg-Marquardt 역전파 알고리즘

본 논문에서는 신경망의 학습을 위한 역전파 과정에서 Levenberg-Marquardt 알고리즘을 적용하였다.(Hagan, 1994) 이 경우 헤시안 행렬을 계산하지 않고 뉴튼방법을 근사할 수 있어 국소최적해를 빨리 찾을 수 있다. 신경망 학습과정은 오차함수 $V(\mathbf{w})$를 $\mathbf{w}$ 에 관해서 최소화하는 과정으로, 이때 뉴튼방법을 이용하여 연결가중치를 변화시키면

이 된다. 이 때 $\nabla^2 V(\mathbf{w})$ 은 헤시안 행렬이 되고, $\nabla V(\mathbf{w})$ 은 경사도가 된다.

$$\nabla^2 V(\mathbf{w}) = J^T(\mathbf{w}) J(\mathbf{w}) \tag{10}$$

Levenberg-Marquardt 알고리즘은 이 헤시안 행렬과 경사도 벡터를 야코비안(Jacobian) 행렬을 이용해서 추정한다. 따라서 야코비안 행렬을 계산하는 것이

$$\nabla V(\mathbf{w}) = J^T(\mathbf{w}) e(\mathbf{w}) \tag{11}$$

Levenberg-Marquardt 알고리즘을 적용하는데 중요한 부분이 된다. 만약 출력값의 차원이 S이고 학습 데이터의 개수가 Q이고 연결가중치의 개수가 n일때, 야코비안 행렬은 다음과 같다. ($N = Q \times S$)

이렇게 계산한 야코비안 행렬을 Levenberg -Marquardt 알고리즘에 적용해서 연결가중치를 변화시키게 된다.

이때 매개변수 μ가 0인 경우는 뉴튼방법과 같은 방

$$J(\mathbf{w}) = \begin{bmatrix} \frac{\partial e_1(\mathbf{w})}{\partial \mathbf{w}_1} & \frac{\partial e_1(\mathbf{w})}{\partial \mathbf{w}_2} & \cdots & \frac{\partial e_1(\mathbf{w})}{\partial \mathbf{w}_n} \\ \frac{\partial e_2(\mathbf{w})}{\partial \mathbf{w}_1} & \frac{\partial e_2(\mathbf{w})}{\partial \mathbf{w}_2} & \cdots & \frac{\partial e_2(\mathbf{w})}{\partial \mathbf{w}_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial e_N(\mathbf{w})}{\partial \mathbf{w}_1} & \frac{\partial e_N(\mathbf{w})}{\partial \mathbf{w}_2} & \cdots & \frac{\partial e_N(\mathbf{w})}{\partial \mathbf{w}_n} \end{bmatrix} \tag{12}$$

법이 되고, 매개변수 μ가 큰 값을 가질 경우는 기울기 하강법(gradient descent)에 근사하게 된다. 연결가중치

$$\Delta \mathbf{w} = -\left[J^T(\mathbf{w}) J(\mathbf{w}) + \mu \mathbf{I} \right]^{-1} J^T(\mathbf{w}) e(\mathbf{w}) \tag{13}$$

를 변화 시켜 나갈 때 오차함수의 값이 감소하면, μ를 작게해서 뉴튼방법에 가깝게 하여 학습 속도를 빠르게 한다. 반대로 오차함수의 값이 증가하면 μ를 크게해서 기울기 하강법에 가깝게 변화시켜 발산하지 않도록 한다. 이렇게 연결가중치를 변화시켜 나가면 오차함수의 값이 결국 국소최소값에 도달하게 된다.

3.2 선형탐색 기반의 터널링 기법

(Roychowdhury, 1999)에서 제안한 동적 터널링 기법은 Hooke-Jeeves 패턴 탐색 방법의 변형으로 생각할 수 있다.(Deb, 1995) 임의의 동적 시스템 $g(u)$ 의 행렬 $\mathbf{A}$의 고유치(eigenvalue)가 모두 양의 실수를 가질때, 평형점 u_{eq} 를 인력체(attractor)라 한다.

$$\mathbf{A} = \partial g(u_{eq}) / \partial u \tag{14}$$

이러한 동적 시스템은 일반적으로 다음의 Lipschitz 조

$$g(u) = du / dt \tag{14}$$

건을 만족한다.

$$\left| \partial g(u_{eq}) / \partial u < \infty \right| \tag{15}$$

이 조건은 임의의 초기 조건 u_0 에서 유일한 해가 존재하는 것을 보장한다. 동적 터널링 기법은 이러한 Lipschitz 조건의 위반을 기반으로 한다. 예를들어 다음

과 같은 동적 시스템을 고려해 보자.

$$du\ /\ dt = -u^{1/3} \tag{16}$$

위 식은 $u = 0$ 에서 평형점을 갖지만 이점은 Lipschitz 조건에 위배된다.

$$\left|\frac{d(du/dt)}{du}\right| = \left|-1/3u^{-2/3}\right| \to \infty, \quad \text{as } u \to 0 \tag{17}$$

따라서 평형점에서 작은 진동을 주게되면 유한한 시간안에 다른 점으로 이동하게 된다. 이 동적 터널링은 다음의 미분방적식을 푸는 것으로 구현된다.

$$d\mathbf{w}\ /\ dt = \rho\left(\mathbf{w} - \mathbf{w}^*\right)^{1/3} \tag{18}$$

여기서 ρ 는 학습률을, $\mathbf{w}^*$ 마지막 국소최적해를 나타낸다.

(Roychowdhury, 1999)와 비교하여 본 논문에서 제안한 터널링 기법은 크게 두가지 점에서 향상된 결과를 얻었다. 첫째, 터널링 과정에서 적절한 방향을 제시하였다. 신경망 학습에서 일반적으로 초기 연결가중치의 값은 -1~1사이의 값을 갖는다. 그러나 국소최적해에 도달했을때 연결가중치 값들은 실험적으로 -10~10정도 값을 갖는다. 이 경우 연결가중치 값들이 포화되어 작은 잡음(noise)에 영향을 받지 않는다. 따라서 연결가중치 값들이 큰 값을 가지고 있을때 동적 터널링기법을 이용할 경우 국소최적해를 벗어나기 어렵다. 따라서 본 논문에서 제안하는 선형탐색 기반의 터널링 기법은 국소최적해에서의 연결가중치로부터 연결가중치의 값이 모두 0이 되는 원점으로 터널링하는 방법을 제안하였다. 연결가중치 값의 변화 방법은,

$$\mathbf{w}(k+1) = \alpha \times \mathbf{w}(k) \tag{19}$$

이다. $(0 < \alpha < 1)$ 이런 방법으로 터널링을 할 경우 연결가중치 값들이 점차 -1~1 사이 값을 갖게되어 연결가중치 값의 포화를 막게되고, 국소최적해로부터 벗어날 수 있다. 터널링 과정은 연결가중치 값들이 거의 0(-0.1~0.1)의 값을 가지게 되면 끝나게 된다.

둘째, 현재 국소최적해로부터 벗어날 가능성을 증가시켰다. (Roychowdhury, 1999)의 동적 터널링 기법은 현재 국소최적해보다 작은 오차값을 가지는 경우에 역전파 알고리즘 단계로 넘어갔다. 그러나 이 경우 새로운 초기값을 찾는데 어려움이 많으며 결국 국소최적해를 벗어나지 못하는 경우가 많았다. 본 논문에서 제안하는 방법은 터널링 과정에서 경사도가 증가하다가 감소하게 되는 즉 새로운 국소최적해를 가지는 곳에서 역

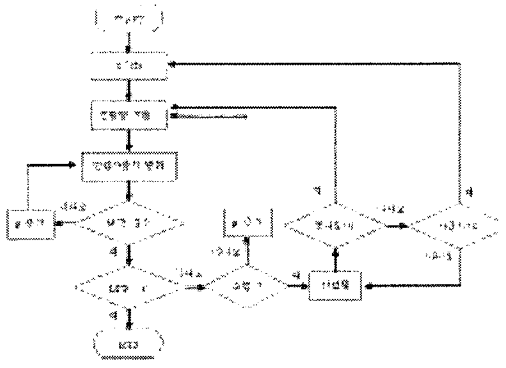

그림 2. 새로운 학습 알고리즘

전파 알고리즘 단계로 넘어갔다.

$$V\left(\mathbf{w}(k+1)\right) < V\left(\mathbf{w}(k)\right) \tag{20}$$

연결가중치를 변화시며 나가면 초기에는 오차함수가 증가할 것이다. 그러나 새로운 국소최적해를 갖는 영역이 나올 경우 오차함수는 감소하게 된다. 이때 역전파 알고리즘 단계로 돌아가서 국소최적해를 찾게 된다.

3.3 계산 알고리즘

본 논문에서 제안한 새로운 신경망 학습 방법의 계산 알고리즘 순서도는 그림 2.에 나타내었다.

알고리즘. Levenberg-Marquardt 알고리즘과 동적터널링 기법을 적용한 다층 퍼셉트론의 학습 알고리즘

초기값 설정: 신경망 구조(층 개수, 뉴런 개수), 목표 오차값, 최소경사도, 최대 훈련횟수(Epoch), α

PHASE I : Levenverg-Marquadt 알고리즘을 이용한 역전파 학습

I.1. 전방향 흐름을 통해 전체 입력값의 오차를 구하고, 오차의 제곱합을 계산한다.

I.2. 구한 오차값을 이용해 야코비안 행렬을 계산한다.

I.3. (13)식을 이용해 연결가중치의 변화량을 구한다.

I.4. 변화시킨 연결가중치 값으로 오차의 제곱합을 다시 계산한다. 오차의 제곱합이 감소했을 경우 μ 값을 작게하고, I.1과정을 반복한다. 반대로 오차의 제곱합이 증가했을 경우 μ 값을 크게하고 I.3 과정을 반복한다.

I.5. 오차가 목표 오차값보다 도달한 경우, 학습을

끝낸다. 목표 오차값에 도달하지 않고, 경사도가 최소경사도보다 작아지나 최대 훈련횟수에 도달한 경우, Phase II를 실행한다.

PHASE II : 국소 최적해를 빠져나오기 위한 터널링

II.1. 연결가중치에 α 를 곱해 가면서 변화시킨다.

II.2. 변화된 연결가중치로 네트웍의 오차의 제곱합을 구한다. 이 값이 국소최적해로부터 증가하다가 다시 감소(경사도가 감소)하는 점이 나올때 까지 II.1을 반복한다.

II.3. 연결가중치를 변화시켜 나가다가 경사도가 감소하는 점이 나타나면 그 점에서의 연결 가중치 값으로 Phase을 수행한다. 그러나 경사도가 감소하는 점을 찾지 못하고 연결가중치 값이 0에 도달할 경우, 새로운 초기 연결가중치 값으로 PahseI을 수행한다.

4. 실험결과와 검토

본 논문에서 제안한 알고리즘의 성과를 측정하기 위해서 5-Parity, 2-Spirals, Vowel, Iris, 4개의 벤치마크 문제에 적용해 기존의 알고리즘과 비교 실험을 하였다. 각 문제에서 사용된 신경망 구조와 목표오차값, 최소경사도, 최대 훈련횟수 등은 표.1에 나타내었다. 제안된 알고리즘(Proposed)은 모멘텀을 이용한 역전파 알고리즘(EBPM)과 터널링을 적용한 역전파 알고리즘(EBPT), Levenberg- Marquardt 알고리즘을 이용한 역전파 알고리즘(LM)과 비교하였다.

표 1. 각 벤치마크 문제의 신경망 구조

	5- arity	Iris	2-Spirals	Vowel
신경망 구조	5-4-1	4-6- -3	2-10-10-1	10-20-20-11
목표 오차값	0.01	0.01	0.001	0.005
최소 경사도	0.0001	0.0001	0.00001	0.0001
최대 훈련횟수	100000	100000	100000	200000
α	0.9	0.9	0.9	0.9

4.1 벤치마크 문제

5-Parity는 2진의 5차원 입력값과 2진의 1차원 출력값을 갖는 총 32개의 데이터로 구성되어 있다. **5-Parity** 문제는 5개의 입력값을 합한 값이 짝수인 경우 -1을 출력하고, 홀수인 경우 1을 출력하는 문

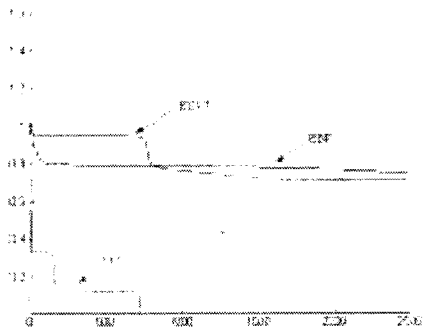

그림 3. Iris 데이터의 학습오차 변화 비교

제이다. 예를들어, XOR 문제는 입력값이 2개인 Parity 문제 할 수 있다.

Iris는 1973 Fisher의 *'The use of multiple measurements in taxonomic problems'*에 발표된 후로 지금까지 많이 사용되고는 패터인식 데이터 중의 하나이다. 4차원 입력값과 3차원의 출력값을 갖는 총 100개의 데이터로 구성되어 있다. 각각의 데이터는 3개의 집합(Sotosa, Versicolour, Virginica)으로 분류된다.

2-Spiral는 2차원 입력값과 1차원 출력값을 갖는 총 388개의 데이터이다. **2-Spiral** 데이터는 2차원 평면상에 놓인 2개의 서로 다른 소용돌이(spiral) 모양의 입력 데이터를 분류하는 문제이다. 2개의 소용돌이는 원점을 출발점으로 하여 서로 감겨있는 형태로 매우 어려운 분류 문제 중에 하나이다.

Vowel은 10차원 입력값과 2진의 11차원 출력값을 갖는 총 528개의 데이터이다. **Vowel** 데이터는 15명의 11가지 영어 모음(Vowel) 발음에 대한 데이터이다. 각각의 발음을 10가지 음향요소(acoustic feature)로 나타내었다.

4.2 실험 결과

각각의 벤치마크 문제에 제안된 알고리즘과 기존의 알고리즘들을 실험을 통해 비교하였다.(표2.) 4개 데이터 모두 제안된 알고리즘이 기존의 알고리즘 보다 훈려횟수, 학습시간, 오차율에서 향상된 결과를 보였다. 그리고 실험적으로 Levenberg-Marquardt 알고리즘을 적용했을때가 기존의 역전파 알고리즘을 적용했을때 보다 국소최적해에 더 많이

표 2. 벤치마크 문제 실험 결과

		EBPM	LM	EBPT	Proposed
5-Parity	평균 훈련횟수	131520	761	11544	99
	평균 학습시간	626.4	8.7	524.7	3.0
	평균 오차율	0.0%	0.0%	0.0%	0.0%
Iris	평균 훈련횟수	42378	1358	21809	669
	평균 학습시간	340.5	35.8	183.6	17.5
	평균 오차율	3.7%	2.9%	3.7%	3.2%
2-Spirals	평균 훈련횟수	-	1574	-	1062
	평균 학습시간	-	114.9	-	76.1
	평균 오차율	-	0.0%	-	0.0%
Vowel	평균 훈련횟수	232890	362	232890	248
	평균 학습시간	8453.6	5425.8	8453.6	2505.0
	평균 오차율	2.2%	1.8%	2.2%	1.6%

빠졌다. 2-Spirals의 경우 주어진 2-10-10-1의 신경망 구조에서는 기존의 역전파 알고리즘이 목표 오차값에 도달하지 못하였고, Vowel은 Levenberg-Marquardt 알고리즘을 적용했을 경우 터널링이 잘 되나 기존의 역전파 알고리즘을 적용했을때는 터널이 되지 않아 EBP와 EBPT가 같은 결과가 나왔다.

5. 결론 및 토의

본 논문에서는 Levenberg-Marquardt 알고리즘과 선형탐색 기반의 터널링 기법을 결합한 새로운 다층 퍼셉트론 신경망의 학습 방법을 제안하였다. 이 알고리즘은 기존의 역전파 알고리즘이 가지고 있던 느린 수렴속도와 국소최적해의 문제를 극복하였다. 본 논문에서 제안한 알고리즘 역시 초기 연력가중치에 따라 영향을 받지만 기존의 역전파 알고리즘과 터널링 알고리즘에 비해 빠른 수렴속도와 낮은 오차율을 보였다. 연결가중치의 차원이 높은 경우 터널링이 잘 되지 않아 원하는 결과를 얻지 못할 수도 있다. 이 경우 터널링의 방향과 간격이 터널링 알고리즘의 성능을 향상시키는데 중요한 요소가 될 것이다. 그리고 본 논문에서 제안한 알고리즘을 이용해서 신경망의 크기를 적절히 선택하고 테스트 오차를 줄이는 방법도 추가 연구 목표가 될 것이다.

참고 문헌

Battiti R. (1992), *First- and Second-Order Methods for Learning: Between Steepest Descent and Newton's Methods* Neural Computation, vol. 4, no. 2, pp. 141-166.

Battiti R. and Masulli F. (1990), *BFGS Optimization for Faster Autimated Supervied Learning* in Int. Neural-Netwo Conf., vol. 2, pp. 757-760.

Becker S. and Cun Y. Le (1989), *Improving the Convergence of Backpropagation Learning with Second Order Methods* in Pro. of the 1988 Connectionist Models Summer School, pp. 29-37.

Charalambous C. (1992), *Conjugate Gradient Algorithm forEfficient Training of Artificial Neural Networks* IEE Proceedings Part G, vol. 139, no. 3, pp. 301-310.

Deb K. (1995), *Optimization For Engineering Design, Algorithms and Examples*. New Delhi, India: Prentice-Hall of India.

Fahlman S. E.,(1989) *Fast Learning Variations on Back-Propagation: An Empirical Study*. in Pro. of the 1988 Connectionist Models Summer School, pp. 38-51.

Hagan Martin T. and Menhaj Mohammad B. (1994), *Training Feedforward Networks with the Marquardt Algorithm* IEEE Trans. on Neural Networks, vol.5, No. 6, pp.989-993.

Haykin S. (1999), *Neural Netwoks -A Comprehensive Foundation-* New Delhi, India: Prentice- Hall of India.

Jacobs R. A. (1988), *Increased rates of convergence through learning rate adaptation*. Nueral Networks vol.1, pp. 295-307.

Magoulas G. D., Vrahatis M. N., and Androulakis G. S. (1997), *Effective backpropagation training with variable stepsize*. Neural Networks vol.10, no. 1, pp. 69-82.

More J.J. (1997), *The Levenberg-Marquardt Algorithm: Implemntation and Theory*. Numerical Analysis, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer

Verleg, pp. 105-116.

Roychowdhury Pinaki , Singh Y. P., and Chasarkar R. A. (1999), *Dynamic Tunneling Technique for Efficient Training of Multilayer Perceptrons*. IEEE Trans. on Neural Networks, vol.10, No. 1, pp.48-55.

Rumerhart D. E., Hinton G. E., and Wiliams R. J. (1986), *Learning Internal Representations by Back-Propagation Error*. Nature, vol.323, pp.535-536.