「經營情報硏究」 第2號
大韓經營情報學會
1998年 6月

개선된 역전파법:알고리즘과 수치예제

한 홍 수[1] · 최 상 웅[2]
정 현 식[3] · 노 정 구[4]

〈목 차〉

Ⅰ. 서 론

Werbos[1], McClelland and Rumelhart[2]에 의해 정형화된 역전파법(backpropagation)은 주로 패턴인식, 제어, 예측 등에 이용되며 다층 인공신경망을 가진다. 외형적으로는 최소화문제이지만 본래의 목적은 대상의 특징을 파악하는 것이다. 오차의 최소화로 정의되는 학습은 수단에 불과하다. 모든 최소화문제(minimization problem)에서 새로운 탐색점을 결정할 때는 강하방향(함수값이 감소하는 방향)과 이 방향으로 옮겨가는 폭(step size)에 대한 정보가 요구된다. 강하방향을 설정하는 방법에 의해 여러 가지 알고리즘으로 구별되며 스텝폭을 결정하는 문제를 일차원탐색(linear search)이라 한다.

기존 역전파법은 최적화 알고리즘으로 최급강하법(steepest descent method)을 채택하고 있으며 한 점에서의 (−)기울기를 강하방향으로 설정한다. 그리고 스텝폭(학습계수)을 주어지는 상수로 간주하기 때문에 일차원탐색을 고려하지 않는다. 따라서 수치해석 문헌에서 볼 수 있는 엄격한 의미의 최급강하법이라고 말할 수는 없다. 역전파법과 관련

1) 포항전문대학 경영정보과 교수 2) 포항전문대학 경영정보과 전임강사
3) 포항전문대학 경영정보과 전임강사 4) 포항전문대학 경영정보과 전임강사

된 많은 연구결과들이 소개되었지만, 주로 응용위주였고 충분한 기초연구는 다소 부족했던 것 같다. 본 논문에서는 기존의 역전파법에 내재한 한계점들을 극복할 수 있는 개선된 알고리즘을 제시하고 응용가능성을 수치예제로 살펴보고자 한다.

개선된 알고리즘은 공액구배법(conjugate gradient method)을 토대로 한다. 공액구배법은 1952년 Hestenes and Stiefel[3]에 의해서 계수행렬이 양정치인 연립방정식을 풀기 위해서 고안되었다. 그 후 다수의 연구자들에 의해서 다양한 버전이 소개되었는데 그 중 대표적인 것으로 Fletcher and Reeves[4], Polak and Ribiere[5]의 연구를 들 수 있다.

공액(conjugacy)이란 직교(orthogonality)의 개념을 일반화시킨 것이다. 어떤 임의의 두 열벡터 x, $y(\in R^n)$가 대칭인 양정치 행렬[5](symmetric positive definite matrix) Q에 공액이란 $x^TQy=0$을 의미한다. 따라서 두 열벡터 x, $y(\in R^n)$가 직교한다는 것은 Q가 항등행렬(identity matrix)이라는 사실을 의미하며, 달리 표현하여 두 열벡터 x, $y(\in R^n)$는 항등행렬 I에 공액이라고 할 수 있다.

최급강하법에서는 한 점에서의 (−)기울기만을 정보로 하여 최적점을 탐색하며 이 탐색경로가 zigzag로 되어 비효율적인 측면을 보유하고 있다. 이러한 단점을 극복하기 위해서 한 점이 아닌 다수의 점(과거정보)을 이용하여 보다 정확한 최적점의 정보를 얻고자 하는 것이 공액구배법의 기본원리이다. 즉, 개선된 알고리즘의 강하방향은 이전의 모든 강하방향과 상호공액이 되도록 설정된다. 그리고 학습계수의 결정, 즉 일차원탐색은 먼저 학습계수에 포함된 추정요소를 순위통계량(order statistics)[6]으로 제거하여 학습계수의 폐형식(closed form)을 도출한다. 이 폐형식에 포함된 2개의 변수를 반복단계마다 황금분할법(golden section)을 적용하여 1회 계산하면, 각 층의 학습계수를 결정할 수 있다. 황금분할법을 이용한 일차원탐색은 함수값(오차)의 비교만으로 주어진 구간내에서 최적 스텝폭을 결정할 수 있다. 계산량의 측면에서 상당히 효율적인 방법이며 반복수의 증가와 함께 그 구간도 점점 축소된다.

기존 역전파 알고리즘과 비교하여, 개선된 알고리즘의 3가지 장점을 일반성, 월등한 성과, local minima의 극복으로 표현할 수 있다. 즉, 관성항(momentum)을 가지는 기존 역전파모형이 제안된 알고리즘의 특수형태라는 사실을 확인할 수 있고 수렴속도 및 정확도의 측면에서, 기존 역전파모형에 비해서 월등히 우수하며 기존 역전파모형이 갖고

5) 모든 $x\neq 0(\in R^n)$에 대해서 $x^TQx>0$이면, Q는 대칭인 양정치 행렬

6) 상호 독립적이며 동일한 분포를 가지는 확률변수들을 크기 순서로 나열한 통계량

있는 최대의 한계점이라 할 수 있는 local minima의 문제를 극복할 수 있다는 가능성을 시사한다.

*이하에서는 기존 역전파모형(혹은 알고리즘)을 *O_BP*로 개선된 알고리즘을 *N_BP*로 칭한다.

II. *O_BP*에 대한 개괄

*O_BP*의 핵심은 확정적인 부분과 확률적인 부분으로 구성되는 각 층의 weight를 갱신하는 방정식이며 식 [1]로 나타낼 수 있다.

$$W^r_{k+1} = W^r_k - \alpha G^r_k - \beta G^r_{k-1} + S^r_k \quad \text{혹은}$$

$$Vec(W^r_{k+1}) = Vec(W^r_k) - \alpha Vec(G^r_k) - \beta Vec(G^r_{k-1}) + Vec(S^r_k) \qquad [1]$$

식 [1]에서,

$*r$: 층(layer), 입력층($r=0$), 출력층($r=L$), 중간층($r=1\sim(L-1)$)

$*k$: 반복단계, $k\geq 0$

$*\alpha$: 학습계수, β : momentum계수

$W^r = [w^r_{n_r n_{r-1}}]_{N_r\times(N_{r-1}+1)}$, $S^r = [s^r_{n_r n_{r-1}}]_{N_r\times(N_{r-1}+1)}$, $G^r = [g^r_{n_r n_{r-1}}]_{N_r\times(N_{r-1}+1)}$

: r층의 weight행렬, 백색잡음 행렬, 기울기 행렬

$*N_r$: bias항을 제외한 r층의 노우드(node, 뉴런)수

$*n_r$: r층의 노우드, $n_r=N_r$이면 bias항

$*w^r_{n_r n_{r-1}}$, $s^r_{n_r n_{r-1}}$, $g^r_{n_r n_{r-1}}$: r층의 노우드 n_{r-1}과 $(r-1)$층의 노우드 n_{r-1}을 연결하는 weight, 백색잡음, 기울기

식 [1]은 행렬·벡터의 관점에서 표현한 것이며 우변에서 마지막 항이 확률적 요소를 담고 있는 백색잡음항으로서 $s^r_{n_r n_{r-1}, k}$는 평균 0, 분산이 σ^2_k인 정규분포에 따른다. 분산의 크기를 통계열역학의 유사성으로부터 온도라고 부를 수 있다. 이러한 생각은 온도라고 하는 요동을 시스템에 부가함으로써 local minima에서 벗어날 수 있다는 의도였다. 그리고 Geman형제[6], Gidas[7]의 연구결과에 의하면, k_{th}온도, 즉 σ^2_k는 식[2]의 단조감소함수로 주어진다. 식 [2]에서 초기온도를 0으로 두면, 각 층의 weight를 갱신하는 방정식이 확정적인 부분으로만 구성되게 될 것이다.

$$\sigma_k^2 = \frac{\sigma_0^2}{1+\log(k+1)},\ \forall r.\ \sigma_0^2:\ \text{초기온도} \qquad [2]$$

*O_BP*에서는 학습계수, momentum계수를 별도의 계산에 의해 결정하는 것이 아니라 경험적인 값(1보다 작은 양수)을 사전에 설정한다. 대개 관련문헌[8]에서는 [0.05, 0.25] 범위의 값을 사용할 것을 권한다. momentum항의 이용, 백색잡음항의 이용, 중간층의 노우드 수를 조정, 학습계수 및 momentum계수의 조정 그리고 초기 weight의 변경 등으로 local minima를 쉽게 해결할 수 있다는 생각을 가진 이 분야의 연구자들이 의외로 많다는 사실을 접할 때, 그러한 접근법의 한계성을 지적해 두고 싶다. 문제점의 근원은 최급강하법이 갖는 한계에서 파생된다는 사실을 강조하고 싶다.

III. *N_BP*의 수학적 배경

1. 알고리즘

*N_BP*에서는 각 층에 대해서 행렬 $H(\theta)$에 공액인 구배를 개별적으로 결정한 후, 이를 바탕으로 각 층의 weight를 갱신하는 방정식을 유도하며, 각 층에 대해서 일차원탐색을 개별적으로 수행한다. 먼저 필요한 용어를 아래와 같이 표기한다.

* $D^r = [d^r_{n_r n_{r-1}}]_{N_r \times (N_{r-1}+1)}$, $O^r = [0]_{N_r \times (N_{r-1}+1)}$: r층의 방향행렬, 영행렬

* $d^r_{n_r n_{r-1}}$: r층의 노우드 n_r과 $(r-1)$층의 노우드 n_{r-1}를 연결하는 방향

* $E(\theta) = E(Vec(W^1), Vec(W^2), \cdots, Vec(W^{L-1}), Vec(W^L))$: 목적(오차)함수

* $\nabla E(\theta) = \rho = \left[\dfrac{\partial E(\theta)}{\partial Vec(W^r)}\right]_{L\times 1}$, $\dfrac{\partial E(\theta)}{\partial Vec(W^r)} = Vec(G^r)$: $E(\theta)$의 기울기벡터

* $H(\theta) = [H_{ab}]_{L\times L}$, $H_{ab} = \dfrac{\partial^2 E(\theta)}{\partial Vec(W^b)\partial Vec(W^a)} = \left[\dfrac{\partial^2 E(\theta)}{\partial w^b_{n_b n_{b-1}}\ \partial w^a_{n_a n_{a-1}}}\right]_{p\times q}$

  : $E(\theta)$의 헤세행렬, $p = N_a \times (N_{a-1}+1)$, $q = N_b \times (N_{b-1}+1)$

  $H(\theta)$는 양정치(positive definite)인 대칭행렬

* $\| x \| = (x^T x)^{\frac{1}{2}}$

* $\theta^r = [Vec(O^1), Vec(O^2), \cdots, Vec(W^r), Vec(O^{r+1}), \cdots, Vec(O^L)]^T$

$*\lambda^r = [Vec(O^1),\ Vec(O^2),\ \cdots,\ Vec(D^r),\ Vec(O^{r+1}),\ \cdots,\ Vec(O^L)]^T$

$*\rho^r = [Vec(O^1),\ Vec(O^2),\ \cdots,\ Vec(G^r),\ Vec(O^{r+1}),\ \cdots,\ Vec(O^L)]^T$

$*\sum_r \theta^r = \theta,\ \sum_r \lambda^r = \lambda,\ \sum_r \rho^r = \rho,\ 1 \le r \le L$

$*\langle\theta^r\rangle^T = [Vec^T(O^1),\ Vec^T(O^2),\ \cdots,\ Vec^T(W^r),\ Vec^T(O^{r+1}),\ \cdots,\ Vec^T(O^L)]$

$*\langle\lambda^r\rangle^T = [Vec^T(O^1),\ Vec^T(O^2),\ \cdots,\ Vec^T(D^r),\ Vec^T(O^{r+1}),\ \cdots,\ Vec^T(O^L)]$

$*\langle\rho^r\rangle^T = [Vec^T(O^1),\ Vec^T(O^2),\ \cdots,\ Vec^T(G^r),\ Vec^T(O^{r+1}),\ \cdots,\ Vec^T(O^L)]$

$*\sum_r \langle\theta^r\rangle^T = \theta^T,\ \sum_r \langle\lambda^r\rangle^T = \lambda^T,\ \sum_r \langle\rho^r\rangle^T = \rho^T,\ 1 \le r \le L$

정규화된(normalized) 방향벡터 λ_k^r는 λ_{k-1}^r와 행렬 $H(\theta_{k-1})$에 공액이 되도록 생성되며 식[3]으로 표현된다. $1 \le r \le L$, η_{k-1}^r : scalar

$$\lambda_k^r = \frac{1}{\| -\rho_k^r + \eta_{k-1}^r \lambda_{k-1}^r \|}(-\rho_k^r + \eta_{k-1}^r \lambda_{k-1}^r),\ k \ge 1$$
$$\lambda_0^r = \frac{1}{\| -\rho_0^r \|}(-\rho_0^r) \qquad [3]$$

그러므로,

$$Vec(D_k^r) = \frac{1}{\| -Vec(G_k^r) + \eta_{k-1}^r Vec(D_{k-1}^r) \|}(-Vec(G_k^r) + \eta_{k-1}^r Vec(D_{k-1}^r)),\ k \ge 1$$
$$Vec(D_0^r) = \frac{1}{\| -Vec(G_0^r) \|}(-Vec(G_0^r)) \qquad [4]$$

λ_k^r와 λ_{k-1}^r는 행렬 $H(\theta_{k-1})$에 대해서 상호공액이므로,

$$\eta_{k-1}^r = \frac{\langle\rho_k^r\rangle^T H(\theta_{k-1})\lambda_{k-1}^r}{\langle\lambda_{k-1}^r\rangle^T H(\theta_{k-1})\lambda_{k-1}^r} = \frac{Vec^T(G_k^r) H_{rr}(\theta_{k-1}) Vec(D_{k-1}^r)}{Vec^T(D_k^{r-1}) H_{rr}(\theta_{k-1}) Vec(D_{k-1}^r)},\ k \ge 1 \qquad [5]$$

θ_{k+1}^r는 θ_k^r로부터 주어진 방향 λ_k^r으로 일차원탐색에 의해 식 [6]처럼 생성된다. τ_k^r: scalar

$$\theta_{k+1}^r = \theta_k^r + \tau_k^r \lambda_k^r,\ k \ge 0 \qquad [6]$$

즉,

$$Vec(W_{k+1}^r) = Vec(W_k^r) + \tau_k^r Vec(D_k^r),\ k \ge 0 \qquad [7]$$

τ_k^r는 step-size(N_BP의 학습계수)를 나타내며 다음의 식 [8]을 만족시킨다.

$$Min\tau_k^r\left\{E\left(\theta_k^{+r}=\sum_{i\neq r}^{L}\theta_k^i+\theta_{k+1}^r\right)=\right.$$
$$\left.E(Vec(W_k^1),\ \cdots,\ Vec(W_k^{r-1}),\ Vec(W_{k+1}^r),\ Vec(W_k^{r+1}),\ \cdots,\ Vec(W_k^L))\right\} \quad [8]$$

$E(\theta_k^{+r})$를 $\theta_k^{+r}=\theta_k$근방에서 2차 Taylor 전개하면,

$$Min\tau_k^r E(\theta_k^{+r})\cong Min\tau_k^r\left\{E(\theta_k)+\langle\rho_k^r\rangle^T\lambda_k^r\tau_k^r+\frac{1}{2}\langle\lambda_k^r\rangle^T H(\theta_k)\lambda_k^r(\tau_k^r)^2\right\} \quad [9]$$

식 [9]로부터

$$\tau_k^r=\frac{-\langle\rho_k^r\rangle^T\lambda_k^r}{\langle\lambda_k^r\rangle^T H(\theta_k)\lambda_k^r}=\frac{-Vec^T(G_k^r)Vec(D_k^r)}{Vec^T(D_k^r)H_{rr}(\theta_k)Vec(D_k^r)} \quad [10]$$

$\frac{1}{\tau_{k-1}^r}(\rho_k^r-\rho_{k-1}^r)\cong H(\theta_{k-1})\lambda_{k-1}^r$이므로 식 [5]에 대입하면,

$$\eta_{k-1}^r\cong\frac{\langle\rho_k^r\rangle^T(\rho_k^r-\rho_{k-1}^r)}{\langle\lambda_{k-1}^r\rangle^T(\rho_k^r-\rho_{k-1}^r)}=\frac{Vec^T(G_k^r)(Vec(G_k^r)-Vec(G_{k-1}^r))}{Vec^T(D_{k-1}^r)(Vec(G_k^r)-Vec(G_{k-1}^r))} \quad [11]$$

$\langle\rho_k^r\rangle^T\lambda_{k-1}^r\cong 0$[7]를 식 [11]에 적용하여,

$$\eta_{k-1}^r\cong\frac{\langle\rho_k^r\rangle^T(\rho_{k-1}^r-\rho_k^r)}{\langle\rho_{k-1}^r\rangle^T\lambda_{k-1}^r}=\frac{Vec^T(G_k^r)(Vec(G_{k-1}^r)-Vec(G_k^r))}{Vec^T(G_{k-1}^r)Vec(D_{k-1}^r)} \quad [12]$$

$\langle\rho_k^r\rangle^T\rho_{k-1}^r\cong 0$[8]를 식 [12]에 적용하여,

$$\eta_{k-1}^r\cong\frac{-\langle\rho_k^r\rangle^T\rho_k^r}{\langle\rho_{k-1}^r\rangle^T\lambda_{k-1}^r}=\frac{-Vec^T(G_k^r)Vec(G_k^r)}{Vec^T(G_{k-1}^r)Vec(D_{k-1}^r)} \quad [13]$$

이상에서 η_{k-1}^r과 관련된 3가지 유형의 식들이 소개되었는데 이를 정리한 것이 〈표 1〉이다. *N_BP* 에서는 〈표 1〉의 유형 B를 표준형으로 채택한다.

---

7) $\frac{(\rho_k^r-\rho_{k-1}^r)}{\tau_{k-1}^r}\cong H(\theta_{k-1})\lambda_{k-1}^r$와 식 [10]을 차례로 이용하여 도출

8) $\frac{(\rho_k^r-\rho_{k-1}^r)}{\tau_{k-1}^r}\cong H(\theta_{k-1})\lambda_{k-1}^r$와 $\tau_{k-1}^r\langle\lambda_{k-1}^r\rangle^T H(\theta_{k-1})\rho_{k-1}^r\cong\eta_{k-2}^r\tau_{k-1}^r\langle\lambda_{k-1}^r\rangle^T H(\theta_{k-1})\lambda_{k-2}^r-\langle\rho_{k-1}^r\rangle^T\rho_{k-1}^r$를 이용하여 도출

〈표 1〉 η^r_{k-1} 유형

유 형	식	$\langle\rho^r_k\rangle^T\lambda^r_{k-1}\cong 0$	$\langle\rho^r_k\rangle^T\rho^r_{k-1}\cong 0$	관 련 성
A	[11]	不 可	不 可	Hestenes and Stiefel version
B	[12]	可	不 可	Polak and Ribiere version
C	[13]	可	可	Fletcher and Reeves version

2. 학습계수 결정

2.1 폐형식 도출

$E(\theta^{+r}_k)\cong E(\theta_k)+\frac{1}{2}(\langle\rho^r_{k+1}\rangle^T\lambda^r_k+\langle\rho^r_k\rangle^T\lambda^r_k)\tau^r_k$이므로,

$$\tau^r_k \cong \frac{2(E(\theta^{+r}_k)-E(\theta_k))}{\langle\rho^r_{k+1}\rangle^T\lambda^r_k+\langle\rho^r_k\rangle^T\lambda^r_k} \qquad [14]$$

식 [14]는 ρ^r_{k+1}와 $E(\theta^{+r}_k)$에 대한 추정을 요구한다. 식 [14]에 의하면 τ^r_k는 확률변수이고 이 확률변수에 대한 바람직한 추정치 $(\widehat{\tau^r_k})$의 평균$-m(\widehat{\tau^r_k})-$을 τ^r_k에 대한 훌륭한 근사치로서 사용할 수 있다. N_BP에서는 $m(\widehat{\tau^r_k})$를 r층의 k단계 학습계수로 사용한다. 그러나 직접 추정절차를 요하는 것은 아니며 아래의 바람직한 3가지 특성을 충족시키는 $m(\widehat{\tau^r_k})$를 분석적으로 도출할 수 있다.

[특성 1] θ^r_{k+1}이 θ^r_k로부터 주어진 방향 λ^r_k로 존재한다.

[특성 2] L개의 확률변수 τ^1_k, τ^2_k, $\cdots$, τ^{L-1}_k, τ^L_k는 독립적이며 동일한 구간(L_k, U_k)에서 랜덤하게(일양적으로) 분포한다.

[특성 3] $\widehat{\tau_k^1} < \widehat{\tau_k^2} < \cdots < \widehat{\tau_k^{L-1}} < \widehat{\tau_k^L}$

요약하여, 식 [14]가 [특성 1, 2, 3]을 모두 만족시키면 식 [15]로 간단히 표현할 수 있다.

$$\widehat{\tau_k^r} \cong \frac{2(\widehat{E(\theta_k^{+r})} - E(\theta_k))}{\langle \rho_k^r \rangle^T \lambda_k^r}, \ L_k < \widehat{\tau_k^1} < \widehat{\tau_k^2} < \cdots < \widehat{\tau_k^{L-1}} < \widehat{\tau_k^L} < U_k \quad [15]$$

식 [15]로부터,

$$\tau_k^r \approx m(\widehat{\tau_k^r}) \cong \frac{2\{m(\widehat{E(\theta_k^{+r})}) - E(\theta_k)\}}{\langle \rho_k^r \rangle^T \lambda_k^r}$$
$$L_k < \tau_k^1 \approx m(\widehat{\tau_k^1}) < \tau_k^2 \approx m(\widehat{\tau_k^2}) < \cdots\cdots < \tau_k^{L-1} \approx m(\widehat{\tau_k^{L-1}}) < \tau_k^L \approx m(\widehat{\tau_k^L}) < U_k \quad [16]$$

$\tau_k^{(r)}$를 $(\tau_k^1, \tau_k^2, \cdots, \tau_k^{L-1}, \tau_k^L)$ 중에서 r_{th}로 작은 값이라 정의한다. 그러면 이 정의로부터 $\tau_k^{(1)} < \tau_k^{(2)} < \cdots\cdots < \tau_k^{(L-1)} < \tau_k^{(L)}$는 $(\tau_k^1, \tau_k^2, \cdots, \tau_k^{L-1}, \tau_k^L)$에 대한 순위통계량(order statistics)이다. 따라서 $\tau_k^{(r)}$의 밀도함수 및 분포함수는 식 [17]과 같다.

$$f_{\tau_k^{(r)}}(\tau) = \frac{L!}{(L-r)!(r-1)!}\left(\frac{1}{U_k - L_k}\right)^L (\tau - L_k)^{r-1}(U_k - \tau)^{L-r}, \ L_k < \tau < U_k$$
$$F_{\tau_k^{(r)}}(\tau^*) = \int_{L_k}^{\tau^*} f_{\tau_k^{(r)}}(\tau)d\tau, \ L_k \le \tau^* \le U_k \quad [17]$$

[특성 3]으로부터,

$$m(\widehat{\tau_k^r}) = m(\tau_k^{(r)}) \approx \tau_k^r \quad [18]$$

식 [17]을 이용하여

$$m(\tau_k^{(r)}) = \left\{ r_L C_r \sum_{x=0}^{L-r} \frac{{}_{L-r}C_x(-1)^x}{x+r+1} \right\} U_k + \left\{ r_L C_r \sum_{x=0}^{L-r} \frac{{}_{L-r}C_x(-1)^x}{(x+r)(x+r+1)} \right\} L_k \quad [19]$$

식 [16], [18]에 의해

$$r_L C_r \sum_{x=0}^{L-r} \frac{{}_{L-r}C_x(-1)^x}{x+r} = 1 \quad [20]$$

이제 식 [20]과 $\sum_{x=0}^{m} \frac{{}_mCx(-1)^x}{x+1} = \frac{1}{m+1}$를 이용하면,

$$\sum_{x=0}^{L-r} \frac{{}_{L-r}C_x(-1)^x}{x+r+1} = \frac{1}{(L+1)_L C_r}, \quad 1 \le r \le L$$

식 [20]과 [21]로부터 식 [19]는 다음의 폐형식 [22]로 변형된다.

$$m(\tau_k^{(r)}) = m(\widehat{\tau_k^r}) = \left(\frac{r}{L+1}\right)U_k + \left(\frac{L-r+1}{L+1}\right)L_k \approx \tau_k^r \qquad [22]$$

식 [22]는 반복단계 k에서 r층에 이용되는 학습계수의 최적 근사치이다.

2.2 황금분할법에 의한 L_k, U_k결정

식 [22]의 L_k, U_k는 $E(\theta_k)$, $E(\theta_{k-1})$의 비교를 통해서, 존재 구간이 각각 설정된 다음, 이 각 구간에서 최종 결정된다. 반복수의 증가와 함께 오차는 0으로 수렴할 것이고 또한 그렇게 되도록 각 구간이 설정되어 그 각 구간에서 L_k, U_k는 결정될 것이다. 먼저 L_k, U_k가 존재하는 구간을 각각 (b_k^L, e_k^L), (b_k^U, e_k^U)로 정의한다.

① $E(\theta_{k-1} \ge E(\theta_k)$ for all $k \ge 2$,

(1) $(b_k^L, e_k^L) = (p_{1.k-1}^L, e_{k-1}^L)$, $(b_k^U, e_k^U) = (p_{1.k-1}^U, e_{k-1}^U)$

(2) $p_{1.k-1}^L = b_{k-1}^L + \frac{3-\sqrt{5}}{2}(e_{k-1}^L - e_{k-1}^L)$, $\quad p_{1.k-1}^U = b_{k-1}^U + \frac{3-\sqrt{5}}{2}(e_{k-1}^U - b_{k-1}^U)$

(3) $p_{1.k}^L = p_{2.k-1}^L = b_{k-1}^L + \frac{\sqrt{5}-1}{2}(e_{k-1}^L - b_{k-1}^L)$, $p_{1.k}^U = p_{2.k-1}^U = b_{k-1}^U + \frac{\sqrt{5}-1}{2}(e_{k-1}^U - b_{k-1}^U)$

(4) $p_{2.k}^L = L_k = b_k^L + \frac{\sqrt{5}-1}{2}(e_k^L - b_k^L)$, $\quad p_{2.k}^U = U_k = b_k^U + \frac{\sqrt{5}-1}{2}(e_k^U - b_k^U)$

$$\tau_k^r \approx \left(\frac{r}{L+1}\right) p_{2,k}^U + \left(\frac{L-r+1}{L+1}\right) p_{2.k}^L \qquad [23]$$

② $E(\theta_{k-1}) < E(\theta_k)$ for all $k \ge 2$,

(1) $(b_k^L, e_k^L) = (b_{k-1}^L, p_{2,k-1}^L)$, $(b_k^U, e_k^U) = (b_{k-1}^U, p_{2,k-1}^U)$

(2) $p_{2.k-1}^L = b_{k-1}^L + \frac{\sqrt{5}-1}{2}(e_{k-1}^L - b_{k-1}^L)$, $\quad p_{2.k-1}^U = b_{k-1}^U + \frac{\sqrt{5}-1}{2}(e_{k-1}^U - b_{k-1}^U)$

(3) $p_{2.k}^L = p_{1.k-1}^L = b_{k-1}^L + \frac{3-\sqrt{5}}{2}(e_{k-1}^L - b_{k-1}^L)$, $\quad p_{2.k}^U = p_{1.k-1}^U = b_{k-1}^U + \frac{3-\sqrt{5}}{2}(e_{k-1}^U - b_{k-1}^U)$

(4) $p^L_{1,k}=L_k=b^L_k+\frac{3-\sqrt{5}}{2}(e^L_k-b^L_k)$, $p^U_{1,k}=U_k=b^U_k+\frac{3-\sqrt{5}}{2}(e^U_k-b^U_k)$

$$\tau^r_k\approx\left(\frac{r}{L+1}\right)p^U_{1,k}+\left(\frac{L-r+1}{L+1}\right)p^L_{1,k} \qquad [24]$$

③ $k=0$

$$\tau^r_0\approx\left(\frac{r}{L+1}\right)\left(b^U_0+\frac{3-\sqrt{5}}{2}(e^U_0-b^U_0)\right)+\left(\frac{L-r+1}{L+1}\right)\left(b^L_0+\frac{3-\sqrt{5}}{2}(e^L_0-b^L_0)\right) \qquad [25]$$

④ $k=1$

$$\tau^r_1\approx\left(\frac{r}{L+1}\right)\left(b^U_1+\frac{\sqrt{5}-1}{2}(e^U_1-b^U_1)\right)+\left(\frac{L-r+1}{L+1}\right)\left(b^L_1+\frac{\sqrt{5}-1}{2}(e^L_1-b^L_1)\right) \qquad [26]$$

그리고 $b^U_1=b^U_0$, $e^U_1=e^U_0$, $b^L_1=b^L_0$, $e^L_1=e^L_0$.

이제 초기치를 결정하기 위해 $Max_{k,r}\,\tau^r_k=Max_k\,\tau^r_k<C$라고 가정하자. 여기서 C는 주어지는 양의 상수이다.(대개 1) 그러면, $\frac{U_0}{L+1}+\frac{L\cdot L_0}{L+1}>0$, $U_0>L_0$, $\frac{L\cdot U_0}{L+1}+\frac{L_0}{L+1}<C$제약하에서 영역 $(b^U_0<U_0<e^U_0)\cap b^L_0<L_0<e^L_0)$를 최대화시키는 초기치를 다음과 같이 구할 수 있다.

$$b^L_0=0,\ b^U_0=e^L_0=\frac{C}{2},\ e^U_0=C\left(1+\frac{1}{2L}\right) \qquad [27]$$

Ⅳ. 수치예제

이용되는 수치예제는 XOR문제와 수치해석 문헌에서 자주 언급되는 Rosenbrock 함수[9]를 이용하여 함수값을 추정하는 문제이다. XOR 문제는 교사입력에는 2bit, 교사출력에는 1bit가 필요하다. 반면에, Rosenbrock 함수를 이용하여 함수값을 추정하는 문제는 그 자체가 패턴으로 간주될 수 없는 경우로서 다음의 절차를 이용한 변환과정이 필요하다.

구간 $[I_1, I_2]$에서 정밀도(소수이하 자리수) t를 가지는 임의의 실수 R를 이진수로 고칠 수 있는데 2가지 계산이 요구된다.

① 실수 R을 십진수 R^*으로 전환시킨다.

9) $f(x_1, x_2)=100(x_2-x_1^2)^2+(1-x_1)^2$

$$R^* = NINT\left(\frac{(R-I_1)(2^b-1)}{I_2-I_1}\right),\ NINT(*)\ :\ *\text{를 반올림하여 정수형으로 전환}$$

$$2^{b-1} < (I_2-I_1)10^t \leq 2^b,\ b\text{는 정수.} \qquad [28]$$

② 식 [28]에서 구한 십진수를 이진수로 고친다.

따라서 Rosenbrock 함수값 추정문제에서는 교사입력에 2b bit, 교사출력에는 b bit가 필요하다. 그리고 b bit로 표현되는 이진수를 구간 [I_1, I_2]상의 정밀도(소수이하 자리수) t를 가지는 임의의 실수 R로 역변환시킬 수 있다.

*XOR*문제의 실험환경은 〈표 2〉, Rosenbrock 함수값 추정문제의 실험환경은 〈표 3〉에 각각 제시되어 있다. 특히, Rosenbrock 함수값 추정문제에 사용된 학습자료는 구간 [0, 1]에서 정밀도 2를 가지는 난수를 발생시켜 마련하였다. 이 경우 처음 계산된 함수값은 0과 1 사이의 값으로 scaling 과정을 거친 후, 십진수로 변환되어야 한다. 그 이유는 표준형의 sigmoid 출력함수를 사용하기 때문이다. 그리고 주어진 환경하에서 2가지 문제를 〈표 4〉의 4가지 상황을 가정하여 실험을 수행한다.

전술한 실험환경을 Borland C++ 컴파일러(Ver. 3.1)를 이용하여 프로그래밍하였다. 정규화된 초기 weight를 사용자가 직접 정의하거나 난수를 이용하여 발생시키도록 하였다. 십진수 및 이진수 계산 그리고 패턴을 생성하는 난수발생, 함수값 scaling 등의 부수적인 프로그램은 별도의 코드로 처리하였다.

〈표 2〉 *O_BP*, *N_BP* 실험환경(*XOR* 문제)

층	3층	출력함수	표준형의 sigmoid 함수
입력층 노우드	2개	오차수준	10^{-12}
중간층 노우드	2개	최대반복수	3,000
출력층 노우드	1개	결과 출력간격	100
bias항	포함	학습자료	4개

〈표 3〉 *O_BP*, *N_BP* 실험환경(Rosenbrock 함수값 추정문제)

층	3층	출력함수	표준형의 sigmoid 함수
입력층 노우드	14개	오차수준	10^{-12}
중간층 노우드	10개	최대반복수	5,000
출력층 노우드	7개	결과 출력간격	100
bias항	포함	학습자료	20개

〈표 4〉 2가지 사례의 실험상황

실험	O_BP	N_BP
1	학습계수 : 0.3, momentum계수 : 0.15 초기온도 : 0.000	상수 C : 0.3
2	학습계수 : 0.2, momentum계수 : 0.1 초기온도 : 0.000	상수 C : 0.2
3	학습계수 : 0.3, momentum계수 : 0.15 초기온도 : 0.001	상수 C : 0.15
4	학습계수 : 0.2, momentum계수 : 0.1 초기온도 : 0.001	상수 C : 0.1

V. 실험결과 및 결론

XOR 문제의 〈표 5〉와 Rosenbrock 함수값 추정문제의 〈표 6〉은 *O_BP* 및 *N_BP*에 대한 4가지 실험상황의 결과를 나타내고 있다. 계산종료시의 반복수, 오차 그리고 소요된 계산시간 등이 기록되어 있다. 이 실험결과로 3가지 사실들을 알 수 있다. 첫째, 확률항을 가지는 *O_BP*가 그렇지 않는 경우보다 좋은 성과를 가져다 줄 수 있다. 둘째, *N_BP*에서는 양의 상수 *C*가 클수록 더 좋은 결과를 얻었지만, *C*값이 지나치게 커지면 역효과가 발생할 수도 있다. 대체로 0과 1사이의 값이면 무난할 것이다. 셋째, *N_BP*는 수렴속도 및 정확도 모두에서 *O_BP*를 능가하였다. 실험에 이용된 상황보다 더 좋은 조건에서 매우 큰 반복을 수행하더라도, *N_BP*에서 얻을 수 있는 정밀도의 오차를 *O_BP*에서 확보할 수는 없었다. 즉, *N_BP*가 local minima의 문제를 해결할 수 있다는 사실을 간접적으로 암시하고 있다. 그 근거는 다음과 같다.

역전파모형의 목적함수는 오차제곱의 합에 상수배를 한 것이므로 당연히 최소값은 0이다. 따라서 유한한(가급적 적은) 반복을 거쳐서 거의 0이라고 할 수 있는 오차를 확보해야 한다. 그렇지 않으면, 비록 확률적 요소를 가미한다 하더라도, 소위 local minima에 빠져서 벗어날 수 없다. *O_BP*는 반복수를 증가시켜도 *N_BP*에서 확보할 수 있는 수준의 오차에는 도달할 수 없다. 이 점은 *O_BP*의 알고리즘 그 자체의 한계라 할 수 있다. 반면에 *N_BP*에서는 적은 수의 반복만으로도 거의 0이라 간주할 수 있는 수준의 오차에 도달할 수 있다. 이러한 사실들을 정리하면, *N_BP*가 local minima의 문제를 실질적으로 해결할 수 있다는 것을 의미한다. 이러한 맥락에서, 최대반복수를 10^5(십만)으로 하여 *N_BP*에 대해서는 오차수준을 10^{-12}에서 10^{-16}으로, *O_BP*에 대해서는 오차수준을 10^{-12} 10^{-6}으로 하여 추가실험을 수행하였다. 기타의 실험환경은 〈표 2〉, 〈표 3〉과 변함이 없다. 실험상황은 〈표 5〉, 〈표 6〉에서 가장 좋은 성과를 보여준 것을 선택했다. 〈표 5-1〉과 〈표

6-1〉의 추가 실험결과로 다음과 같은 사실들을 발견하였다. 첫째, *O_BP*는 반복수를 증가시키고 오차수준을 낮추더라도, 0에 가까운 오차를 확보할 수 없었다. 둘째, *N_BP*는 비록 오차수준을 높이더라도 상대적으로 적은 반복수만 추가로 요구될 뿐, 거의 0에 가까운 오차를 확보할 수 있었다. 결국, *N_BP*는 local minima의 문제를 충분히 해결할 수 있다는 것을 간접적으로 재확인할 수 있었다.

*O_BP*에서는 가장 좋은 성과를 보인 실험 3 그리고 *N_BP*에서는 실험 1, 2, 3, 4 모두의 테스트 결과를 *XOR* 문제의 〈표 7〉, Rosenbrock 함수값 추정문제의 〈표 8〉, 〈표 9〉에 요약하였다. *XOR* 문제에서는 정밀도 6[10]으로 테스트 결과를 기록하였고, Rosenbrock 함수값 추정문제에서는 정밀도 6의 테스트 결과와 이에 해당하는 십진수 그리고 이 십진수의 0과 1사이의 수치(정밀도 2)[11] 를 모두 기록하였다. 〈표 8〉의 *O_BP*에서는 테스트 성공률이 65%이지만, 〈표 9〉의 *N_BP*에서는 4가지 실험 모두에서 100%의 테스트 성공률을 보이고 있다.

정밀도와 테스트 성공률과 관련하여, Rosenbrock 함수값 추정문제에서 정밀도 4 및 5인 경우의 테스트 결과를 〈표 8-1〉, 〈표 9-1〉, 〈표 8-2〉, 〈표 9-2〉에 요약하였다. 테스트 성공률은 정밀도 6의 경우와 모두 동일하였다.

이상의 실험결과들을 정리하여, 첫째, *O_BP*의 한계는 분명히 알고리즘 그 자체의 한계성이라는 사실을 간접적으로 확인할 수 있었고 지엽적 문제해결 방식들은 결코 이 한계를 극복할 수 없다는 사실이 명백해졌다. 둘째, 이러한 *O_BP*가 갖는 한계를 극복하는 하나의 대안으로서 *N_BP*는 충분한 응용가치가 있다는 것을 확인하였다. 마지막으로, *N_BP*와 관련하여 초기 weight를 어떻게 추출하여 정규화 혹은 그대로 사용할 것인가라는 문제는 앞으로 좀더 논의되어야 할 부분이다.

10) 원시(source)프로그램은 대부분의 실험결과를 파일로 출력하도록 코딩되어 있다. 이러한 이유로 오차(정밀도 12)를 제외한 나머지 실험결과치들은 모두 정밀도 6으로 기록되도록 하였다.

11) 별도의 코드로 계산

〈표 5〉 실험결과(*XOR* 문제)

실험상황	*O_BP*			*N_BP*		
	실행시간	반복수	오차	실행시간	반복수	오차
1	1.43	3000	0.008530726532	0.16	500	0.000000000000
2	1.43	3000	0.095760896064	0.27	800	0.000000000000
3	1.43	3000	0.008250918171	0.38	1000	0.000000000000
4	1.43	3000	0.084207681382	0.49	1400	0.000000000001

〈표 5-1〉 추가 실험결과(*XOR* 문제)

실험상황	*O_BP*			*N_BP*		
	실행시간	반복수	오차	실행시간	반복수	오차
3(*O_BP*) 1(*N_BP*)	54.84	100,000	0.000070653689	0.27	700	0.0000000000000000

〈표 6〉 실험결과(Rosenbrock 함수값 추정문제)

실험상황	*O_BP*			*N_BP*		
	실행시간	반복수	오차	실행시간	반복수	오차
1	95.00	5000	0.007649889714	19.89	1200	0.000000000001
2	95.00	5000	0.012565657528	29.62	1800	0.000000000000
3	94.84	5000	0.007521271519	37.80	2300	0.000000000001
4	94.84	5000	0.011901285573	57.47	3500	0.000000000000

〈표 6-1〉 추가 실험결과(Rosenbrock 함수값 추정문제)

실험상황	*O_BP*			*N_BP*		
	실행시간	반복수	오차	실행시간	반복수	오차
3(*O_BP*) 1(*N_BP*)	2293.13	100,000	0.000338804223	37.53	2000	0.0000000000000000

〈표 7〉 테스트 결과(*XOR* 문제)

입력	교사출력	*O_BP*(실험 3)	*N_BP*(실험 1, 2, 3, 4)
0 0	1	0.934254	1.000000
0 1	0	0.061736	0.000000
1 0	0	0.062078	0.000000
1 1	1	0.932931	1.000000

〈표 8〉 테스트 결과(Rosenbrock 함수값 추정문제 $-O\_BP$(실험 3)) : 정밀도 6

입력	교사출력			O_BP(실험 3) 출력		
0000001 0000000	0000001	1	0.01	0.000022 0.004794 0.002174 0.004470 0.000399 0.006432 0.999620	1	0.01
0101011 0000100	0000011	3	0.02	0.000012 0.010878 0.001369 0.013687 0.000210 0.992575 0.984732	3	0.02
0101100 0011100	0000011	3	0.02	0.000004 0.005105 0.000226 0.005875 0.000442 0.990595 0.997486	3	0.02
1000101 0011001	0000001	1	0.01	0.000001 0.008100 0.000036 0.006517 0.009125 0.005344 0.989026	1	0.01
1011001 1111001	0101011	43	0.34	0.011970 0.985131 0.004790 0.983314 0.012152 0.993404 0.999958	43	0.34
0100010 0111000	0011101	29	0.23	0.012452 0.003740 0.982409 0.986373 0.994472 0.015589 0.999999	30	0.24
0001110 1011001	1100011	99	0.78	0.985364 0.999974 0.014773 0.010458 0.000218 0.989936 0.992906	98	0.77
1000111 0000101	0001111	15	0.12	0.000096 0.009453 0.014000 0.986843 0.987209 0.990537 0.994914	15	0.12
0010110 1100111	1111111	127	1.00	0.980148 0.990029 0.983549 0.997152 0.996583 0.985134 0.979257	125	0.98
1011000 1100001	0010001	17	0.13	0.007159 0.008156 0.987498 0.010257 0.007069 0.006364 0.991524	18	0.14
1101001 1111010	0001111	15	0.12	0.000027 0.010589 0.011263 0.993866 0.990300 0.989811 0.988065	15	0.12
0011100 0110111	0011110	30	0.24	0.012541 0.007921 0.975455 0.993836 0.979083 0.999933 0.011296	31	0.24
1111001 1101011	0000000	0	0.00	0.000014 0.001129 0.003716 0.009150 0.012355 0.002996 0.020099	0	0.00
1110101 1100111	0000000	0	0.00	0.000006 0.000145 0.009908 0.007524 0.019025 0.007585 0.010007	0	0.00
0111001 1001100	0100001	33	0.26	0.000677 0.990465 0.000172 0.001840 0.000380 0.007084 0.999922	33	0.26
1010100 1001100	0000101	5	0.04	0.000001 0.004409 0.001341 0.007789 0.982854 0.000022 0.991379	5	0.04
1000110 1011011	0100100	36	0.28	0.006218 0.986991 0.007432 0.006742 0.991621 0.014011 0.012568	36	0.28
0001110 0110100	0100010	34	0.27	0.016987 0.974574 0.005319 0.015670 0.017997 0.999991 0.020200	35	0.28
0001111 1010101	1011010	90	0.71	0.981223 0.022403 0.990321 0.981805 0.002299 0.999973 0.008632	89	0.70
0111100 0111110	0001111	15	0.12	0.000011 0.013465 0.017989 0.993010 0.993609 0.998208 0.993103	16	0.13

〈표 8-1〉 테스트 결과(Rosenbrock 함수값 추정문제 $-O\_BP$(실험 3)) : 정밀도 4

입력	교사출력			O_BP(실험 3) 출력		
0000001 0000000	0000001	1	0.01	0.0000 0.0048 0.0022 0.0045 0.0004 0.0064 1.0000	1	0.01
0101011 0000100	0000011	3	0.02	0.0000 0.0109 0.0014 0.0137 0.0002 0.9926 0.9847	3	0.02
0101100 0011100	0000011	3	0.02	0.0000 0.0051 0.0002 0.0059 0.0004 0.9906 0.9975	3	0.02
1000101 0011001	0000001	1	0.01	0.0000 0.0081 0.0000 0.0065 0.0091 0.0053 0.9890	1	0.01
1011001 1111001	0101011	43	0.34	0.0120 0.9851 0.0048 0.9833 0.0122 0.9934 1.0000	43	0.34
0100010 0111000	0011101	29	0.23	0.0125 0.0037 0.9824 0.9864 0.9945 0.0156 1.0000	30	0.24
0001110 1011001	1100011	99	0.78	0.9854 1.0000 0.0148 0.0105 0.0002 0.9899 0.9929	98	0.77
1000111 0000101	0001111	15	0.12	0.0001 0.0095 0.0140 0.9868 0.9872 0.9905 0.9949	15	0.12
0010110 1100111	1111111	127	1.00	0.9801 0.9900 0.9835 0.9972 0.9966 0.9851 0.9793	125	0.98
1011000 1100001	0010001	17	0.13	0.0072 0.0082 0.9875 0.0103 0.0071 0.0064 0.9915	18	0.14
1101001 1111010	0001111	15	0.12	0.0000 0.0106 0.0113 0.9939 0.9903 0.9898 0.9881	15	0.12
0011100 0110111	0011110	30	0.24	0.0125 0.0079 0.9755 0.9938 0.9791 0.9999 0.0113	31	0.24
1111001 1101011	0000000	0	0.00	0.0000 0.0011 0.0037 0.0092 0.0124 0.0030 0.0201	0	0.00
1110101 1100111	0000000	0	0.00	0.0000 0.0001 0.0099 0.0075 0.0190 0.0076 0.0100	0	0.00
0111001 1001100	0100001	33	0.26	0.0007 0.9905 0.0002 0.0018 0.0004 0.0071 0.9999	33	0.26
1010100 1001100	0000101	5	0.04	0.0000 0.0044 0.0013 0.0078 0.9829 0.0000 0.9914	5	0.04
1000110 1011011	0100100	36	0.28	0.0062 0.9870 0.0074 0.0067 0.9916 0.0140 0.0126	36	0.28
0001110 0110100	0100010	34	0.27	0.0170 0.9746 0.0053 0.0157 0.0180 1.0000 0.0202	35	0.28
0001111 1010101	1011010	90	0.71	0.9812 0.0224 0.9903 0.9818 0.0023 1.0000 0.0086	89	0.70
0111100 0111110	0001111	15	0.12	0.0000 0.0135 0.0180 0.9930 0.9936 0.9982 0.9931	16	0.13

〈표 8-2〉 테스트 결과(Rosenbrock 함수값 추정문제 – *O_BP*(실험 3)) : 정밀도 5

입력	교사출력			O_BP(실험 3) 출력		
0000001 0000000	0000001	1	0.01	0.00002 0.00479 0.00217 0.00447 0.00040 0.00643 0.99962	1	0.01
0101011 0000100	0000011	3	0.02	0.00001 0.01088 0.00137 0.01369 0.00021 0.99258 0.98473	3	0.02
0101100 0011100	0000011	3	0.02	0.00000 0.00511 0.00023 0.00588 0.00044 0.99060 0.99749	3	0.02
1000101 0011001	0000001	1	0.01	0.00000 0.00810 0.00004 0.00652 0.00913 0.00534 0.98903	1	0.01
1011001 1111001	0101011	43	0.34	0.01197 0.98513 0.00479 0.98331 0.01215 0.99340 0.99996	43	0.34
0100010 0111000	0011101	29	0.23	0.01245 0.00374 0.98241 0.98637 0.99447 0.01559 1.00000	30	0.24
0001110 1011001	1100011	99	0.78	0.98536 0.99997 0.01477 0.01046 0.00022 0.98994 0.99291	98	0.77
1000111 0000101	0001111	15	0.12	0.00010 0.00945 0.01400 0.98684 0.98721 0.99054 0.99491	15	0.12
0010110 1100111	1111111	127	1.00	0.98015 0.99001 0.98355 0.99715 0.99658 0.98513 0.97926	125	0.98
1011000 1100001	0010001	17	0.13	0.00716 0.00816 0.98750 0.01026 0.00707 0.00636 0.99152	18	0.14
1101001 1111010	0001111	15	0.12	0.00003 0.01059 0.01126 0.99387 0.99030 0.98981 0.98807	15	0.12
0011100 0110111	0011110	30	0.24	0.01254 0.00792 0.97546 0.99384 0.97908 0.99993 0.01130	31	0.24
1111001 1101011	0000000	0	0.00	0.00001 0.00113 0.00372 0.00915 0.01236 0.00300 0.02010	0	0.00
1110101 1100111	0000000	0	0.00	0.00001 0.00015 0.00991 0.00752 0.01903 0.00759 0.01001	0	0.00
0111001 1001100	0100001	33	0.26	0.00068 0.99047 0.00017 0.00184 0.00038 0.00708 0.99992	33	0.26
1010100 1001100	0000101	5	0.04	0.00000 0.00441 0.00134 0.00779 0.98285 0.00002 0.99138	5	0.04
1000110 1011011	0100100	36	0.28	0.00622 0.98699 0.00743 0.00674 0.99162 0.01401 0.01257	36	0.28
0001110 0110100	0100010	34	0.27	0.01699 0.97457 0.00532 0.01567 0.01800 0.99999 0.02020	35	0.28
0001111 1010101	1011010	90	0.71	0.98122 0.02240 0.99032 0.98181 0.00230 0.99997 0.00863	89	0.70
0111100 0111110	0001111	15	0.12	0.00001 0.01347 0.01799 0.99301 0.99361 0.99821 0.99310	16	0.13

〈표 9〉 테스트 결과(Rosenbrock 함수값 추정문제 – *N_BP*(실험 1, 2, 3, 4)) : 정밀도 6

입력	교사출력			N_BP(실험 1, 2, 3, 4) 출력		
0000001 0000000	0000001	1	0.01	0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 1.000000	1	0.01
0101011 0000100	0000011	3	0.02	0.000000 0.000000 0.000000 0.000000 0.000000 1.000000 1.000000	3	0.02
0101100 0011100	0000011	3	0.02	0.000000 0.000000 0.000000 0.000000 0.000000 1.000000 1.000000	3	0.02
1000101 0011001	0000001	1	0.01	0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 1.000000	1	0.01
1011001 1111001	0101011	43	0.34	0.000000 1.000000 0.000000 1.000000 0.000000 1.000000 1.000000	43	0.34
0100010 0111000	0011101	29	0.23	0.000000 0.000000 1.000000 1.000000 1.000000 0.000000 1.000000	29	0.23
0001110 1011001	1100011	99	0.78	1.000000 1.000000 0.000000 0.000000 0.000000 1.000000 1.000000	99	0.78
1000111 0000101	0001111	15	0.12	0.000000 0.000000 0.000000 1.000000 1.000000 1.000000 1.000000	15	0.12
0010110 1100111	1111111	127	1.00	1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000	127	1.00
1011000 1100001	0010001	17	0.13	0.000000 0.000000 1.000000 0.000000 0.000000 0.000000 1.000000	17	0.13
1101001 1111010	0001111	15	0.12	0.000000 0.000000 0.000000 1.000000 1.000000 1.000000 1.000000	15	0.12
0011100 0110111	0011110	30	0.24	0.000000 0.000000 1.000000 1.000000 1.000000 1.000000 0.000000	30	0.24
1111001 1101011	0000000	0	0.00	0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000	0	0.00
1110101 1100111	0000000	0	0.00	0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000	0	0.00
0111001 1001100	0100001	33	0.26	0.000000 1.000000 0.000000 0.000000 0.000000 0.000000 1.000000	33	0.26
1010100 1001100	0000101	5	0.04	0.000000 0.000000 0.000000 0.000000 1.000000 0.000000 1.000000	5	0.04
1000110 1011011	0100100	36	0.28	0.000000 1.000000 0.000000 0.000000 1.000000 0.000000 0.000000	36	0.28
0001110 0110100	0100010	34	0.27	0.000000 1.000000 0.000000 0.000000 0.000000 1.000000 0.000000	34	0.27
0001111 1010101	1011010	90	0.71	1.000000 0.000000 1.000000 1.000000 0.000000 1.000000 0.000000	90	0.71
0111100 0111110	0001111	15	0.12	0.000000 0.000000 0.000000 1.000000 1.000000 1.000000 1.000000	15	0.12

〈표 9-1〉 테스트 결과(Rosenbrock 함수값 추정문제 $-N\_BP$(실험 1, 2, 3, 4)) : 정밀도 4

입력	교사출력			N_BP(실험 1, 2, 3, 4) 출력		
0000001 0000000	0000001	1	0.01	0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 1.0000	1	0.01
0101011 0000100	0000011	3	0.02	0.0000 0.0000 0.0000 0.0000 0.0000 1.0000 1.0000	3	0.02
0101100 0011100	0000011	3	0.02	0.0000 0.0000 0.0000 0.0000 0.0000 1.0000 1.0000	3	0.02
1000101 0011001	0000001	1	0.01	0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 1.0000	1	0.01
1011001 1111001	0101011	43	0.34	0.0000 1.0000 0.0000 1.0000 0.0000 1.0000 1.0000	43	0.34
0100010 0111000	0011101	29	0.23	0.0000 0.0000 1.0000 1.0000 1.0000 0.0000 1.0000	29	0.23
0001110 1011001	1100011	99	0.78	1.0000 1.0000 0.0000 0.0000 0.0000 1.0000 1.0000	99	0.78
1000111 0000101	0001111	15	0.12	0.0000 0.0000 0.0000 1.0000 1.0000 1.0000 1.0000	15	0.12
0010110 1100111	1111111	127	1.00	1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000	127	1.00
1011000 1100001	0010001	17	0.13	0.0000 0.0000 1.0000 0.0000 0.0000 0.0000 1.0000	17	0.13
1101001 1111010	0001111	15	0.12	0.0000 0.0000 0.0000 1.0000 1.0000 1.0000 1.0000	15	0.12
0011100 0110111	0011110	30	0.24	0.0000 0.0000 1.0000 1.0000 1.0000 1.0000 0.0000	30	0.24
1111001 1101011	0000000	0	0.00	0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000	0	0.00
1110101 1100111	0000000	0	0.00	0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000	0	0.00
0111001 1001100	0100001	33	0.26	0.0000 1.0000 0.0000 0.0000 0.0000 0.0000 1.0000	33	0.26
1010100 1001100	0000101	5	0.04	0.0000 0.0000 0.0000 0.0000 1.0000 0.0000 1.0000	5	0.04
1000110 1011011	0100100	36	0.28	0.0000 1.0000 0.0000 0.0000 1.0000 0.0000 0.0000	36	0.28
0001110 0110100	0100010	34	0.27	0.0000 1.0000 0.0000 0.0000 0.0000 1.0000 0.0000	34	0.27
0001111 1010101	1011010	90	0.71	1.0000 0.0000 1.0000 1.0000 0.0000 1.0000 0.0000	90	0.71
0111100 0111110	0001111	15	0.12	0.0000 0.0000 0.0000 1.0000 1.0000 1.0000 1.0000	15	0.12

〈표 9-2〉 테스트 결과(Rosenbrock 함수값 추정문제 $-N\_BP$(실험 1, 2, 3, 4)) : 정밀도 5

입력	교사출력			N_BP(실험 1, 2, 3, 4) 출력		
0000001 0000000	0000001	1	0.01	0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000	1	0.01
0101011 0000100	0000011	3	0.02	0.00000 0.00000 0.00000 0.00000 0.00000 1.00000 1.00000	3	0.02
0101100 0011100	0000011	3	0.02	0.00000 0.00000 0.00000 0.00000 0.00000 1.00000 1.00000	3	0.02
1000101 0011001	0000001	1	0.01	0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000	1	0.01
1011001 1111001	0101011	43	0.34	0.00000 1.00000 0.00000 1.00000 0.00000 1.00000 1.00000	43	0.34
0100010 0111000	0011101	29	0.23	0.00000 0.00000 1.00000 1.00000 1.00000 0.00000 1.00000	29	0.23
0001110 1011001	1100011	99	0.78	1.00000 1.00000 0.00000 0.00000 0.00000 1.00000 1.00000	99	0.78
1000111 0000101	0001111	15	0.12	0.00000 0.00000 0.00000 1.00000 1.00000 1.00000 1.00000	15	0.12
0010110 1100111	1111111	127	1.00	1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000	127	1.00
1011000 1100001	0010001	17	0.13	0.00000 0.00000 1.00000 0.00000 0.00000 0.00000 1.00000	17	0.13
1101001 1111010	0001111	15	0.12	0.00000 0.00000 0.00000 1.00000 1.00000 1.00000 1.00000	15	0.12
0011100 0110111	0011110	30	0.24	0.00000 0.00000 1.00000 1.00000 1.00000 1.00000 0.00000	30	0.24
1111001 1101011	0000000	0	0.00	0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000	0	0.00
1110101 1100111	0000000	0	0.00	0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000	0	0.00
0111001 1001100	0100001	33	0.26	0.00000 1.00000 0.00000 0.00000 0.00000 0.00000 1.00000	33	0.26
1010100 1001100	0000101	5	0.04	0.00000 0.00000 0.00000 0.00000 1.00000 0.00000 1.00000	5	0.04
1000110 1011011	0100100	36	0.28	0.00000 1.00000 0.00000 0.00000 1.00000 0.00000 0.00000	36	0.28
0001110 0110100	0100010	34	0.27	0.00000 1.00000 0.00000 0.00000 0.00000 1.00000 0.00000	34	0.27
0001111 1010101	1011010	90	0.71	1.00000 0.00000 1.00000 1.00000 0.00000 1.00000 0.00000	90	0.71
0111100 0111110	0001111	15	0.12	0.00000 0.00000 0.00000 1.00000 1.00000 1.00000 1.00000	15	0.12

參考文獻

1. P. Werbos, Beyond Regression : "New Tools for Prediction and Analysis in the Behavioral Science," PhD thesis, Havard, Cambridge, MA, August, 1974.
2. James McClelland and David Rumelhart, *Parallel Distributed Processing*, vol. .1 and 2, MIT Press, Cambridge, MA, 1986.
3. M. R. Hestenes and E. Stiefel, "Methods of Conjugate Gradient for Solving Linear Systems," J. Res. Nat. Bur. Standards, 49, pp. 409-436, 1952.
4. R. Fletcher and C. M. Reeves, "Function Minimization by Conjugate Gradients," *The Computer Journal* 7, pp. 149-153, 1964.
5. E. Polak and G. Ribiere, "Note Surla Convergence de Methodes Conjugees," *Revue Francaise Inform.* Rech. Operation 16-R1, pp. 35-43.
6. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," IEEE Proc. Pattern Analysis and Machine Intelligence, PAMI-6, pp. 721-741, 1984.
7. B. Gidas, "Global Optimization via the Langevin Equation," Proc. 24th Conf. on Decision and Control, Ft. Lauderdale, pp. 774-778, 1985.
8. James A. Freeman and David M. Skapura, *Neural Networks : Algorithms, Applications and Programming Techniques*, Addison Wesley, 1992.

Abstract

Enhanced Backpropagation : Algorithm and Numeric Examples

In this paper, we propose a new algorithm(*N_BP*) to be capable of overcoming limitations of the traditional backpropagation(*O_BP*). The *N_BP* is based on the method of conjugate gradients and calculates learning parameters through the line search which may be characterized by order statistics and golden section. Experimental results showed that the *N_BP* was definitely superior to the *O_BP* with and without a stochastic term in terms of accuracy and rate of convergence and might surmount the problem of local minima. Furthermore, they confirmed us that the stagnant phenomenon of learning in the *O_BP* resulted from the limitations of its algorithm in itself and that unessential approaches would never cured it of this phenomenon.