

다이나믹 학습율에 의한 다층 신경망의 학습 속도 개선

Improvement of learning speed of a multilayer neural network by using a dynamic learning rate

김 병호^{†,‡}, 오 상록[†], 서 일홍[‡]

† : 한국과학기술연구원 정보전자연구부 (Tel. (02)958-5748, Fax. (02)958-5709)

‡ : 한양대학교 전자공학과 (Tel. (0345)408-5802, Fax. (0345)408-5803)

Abstract A dynamic learning rate algorithm is proposed in this paper to improve the learning speed of the conventional error backpropagation algorithm of multilayer neural network. The method employs a statistical procedure to adjust the learning rate at each iteration, which is based on the variation of sum of absolute maximum error of network output. The simulation results illustrate its effectiveness in training a four-layered neural network for modelling the given system.

Keyword : *dynamic learning rate algorithm*

1. 서 론

시스템의 모델링, 패턴인식 및 제어 분야 등에 다양하게 응용되고 있는 다층신경망을 이용하여 어떤 시스템을 모델링함에 있어서 오차역전달 학습 알고리즘이 널리 적용되고 있다[1]. 이때 다층신경망의 학습속도와 수렴정도는 사용된 신경망의 층수와 각 층에 있는 뉴런의 갯수 및 학습율에 밀접하게 관련되어 있다. 그런데 대부분의 경우 기존의 오차역전달 학습 알고리즘에서는 신경망의 학습율을 임의로 선정하여 고정시켜 놓고 학습을 수행하는 경우가 많다. 이때 학습패턴에 따라 알맞는 학습율을 선정하기가 어려울 뿐만 아니라 정적인 학습이 이루어지게 되어 신경망의 학습율이 너무 크게 선정된 경우에는 지나친 학습효과로 인하여 신경망의 출력이 발산 하기도 하고, 불필요한 학습이 일어나기도 하여 불안정한 학습이 나타날 수 있고, 너무 작게 적용된 경우에는 신경망의 수렴 속도가 너무 늦게 되는 경향이 있다. 특히 복잡한 비선형 함수나 여러 개의 입력이 동시에 존재하는 학습 패턴 및 학습 패턴이 불규칙한 경우 등에 있어서 신경망의 학습율은 안정적이면서도 빠른 응답 특성을 얻는데 중요한 역할을 한다. 따라서 학습 패턴에 따라 신경망의

크기와 학습율을 보다 효과적으로 결정하는 것이 중요한 과제이며, 이러한 문제를 보다 효과적으로 해결하고 신경망의 학습속도를 향상시키기 위한 연구가 진행되고 있다[2, 3].

최근에 Hsin [2]은 학습이 진행됨에 따라 모든 연결강도의 방향 코사인을 참조하여 매 순간 학습율을 다르게 적용하는 방법을 제안하였는데, 이 방법은 과거 연결강도의 변화량을 근거로 현재 학습율을 결정하는 방법으로서 오차의 변화가 크면 클수록 큰 학습율이 구해지게 된다. 만약 오차가 증가하는 방향인 경우, 가능하면 학습율을 줄여가야 하는데, 오히려 큰 학습율이 적용되게 되어 신경망의 학습을 방해할 가능성을 내포하고 있다. 또한 연결강도의 변화량을 기억해야 하기 때문에 기억용량이 증가하는 문제점이 있고, Jin [3]의 알고리즘은 계산과정이 너무 복잡한 단점이 있다.

본 논문에서는 신경망이 갖게 되는 기억용량의 증가 없이, 단순하게 매 순간 신경망의 출력에 나타난 오차의 변화를 관찰하여 지속적으로 감소하는 경향이면 학습율을 증가시키고, 반대로 신경망의 출력 오차가 증가하는 경향이면 학습율을 감소시킬 필요성이 있다는 점에 착안하여 다층신경망의 초기 학습 결과를 근거

로 신경망의 초기 학습율을 구하고, 신경망의 학습이 진행됨에 따라ダイナミック하게 학습율을 결정하는 알고리즘을 제안하였다. 제안된 알고리즘의 효용성 입증 을 위해 비선형 함수 특성을 갖는 시스템 모델링에 대한 시뮬레이션 결과를 보이고, Hsin [2]의 결과와 비교 하였다.

2. 다이내믹 학습율에 의한 학습 알고리즘

본 장에서는 기존의 오차역전달 학습 알고리즘[1]에 적용할 학습율을 신경망의 학습상황에 따라 적절히 바꾸어 주는 “다이내믹 학습율 알고리즘”에 대해 기술 한다.

step 1) 학습패턴 p 에 대해 신경망의 출력층에서 오차 성분의 절대치를 계산한 후, 최대값을 찾아 내고, 이 값에 의해 신경망의 초기 학습율을 구한다.

$$\eta(t) = -\max\{|E_p(k)|\} \log_{10}(\beta) \quad (1)$$

$$E_p(k) = y_p(k) - o_p(k), \quad k = 1, 2, \dots, N \quad (2)$$

여기서, β 는 초기 학습율의 크기를 조정하기 위한 조정계수로서 $0 < \beta < 1$ 인 실수를 선택한 값이다. P 와 Q 에는 $Q \geq P > 0$ 인 정수, K 에는 $K > 0$ 인 실수를 선정하며, 이 값들은 step 3)과 step 4)에서 이용될 확률 변수이다.

step 2) 매 순간 각 학습패턴 p 에 대한 절대치 오차의 합과 그 변화율을 다음과 같이 계산한다.

$$S_e = \sum_k^N |E_p(k)| \quad (3)$$

$$\delta_E = S_e(t+1) - S_e(t), \quad t = 0, 1, \dots \quad (4)$$

step 3) 다음과 같은 두 경우가 만족되면

case 1 : $\delta_E \leq 0$ 인 경우가 연속적으로 P 번 계속 되는 경우,

case 2 : $\delta_E \leq 0$ 경우와 $\delta_E > 0$ 인 경우가 반복 되고, 그 합이 Q 번이며, $\delta_E > 0$ 인 경우에는 $0 < \xi < 1$ 에서 임의로 선택한 ξ 에 대해

$$\xi > \exp\left(-\frac{\eta(t)}{\delta_E}\right) \quad (5)$$

이 만족되는 경우,

$$\eta(t+1) = (1 + \alpha)\eta(t) \quad (6)$$

$$\alpha = K \log_{10}(1 + \delta_E) \quad (7)$$

에 의해 학습율을 증가시킨다.

step 4) 다음과 같은 두 경우가 만족되면

case 1 : $\delta_E \geq 0$ 인 경우가 연속적으로 P 번 계속 되는 경우,

case 2 : $\delta_E \geq 0$ 경우와 $\delta_E < 0$ 인 경우가 반복되고, 그 합이 Q 번이며, $\delta_E < 0$ 인 경우에는 step 3)에서 선택된 ξ 에 대해

$$\xi > \exp\left(\frac{\eta(t)}{\delta_E}\right) \quad (8)$$

이 만족되는 경우,

$$\eta(t+1) = (1 - \alpha)\eta(t) \quad (9)$$

에 의해 학습율을 감소시킨다. 여기서, α 는 식(7)과 같다.

step 5) step 3), step 4)의 조건이 모두 만족되지 않으면, 학습율은 갱신하지 않고, 이전의 학습율을 그대로 사용하여 학습하게 되며, 모든 학습 입력에 대한 학습이 끝나면, step 2)번 과정으로 되돌아간다.

[Remark 1.] : P 와 Q 는 현재 신경망의 출력오차가 감소하거나 증가하는 경향을 파악하는 확률변수로서, 학습율을 증감시키는 빈도를 조절할 수 있다. 특히 대상 시스템이 진동하는 특성을 가진 경우나 시스템 특성을 잘 모르는 경우에는 비교적 큰 값을 선택할수록 신중하게 학습을 유도하게 되어 안정된 학습효과를 얻을 수 있게 된다.

[Remark 2.] : 식(1)은 신경망의 초기 학습율을 최대 오차의 로그(\log_{10}) 눈금자에 의해, 스케일링(scaling)하여 결정함을 의미한다. 여기서, β 를 1에 가깝게 정할수록 초기 학습율이 작게 되어 미지의 시스템인 경우에도 안정된 학습을 수행할 수 있게 된다. 또한 시

스텝 정보를 어느 정도 알고 있는 경우에는 적절히 작게 선택하여 초기 학습을 촉진시킬 수 있다는 의미를 포함하고 있다.

[Remark 3.] : 식(5)과 식(8)에서 δ_E 는 모든 학습패턴에 대한 오차의 변화율을 나타내는데, 오차가 비록 증가하거나 감소하더라도 학습패턴에 따라 신경망의 전체적인 학습에 있어서는 오히려 학습이 진전되고 있거나 방해받고 있는 경우가 있기 때문에 확률을 이용해 현재의 학습율에 대한 신경망의 오차의 변화율이 작거나 큰 경향으로 인식하여 학습율을 크게 하거나 작게 하겠다는 의미를 갖는다.

[Remark 4.] : 식(7) 에서 K 는 로그함수의 기울기를 조정하게 되어 학습율 증감 비율을 조절한다.

3. 시뮬레이션 결과 및 검토

여기서는 제안한 동적 학습율 알고리즘을 이용한 시뮬레이션 결과를 보이고, 비교적 접근 방법이 유사한 Hsin [2]의 결과와 비교하였다.

시뮬레이션에 사용된 시스템 모델은 다음과 같다[4].

$$y_p(k+1) = 0.6y_p(k) + \sin(2\pi k/25) + \sin(2\pi k/10), \quad 0 \leq k \leq 25 \quad (10)$$

여기서, $y_p(0) = 0$ 이다.

시스템 모델링을 위한 다층신경망은 4 층 신경망으로서 입력층은 입력 뉴런 1 개와 바이어스(bias) 뉴런 1 개, 첫번째 은닉층의 뉴런은 20 개, 두번째 은닉층의 뉴런은 20 개, 출력층의 뉴런의 갯수는 1 개로 각각 구성하였고, 바이어스 뉴런은 모든 은닉층의 뉴런에 연결되어 있다. 입력층은 입력을 그대로 내보내고, 은닉층은 시그모이드(sigmoid) 함수를 사용하였으며, 출력층은 은닉층을 통해 들어온 입력의 합을 계산한다. 이때 바이어스는 1.0, 시그모이드 함수의 기울기는 1.0 으로 초기화 하였다.

Hsin [2]의 알고리즘과 제안한 알고리즘을 적용하여 시스템 모델링한 결과와 학습속도에 대한 성능은 각각 그림 1 과 그림 2 에서 보는 바와 같이 제안한 알고리즘에 의한 모델링 특성이 한층 더 정밀함을 확인할

수 있고, 제안한 알고리즘에 의한 학습오차는 12000 번 학습이 진행된 후 부터, 안정적으로 수렴해 가는데 비해 Hsin 의 알고리즘을 이용한 경우에는 계속 진동하는 것을 볼 수 있다. 이는 학습율을 연결강도의 방향코사인에 의해 구하기 때문에 학습율 변동이 너무 지나치게 이루어지는 결과이며, 그림 3 에 나타낸 학습율 변화 과정을 통해 알 수 있다.

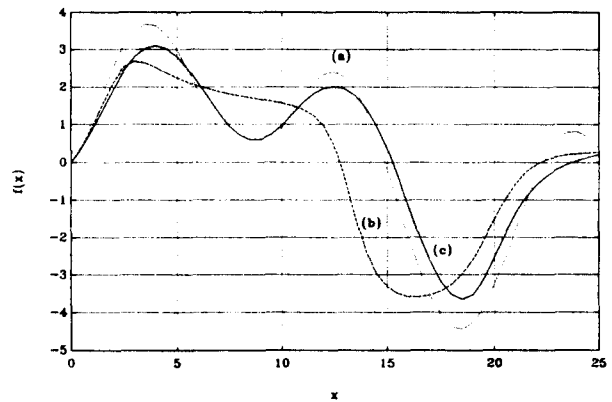


그림 1. 신경망의 출력

- (a) 주어진 시스템 (b) Hsin 의 알고리즘
- (c) 제안한 알고리즘

Fig. 1. The output of the network

- (a) Given system (b) Hsin's algorithm($L=1$)

- (c) The proposed algorithm

($\beta = 0.995, P=15, Q=20, K=0.01$)

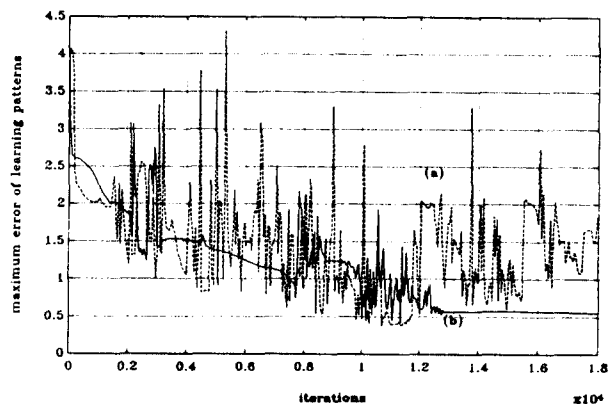


그림 2. 신경망의 학습오차

- (a) Hsin 의 알고리즘 (b) 제안한 알고리즘

Fig. 2. The behavior of learning error

- (a) Hsin's algorithm (b) The proposed algorithm

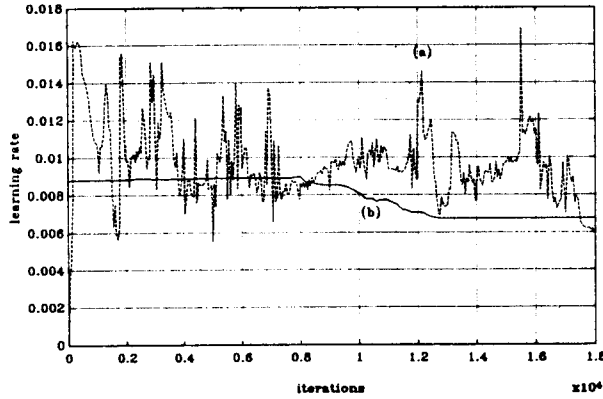


그림 3. 학습율 변화 과정

(a) Hsin의 알고리즘 (b) 제안한 알고리즘

Fig. 3. The behavior of learning rate

(a) Hsin's algorithm (b) The proposed algorithm

4. 결론

본 논문에서는 다층 신경망의 학습 속도를 한층 더 개선하기 위하여ダイナミック하게 학습율을 결정하는 새로운 알고리즘을 제안하였다. 제안된 알고리즘은 모든 학습패턴에 대한 신경망의 출력오차의 합을 근거로 그 변화율에 따라 신경망의 학습율을 조절하기 때문에 학습패턴이나 신경망의 구조에 관계없이 적용할 수 있을 뿐만 아니라 학습율 계산과정이 단순하여 비선형 함수나 여러개의 입력이 동시에 존재하는 학습 패턴 등의 학습에 적용이 용이하며, 한층 더 학습속도를 향상시킬 수 있다.

한편 확률을 결정하는 각 변수들의 효율적인 선정에 관한 연구가 추가적으로 필요하다.

References

- [1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing*, vol. 1, pp. 318-362, MIT press, 1986.
- [2] H.C. Hsin, C.C. Li, and R.J. Scabassi, "An Adaptive Training Algorithm for Back-Propagation Neu-

ral Networks," *IEEE Trans. on Systems, Man, and Cybernetics*, pp. 512-514, March 1995.

- [3] L. Jin, P.N. Nikiforuk, and M.M. Gupta, "Fast Neural Learning and Control of Discrete-Time Nonlinear Systems," *IEEE Trans. on Systems, Man, and Cybernetics*, pp. 478-488, March 1995.
- [4] K. S. Narendra, and K. Parthasarathy, "Gradient Methods for the Optimization of Dynamical Systems Containing Neural Networks," *IEEE Trans. on Neural Networks*, Vol. 2, No. 2, pp. 252-262, 1991.