

## 퍼지 모델을 이용한 신경망의 학습률 조정

라혁주, 서재용\*, 김성주, 전홍태  
중앙대학교 일반대학원 전자전기공학부  
전화 : 02-820-5297  
\*한국기술교육대학교 정보기술공학부

### Tuning Learning Rate in Neural Network Using Fuzzy Model

Hyuk-Ju Ra, Jae-Yong Seo\*, Sung-Joo Kim, Hong-Tae Jeon

School of Electrical and Electronic Engineering, Chung-Ang Univ.

\*School of Information and Technology Engineering, Korea Univ. of Tech. and Edu.

E-mail : rahyukju@wm.cau.ac.kr

#### Abstract

The neural networks are a famous model to learn the nonlinear function or nonlinear system. The main point of neural network is that the difference actual output from desired output is used to update weights. Usually, the gradient descent method is used for the learning process. On training process, if learning rate is too large, neural networks hardly guarantee convergence of neural networks. On the other hand, if learning rate is too small, the training spends much time. Therefore, one major problem in use of neural networks are to decrease the learning time while neural networks are guaranteed convergence.

In this paper, we suggest the model of fuzzy logic to neural networks to calibrate learning rate. This method is to tune learning rate dynamically according to error and demonstrates the optimization of training.

#### I. 서론

신경망은 병행적으로 상호 작용하는 여러 개의 계산 요소들로 이루어져 있으며, 각 계산요소는 가중치 합(weighted sum)과 같은 단순한 계산만을 수행한다. 신경망의 계산능력은 다량의 계산요소들이 병행적으로 계산을 수행함으로써 얻어진다. 신경망은 학습데이터를 통하여 가중치를 학습하고, 그 데이터에 대한 특성을

일반화시킴으로서 학습 기능을 수행한다[1]. 이러한 신경망은 퍼셉트론 모델에 은닉층(hidden layer)과 역전파(back-propagation)알고리즘이 적용되면서 더욱 발전하게 되었다.

은닉층은 기존의 퍼셉트론 모델이 갖는 입력과 출력의 단일 조정층사이에 한 개 혹은 그 이상의 층을 첨가함으로써 복잡한 문제에 대한 근사해를 비교적 신속하게 찾아낼 수 있는 구조를 형성하게 되었다. 또한 역전파 알고리즘은 다층 신경망의 학습이 가능하게 되고, 여러 가지 응용 과제로의 적용이 시도되고 있다. 그렇지만 실용적인 규모의 네트워크 학습에도 매우 시간이 걸린다는 문제와 학습의 과정 중에 네트워크의 연결강도가 오차에 대해 어떻게 변화하는지 알 수 없다는 문제들이 해결 과제로 남아있다.

이와 같은 신경망의 문제를 해결하는데 있어 본 논문에서는 학습률의 능동적 변화에 초점을 맞추고자 한다. 학습률(Learning rate)은 그 크기에 따라 1회의 시냅스 하중의 수정량을 변화시키는 정도를 결정하는 역할을 한다. 학습률이 클 경우에는 빠른 수렴을 이끌어 낼 수 있으나, 불확실한 학습 결과를 갖게 할 수도 있다 [1]. 학습과정에 있어 오차발생과 연계된 학습률의 능동적 변화는 수렴성과 최적 학습의 상쇄관계(trade off)를 해결할 수 있을 것이다.

본 논문에서는 학습률의 능동적 변화를 이끌어 내기 위해서 Sugeno 퍼지 모델을 사용하였다. 이 모델은 수학적으로 다루기가 비교적 쉽고 추론에 의해 도출된 결론은 시스템에 바로 적용이 가능하다는 장점이 있다.

## II. 신경망 이론

신경망은 기존의 방식과는 달리 분산 저장 방식을 갖고, 뛰어난 학습 능력, 일반성, 오류 허용(fault-tolerance)과 같은 특성을 갖는다.

### 2.1 전 방향 네트워크

가장 널리 쓰이는 다층신경망 구조는 그림 1에 보이는 바와 같다. 이 전 방향 네트워크의 뉴런의 출력은 식(1)과 같이 표시할 수 있다.

$$v_k(n) = \sum_{j=0} w_{kj}(n) y_j(n) \quad (1)$$

$$y_k(n) = f(v_k(n)) \quad (2)$$

여기서  $v_k(n)$ 은  $j$ 번째 뉴런에 출력에 연결강도가 곱해서 모두 더한 값이다. 이 값이 활성화함수의 입력으로 들어간다.  $y_k(n)$ 은  $k$ 번째 뉴런의 출력이다.

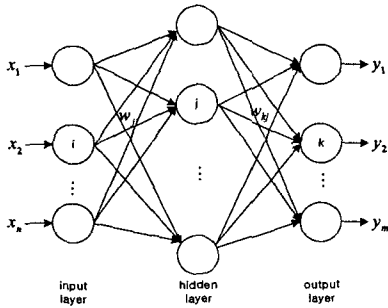


그림 1 다층 신경망 모델

### 2.2 오차 역전파 알고리즘

역전파 학습 방법은 가장 널리 이용되는 감독학습(supervised learning)으로 동일한 입력  $x_j(n)$ 에 대해서 현재 연결강도에 의한 신경망 출력  $\hat{y}_k(n)$ 이 비선형 함수 출력  $y_k(n)$ 을 근사화하도록 한다. 이것은 에러  $e_k(n)$ 가 신경망에 대해 역전파되어 연결강도에 따른 에러 기울기가 계산되고 신경망의 연결강도가 조정된다

그림 1의 다층 신경망의 학습은 오차 역전파 학습알고리즘에 의해 진행되며 출력단에서 전체 오차는 각 출력층 뉴런의 출력과 목표치와의 차이들의 합으로 다음과 같이 구할 수 있다.

$$E = \frac{1}{2} \cdot \sum_k (d_k - y_k)^2 \quad (3)$$

여기서,  $d_k$ 는 출력단  $k$ 번째 뉴런의 목표 값이며,  $y_k$ 는 출력단  $k$ 번째 뉴런의 실제 출력 값을 나타낸다.

연결강도들의 변화량은 Gradient descent 방법에 의해

다음과 같이 구할 수 있다.

$$\begin{aligned} \Delta w_{kj} &= -\eta \cdot \frac{\partial E}{\partial w_{kj}} \\ &= -\eta \cdot \frac{\partial E}{\partial v_k} \cdot y_j \\ &= \eta \cdot \delta_k \cdot y_j \end{aligned} \quad (4)$$

$$\delta_k = (d_k - y_k) \cdot f'(v_k) \quad (5)$$

여기에서,

$$f'(v_k) = \frac{\partial f(v_k)}{\partial v_k} \quad (6)$$

이다.

또한  $\eta$ 는 학습률을 나타내고  $\delta_k$ 는 역방향으로부터 전달되어 오는 오차이다. 위 식에 의한 출력단에서의 연결강도 변화와는 달리 중간층에서의 연결강도 변화량  $\Delta W_{ji}$ 는 다음 식에 의해 구할 수 있다.

$$\begin{aligned} \Delta w_{ji} &= -\eta \cdot \frac{\partial E}{\partial w_{ji}} \\ &= \eta \cdot \delta_j \cdot y_i \end{aligned} \quad (7)$$

$$\delta_j = f'(v_k) \cdot \sum_k (\delta_k \cdot w_{kj}) \quad (8)$$

각 층에서의 새로운 연결강도들은 최종적으로 다음 식에 의해 조정된다.

$$w_{kj}(t+1) = w_{kj}(t) + \Delta w_{kj} \quad (9)$$

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji} \quad (10)$$

여기서, 위의 식을 출력단에서 입력 단까지 계속적으로 반복하면 각 뉴런의 출력 오차가 감소하도록 연결강도가 변하게 되며, 마지막에는 오차가 없게 되어 학습이 완료된다.[4]

### 2.3 학습률에 따른 수렴조건

학습률은 앞서 설명한 바와 같이 1회의 시냅스 하중의 수정량을 결정하는 중요요소가 된다. 학습률이 클 경우 시냅스의 하중에 큰 변화를 유도하게 되며, 반대로 작은 값일 경우에는 수렴속도를 현저하게 감소시키게 되어 학습시간이 길어진다. 학습시간과 최적학습은 서로 맞물려 있어서 상쇄관계를 형성한다. 그러므로 신경망의 적절한 학습률은 최적학습을 보장해야하며 수렴시간을 단축 할 수 있어야 하는 것이다[4].

본 논문에서는 신경망의 학습률이 갖고 있는 중요성을 고려하여 연결강도가 갱신될 때마다 이상적 출력과 실제 출력의 차이를 입력으로 받아 새로운 학습률을 적용하게 하였다.

## III. Sugeno 퍼지 모델

퍼지 기술이란 그 소속정도가 불확실하거나 불분명한 입력에 대하여 결론을 도출해 내는 추론기술을 의미한

다. 퍼지 제어기술은 인간의 언어적 형식을 기반으로 하여 규칙과 조건에 의한 합리적 결론 도출을 유도한다 [3]. 이러한 퍼지 모델들은 정확한 결과를 내는 것보다는 입력의 변화가 크고 빠른 대응성을 필요로 하는 경우에 적합하다.

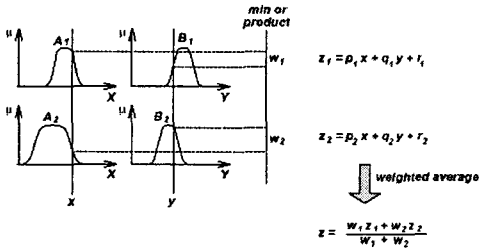


그림 2 Sugeno 퍼지 모델[5]

Sugeno 퍼지 모델은 주어진 입력이 조건부에 의해 퍼지화되고, 그 지수가 결론부의 함수에 입력되어 최종 결과로 출력되는 방식이다. 이 모델의 가장 큰 장점은 앞서 설명한 바와 같이 비퍼지화 과정이 필요 없이 바로 시스템에 적용이 가능한 형식으로 출력된다는 점이다. 결론부의 함수형식은 주어진 입력과 파라미터에 의한 선형 일차 함수의 형식으로 나타난다.

## IV. 실험 및 결과

### 4.1 모델 구현

본 논문에서 구현된 Sugeno 퍼지 모델에 의한 학습률의 최적적용 방식은 오차 역전파 방식의 신경망을 골격으로 하였다.

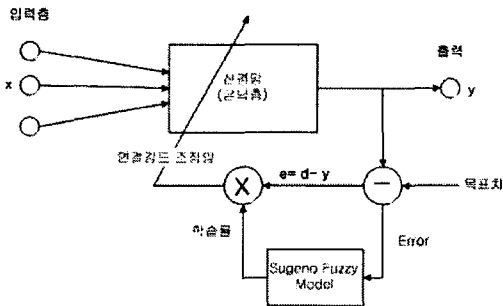


그림 3 구현된 신경망 모델의 개념도

Sugeno 퍼지 모델의 입력은 출력과 목표값과의 차이이며, 출력은 최적화된 학습률이 되어 신경망의 연결강도를 조절하게 된다.

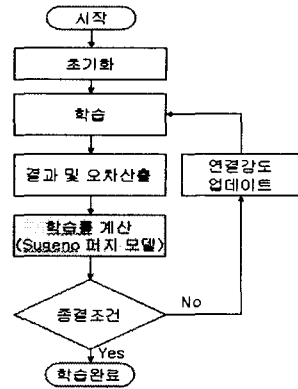


그림 4 구현된 모델의 처리과정

신경망의 학습과정에서 발생하는 오차는 Sugeno 퍼지 모델의 입력이 되어 새로운 학습률을 결과로 나타낸다. 오차의 특성상 절대값을 취하여 Sugeno 퍼지 모델의 소속 함수의 수를 줄이고, 계산량을 감소시켰다. 또한 학습률의 기본적인 범위를 0.01~0.5의 범위로 한정하여 연결강도가 돌발적으로 변화하는 것을 제한하였다.

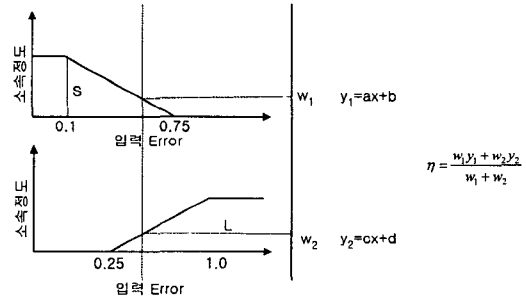


그림 5 Sugeno 퍼지 모델

### 4.2 모의실험

1차원 함수와 2차원 함수에 대하여 근사화 실험을 수행하였으며 비교 대상으로는 기존의 다층 전방향 신경 회로망을 사용하였다[7]. 근사 대상 함수  $f(x)$ 는 다음과 같다.

함 수	구간[-10, 10]
$-2.186x - 12.864$	$[-10, -2)$
$4.246x$	$[-2, 0)$
$10e^{-0.05x - 0.5} \cdot \sin[(0.03x + 0.7)x]$	$[0, 10]$

비교를 위한 기존 다층 전방향 신경 회로망은 학습률  $\eta$ 는 0.05를 사용하였고, 총 1000번의 반복 학습을 수행하였다.

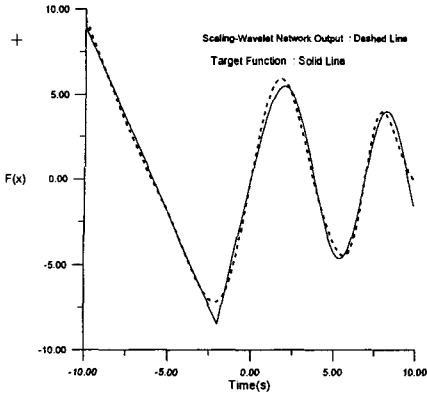


그림 6 제안 모델의 함수 근사화 결과

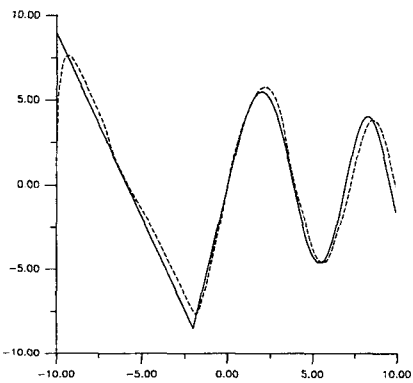


그림 7 기존 전방향 신경 회로망

전반적으로 두 모델들의 근사화 결과는 상당히 유사하였지만, 아래의 RMSE 값을 비교해보면 초기 수렴속도의 차이를 알 수 있다.

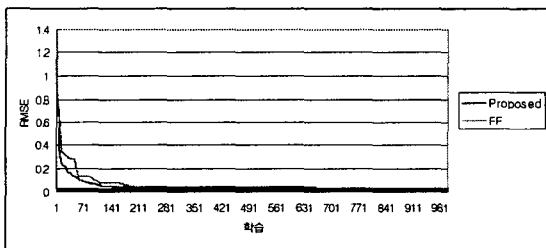


그림 8 두 모델의 RMSE

### V. 결론 및 향후 과제

본 논문에서는 Sugeno 퍼지 모델을 사용하여 신경망의 학습과정에서 발생하는 오차의 변화를 능동적으로 수렴하여 학습률의 변화를 주는 모델을 제시하였다. 앞서 실행한 모의실험의 결과에서처럼 초기 수렴속도의 향상이 입증되었다(그림 8). 그러나 학습 횟수가 증가함에 따라 기존 신경망과 같은 RMSE값으로 수렴하는 것이 확인 되었다. 이는 일정한 학습 횟수가 지남에 따라 오차가 충분히 작기 때문에 제안한 모델에서 적용되는 학습률  $\eta$ 의 차이가 기존 전방향 신경망의 학습률  $\eta$ 와 차이가 크지 않았기 때문으로 사료된다.

감사의 글 : 본 논문은 과학기술부 뇌신경정보학연구사업에 의해 지원받았습니다.

### VI. 참고문헌

- [1]일본계측자동제어학회, 뉴로퍼지인공지능 핸드북, 도서출판 기다리1998. pp.105-108
- [2]오창석, 뉴로컴퓨터, 내하출판사, 2000. pp.149-153, 210-218
- [3]변중남, 퍼지논리 제어, 홍릉과학 출판사, 1997. pp.2-3
- [4]Simon Haykin, Neural Networks, 2nd edition, Prentice Hall, 1999. pp.1-6, 161-173
- [5]J.-S.R.Jang, C.-T.Sun, E.Mizutani, Neuro-Fuzzy and Soft Computing, Prentice Hall, 1997. pp.81-84, 243-238,
- [6]이광열,오길록, 퍼지이론 및 응용, 홍릉과학 출판사, 1997.
- [7]김성주, 서재용, 조현찬, 김성현, 전홍태, 직교 기저 함수 기반의 혼합 신경회로망 구조,
- [8]Q. Zhang and A. Benveniste, "Wavelet Works," IEEE Trans. Neural Networks, vol. 3, pp. 889-898, Nov. 1992.
- [9] Jaideva C. Goswami and Andrew K. Chan, Fundamentals of Wavelets, Wiley Interscience, 1999.