

퍼지 추론에 의한 신경회로망 활성화 함수 형상의 자동 조정

김 인*, 정경권*, 김주용*, 서원호*, 손동설**, 엄기환*
*동국대학교 전자공학과, **유한대 전자과
TEL:02-2260-3332/FAX:2279-1798

Auto-Tuning of Activation Function in Neural Networks Using Fuzzy Inference

Gim Ine*, Jung Kyung-kwon*, Kim Joo-Woong*, Seo Won-Ho*, Son Dong-Seol**, Eom Ki-Hwan*
*Dept. of Electronics Engineering, Dongguk Univ, **Dept. of Electronics, Yuhan College
E-mail: igotome@dgu.ac.kr

Abstract

In this paper, we propose the auto-tuning method for the gradient of activation function using fuzzy inference. In order to verify the usefulness of the proposed method, we simulated it with one link manipulator system and confirmed the excellency.

I. 서 론

다층 신경 회로망은 단순한 기능을 가진 무수히 많은 뉴런 또는 처리 소자들이 병렬 연결된 연산 구조로 학습이나 훈련을 통해 연결 강도를 조정함으로써 유연성이나 적응 능력이 뛰어나다. 이러한 능력은 다층 신경 회로망의 학습 기구에 의해 실현되고 있으며 일반적으로 오차 역

전과 알고리즘을 많이 이용하고 있다. 오차 역전과 알고리즘은 목표 출력과 네트워크 실제 출력 사이의 차의 자승합을 목적함수로 하여 그 목적 함수의 최소점을 최급강하법으로 구한다. 이러한 오차 역전과 알고리즘의 문제점은 학습 소요 시간이 길고, 지역 최소점에 빠질 가능성이 있으며, 적절한 중간층의 층수와

뉴런수를 결정하는 것이다. 이러한 문제점 중에서 학습 소요 시간의 단축은 다층 신경 회로망 학습 전반에 걸쳐 메리트가 가장 크다고 생각된다. 오류 역전과 알고리즘의 중요한 학습 요소들은 초기 연결 강도, 누적 연결 강도 조정과 증분 갱신, 활성화 함수의 기울기, 학습률 등이 있는데, 이러한 요소들은 학습 속도와 수렴성 등 신경망의 학습 능력에 영향을 미친다. 그러나 이러한 요소들은 기존에 임의적이고 경험적인 조작에 의존해 왔다. 그 중에서도 상대적으로 고정된 형태로 사용되어온 활성화 함수(activation function)의 경우, 역전과 알고리즘을 이용하는 다층 신경회로망의 은닉층 활성화 함수로서 비선형 함수이며 미분 가능하고 미분식이 간단하다는 등의 장점으로 인해 시그모이드(sigmoid) 함수가 널리 사용되고 있다.[1][2]

본 논문에서는 오차 역전과 알고리즘의 학습 소요 시간 단축을 위해 신경 회로망 활성화 함수 형상을 퍼지 추론을 이용하여 자동조정하는 방법을 제안한다. 활성화 함수로는 일반적으로 많이 사용되고 있는 시그모이드 함수를 사용하였으며, 오차 함수의 구배와 시그모이드 함수의 기울기의 관계를 조사한 후 퍼지 추론을 이용하여 시그모이드 함수의 기울기를 자동 조절한다. 단일 관절 매니퓰레이터를 제어 대상으로 기존의 신경

회로망, 일반적인 퍼지 제어 방식과의 비교를 통해 제안한 방식의 유용성을 검토한다.

II. 오차 함수의 형상이 학습에 미치는 영향

비선형 활성화 함수 $\psi[\cdot]$ 는 합성값 $net \in [-\infty, \infty]$ 를 신경 출력으로 사상한다. 일반적으로 신경 출력은 단극(unipolar) 신호 $[0, 1]$ 의 범위와 양극(bipolar) 신호 $[-1, 1]$ 의 범위에 있다. 비선형 활성화 함수 연산자는 집합 net 을 신경 출력의 범위 0로 변환한다.

$$O = \psi[net] = \psi[\sum X_i \cdot W_i] \quad (1)$$

단, X_i 는 신경 입력벡터, W_i 는 시냅스 연결 강도이고 합성 연산 \cdot 는 둘 사이의 유사점 측정을 제공한다.

뉴런의 연속 활성화함수는 기울기 λ 에 의해 특성이 결정되고 미분치 $\psi'(net)$ 은 오차 신호 δ 를 만드는 구성요소이므로 이 두 요소는 신경망의 학습 속도에 영향을 미친다.[1]

$$\psi[net] = \frac{2}{1 + \exp(-\lambda net)} - 1 \quad (2)$$

$$\psi'[net] = \frac{2\lambda \exp(-\lambda net)}{[1 + \exp(-\lambda net)]^2} \quad (3)$$

$$hidnet_k = \sum_i x_i \cdot v_{ik} \quad (4)$$

$$hidout_k = \psi(hidnet_k) \quad (5)$$

$$outnet_m = \sum_k hidout_k \cdot w_{km} \quad (6)$$

$$Out_m = \psi(outnet_m) \quad (7)$$

그림 1은 다층 신경망의 일반적인 구조이다.

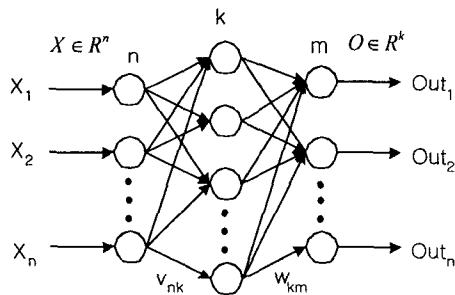


그림 1 다층 신경망의 일반적인 구조
Fig 1. Multilayerd neural network

역전파 알고리즘은 모든 층에서 일반화된 Delta Rule

을 적용한다.

$$\Delta W_{km}(n) = \alpha \Delta W_{km}(n-1) + \eta \delta k(n) Out_m(n) \quad (8)$$

단, α 는 momentum constant, η 은 학습률이다.

$$\delta_m(n) = e(n) \psi'(outnet_m(n)) \quad (9)$$

단, δ 는 제곱근 에러의 합이다.

그림 2에서 알 수 있 듯 λ 는 활성화 함수의 기울기를 결정하는 파라미터로서 λ 가 무한대로 발산할 때, 연속적인 함수는 이산 함수의 모양에 가까워진다.

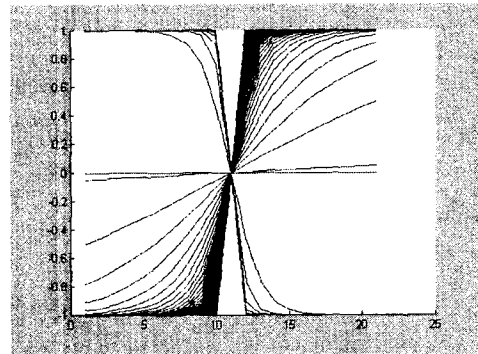


그림 2 λ 의 변화에 따른 함수 형상
fig 2. Shape of function about λ

λ 의 변화에 따른 미분치의 결과는 그림 3과 같다. 그림에서 알 수 있듯이, 미분치는 λ 값이 일정할 때 중간 대역의 값이 가장 크고 입력값이 일정할 때 λ 값이 클수록 커진다.

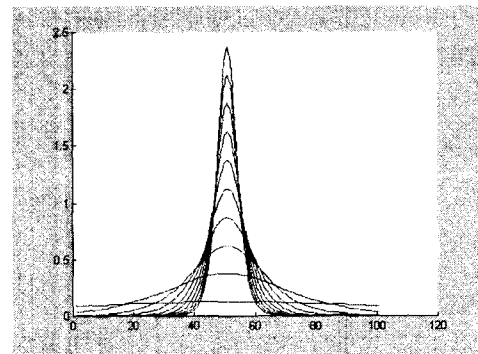


그림 3 활성화 함수의 기울기
fig 3. Gradient of activation function

III. 퍼지 추론에 의한 활성화 함수 형상 조정

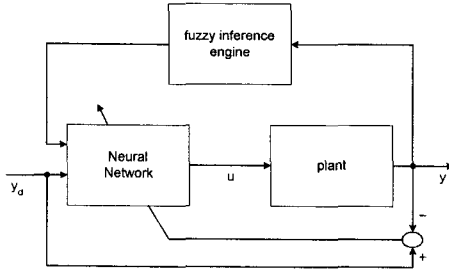


그림 4 제안한 방식
fig 4 . Proposed method

그림 4는 퍼지 추론에 의한 신경회로망의 활성화 함수를 자동 조정하는 신경 회로망을 제어기로 사용한 제안한 제어 방식의 전체적인 블록 선도이다.

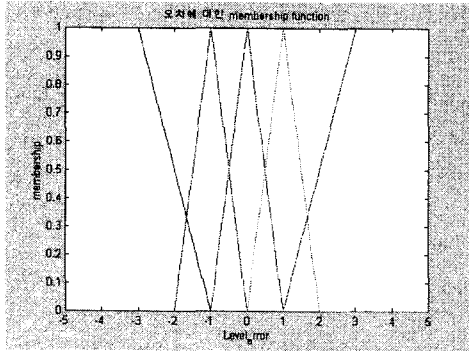


그림 5 오차에 대한 멤버쉽 함수
fig 5. Membership function about error

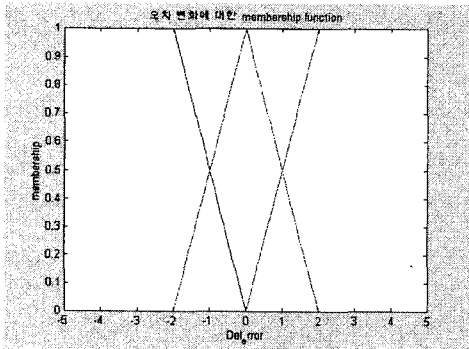


그림 6 오차 변화율에 대한 멤버쉽 함수
fig6. Membership function about error change

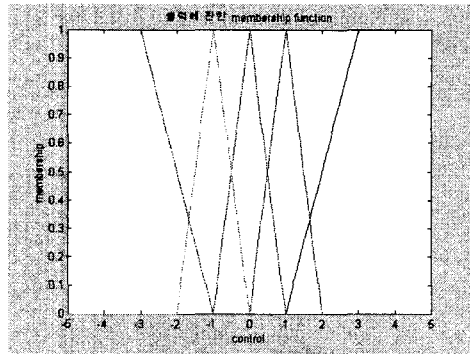


그림 7 출력에 대한 멤버쉽 함수
fig 7. Membership function about output

IV. 시뮬레이션

제어 대상인 단일 관절 매니퓰레이터는 식(10)과 같다.

$$m l^2 \ddot{\theta} + B \dot{\theta} + m g l \cos(\theta) = \tau \quad (10)$$

단, $\theta, \dot{\theta}, \ddot{\theta}$ 는 각각 매니퓰레이터의 각변위, 각속도, 각가속도, τ 는 매니퓰레이터의 입력으로 관절에 가해지는 토크, m, l, B, g 는 각각 매니퓰레이터의 질량, 길이, 마찰 계수, 중력 가속도.

$$m=0.1[\text{kg}], l=0.5[\text{m}], g=9.8[\text{m/s}^2],$$

$$B=0.4[\text{kg} \cdot \text{m}^2/\text{s}]$$

식(10)의 시스템을 상태 공간 모델로 표현하면 식(11)와 같다.

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = \frac{1}{m l^2} \tau - \frac{B}{m l^2} x_2 - \frac{g}{l} \cos x_1 \quad (11)$$

식(11)에 대해 Runge-Kutta fourth 알고리즘을 이용하였고, Sampling time은 10[ms]이며 목표 함수는 식(12)와 같다.

$$y_d(t) = 4 \sin(t) \quad (12)$$

일반적인 신경망의 경우와 제안한 방식의 경우, 신경망의 구조는 입력층 4개, 은닉층 30개, 출력층 1개로 구성되었고, 초기 조건은 동일하게 구성하였으며, 각각 학습률은 0.05, 0.5이다.

연결 강도와 바이어스는 [0, 1] 사이에서 랜덤하게 결정하였다.

일반적인 퍼지 제어의 경우 퍼지 추론의 출력이 플랜트의 입력으로 들어가는 기존의 퍼지 방식을 사용하였다. 또한 세 경우 모두 최적화된 결과이다.

그림 8은 일반적인 신경망을 이용한 제어 결과이며, 그림 9은 일반적인 퍼지 방식을 이용한 결과이고, 그림 10은 제안한 방식의 결과이다.

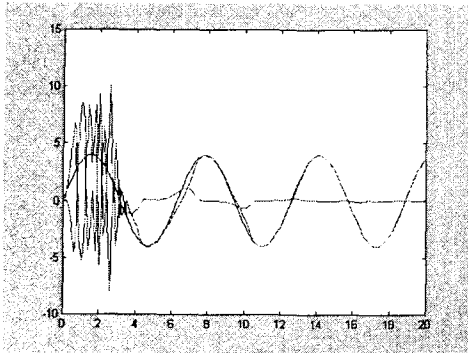


그림 8 일반적인 신경망의 기준계적과 출력
fig 8. Response of general NN

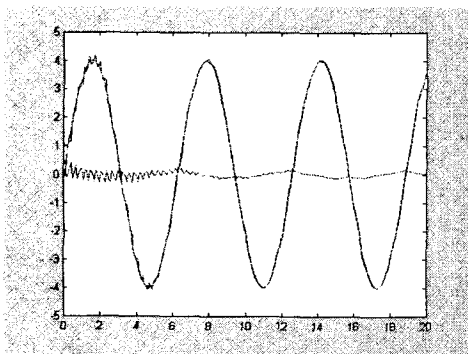


그림 9 일반적인 퍼지 제어의 경우
fig 9. General fuzzy control

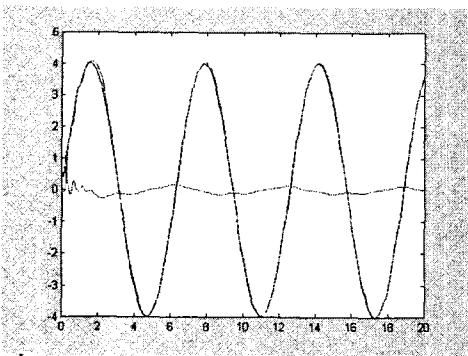


그림 10 제안한 방식의 결과
fig 10. Output of proposed method

V. 결 론

본 논문에서는 다층 신경망의 학습 소요 시간의 단축과 보다 좋은 수렴 특성을 얻기 위해서 은닉층 활성화 함수의 기울기를 자동 조정하는 방식을 제안하였다. 제안한 방식은 활성화 함수의 기울기 파라미터인 λ 를 퍼지 추론에 의해 조정함으로써 함수

형상을 변화시켜 신경망을 최적화시키는 방법이다.

제안한 방식을 단일 관절 매니플레이터에 적용하고 기존의 신경망과 퍼지를 사용하는 제어 방식과 비교해 본 결과, 초기 오차와 오버 슈트가 적고, 빠른 오차 수렴을 확인하였다.

참고 문헌

- [1] 이상배, "퍼지-뉴로제어 시스템", (주) 교학사, 1999
- [2] Chun-Sup Kim, "A Study on the Learning Speed Acceleration of the Fuzzy Neural Network of Back-propagation", IEEE Press, III-499~III-502, 1995
- [3] Tokumitsu Fujita, "A Study on improvement in Learning Efficiency of Multilayered Neural Networks based on Dynamic system", T.IEE japan, Vol.117-C, No.12,1848-1855,1997
- [4] Kevin M. Passino, "Fuzzy Control", Addison-Wesley, 1998