다층 신경회로망 모델 Topology의 최적 구성

이인재* · 정성부** · 임중규* · 이현관* · 정지원* 엄기환*

*동국대학교 전자공학과 **서일대학 전자공학과

The Optimal Construction of Multilayer Neural Network Model Topology

In-jae Lee* · Sung-boo Chung** · Joong-gyu Lim* · Hyun-kwan Lee*

· Jee-won Joeng* · Ki-hwan Eom*

*Dept. of Electronic Eng., Dongguk University · **Dept. fo Electronic Eng., Seoil College

E-mail : baram@cakra.dongguk.ac.kr

요 약

다층 신경회로망의 모델의 크기는 적용분야에 따라서 임의로 선택되어지고, 최적의 네트워크 크기는 긴 시간에 걸친 시행착오를 통하여 결정된다. 본 논문에서는 은닉층의 뉴런 수를 학습 과정에서 유동적으로 결정하는 역전파 알고리즘을 제안한다. 기존의 Narendra의 모델의 동정에 대하여 제안한 알고리즘의 유용성을 비교 검토하였다.

I. 서론

최근 들어 고도로 발달하여 가는 고성능의 로봇, 항공기나 인공위성과 같이 복잡하고 불확실하며 비선형성이 강한 시스템의 제어는 고전적인 선형화 제어방식으로는 적용의 한계가 존재한다. 이에 대처하는 자기동조제어기(self-tunning regulator), 적응예측제어, 강인 제어 등의 방식이 존재하지만[1][2][3], 이러한 방식도 제한된 불확실성 내에서 적용가능하며, 비선형성이 강하거나 동작점의 범위가 넓은 경우에는 적용하는데 문제점이 있다. 이러한 이유로 최근에 지능제어 방식으로 전문가 법칙에서 경계의 불확실성을 해결하고자하는 퍼지 제어나 생물학적 신경계통을 모방한 신경회로망 제어 등이 많이 사용되고 있다[4].

생물체의 신경계를 근사 모델링한 것으로 대표적인 다층 신경회로망은 많은 뉴런들을 3층 이상의 망구조로 연결함으로써 임의의 비 선형 함수를 나타낼 수 있다. 최적의 다층 신경회로망의 모델링은 최적의 뉴런수를 결정하고 적절한 활성화 함수를 선택하는 것이라고 볼수 있다.

다층 신경회로망의 모델 크기는 여러가지 적용환경에 따라 임의로 선택되어지는 경우가 많고, 최적의 모델 크기는 긴 시간에 걸친 반복에 의한 시행착오를 통하여 결정된다. 다층 신경망에 있어서, 출력 뉴런들은 입력된 패턴들과 출력 가중치들에 의해서 다양하고 복잡한 결정 영역을 형성한다. 이러한 가중치들은 출력 뉴런들을 모든 은닉층 뉴런들과 연결한다. 결정 영역의 복잡한 정도에 의해, 모든 출력 뉴런은 정확한 출력을 얻기 위해 일정수의 은닉층 뉴런들을 필요로 한다. 다시 말하자면, 결정 영역의 복잡도가 높아질수록 출력측이 필요로 하는 은닉층 뉴런들이 증가한다는 것이다.

본 논문에서는 다층 신경회로망 모델의 은닉층 뉴런의 개수를 최적화 하는 역전파 알고리즘을 제안한다. 제안한 알고리즘은 모든 출력의 평균자승 합(Mean Square Error:MSE)의 변화를 관찰하고 이것을 이용하여 출력들이 수렴되는 방향으로 학습되는지를 관찰하여 은닉층이 더 확장되는 것을 막는다. 또한, 최적의 은닉층 뉴런수를 결정하고, 동시에 부분최소점(local minima)에 빠지는 것을 피하려고 시도하는 과정에서 학습이 완료되는 동안 은닉층 뉴런의 수가 유동적으로

변하게 된다. 수렴의 기준은 미리 설정된 MSE의 최소 변화율을 기준으로 하여 실제 변화율이 상대적으로 클 경우 수렴중으로 판단하여 학습을 계속하고, 상대적으로 작을 경우 은닉층을 확장하게 되며, 이러한 과정이 MSE가 미리 설정된 오차보다 작아질 때까지 계속된다.

제안한 방식의 유용성을 확인하기 위하여 기존의 Narendra모델의 동정에 대하여 시뮬레이션 한다.

II. 일반적으로 사용되는 신경회로망 모델 topology

일반적으로 사용되는 역전파 신경회로망 (Back-Propagation Neural Network)은 선형 뉴런의 입력층과 출력층 사이에 하나 이상의 중간층, 즉 비선형 뉴런층으로 구성된 은닉(hidden)층을 갖는 그림 1과 같은 다층(multi-layer) 구조의 회로망이다.[4].

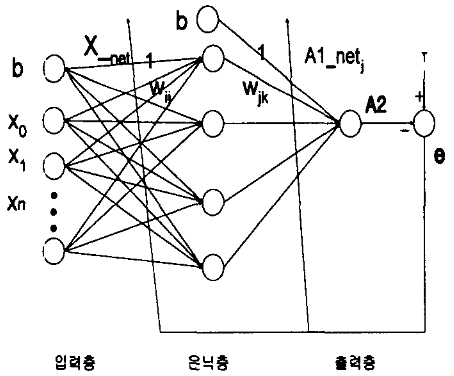

그림1. 다층 신경회로망 모델

목적 패턴이 주어져서 신경회로망의 출력 패턴이 목적 패턴과 근사화 되도록 학습하는 감독학습(upervised learning) 에 주로 사용되는 역전파 알고리즘에 있어서, 주어진 환경에 맞는 모델의 크기를 결정 하는 것은 매우 중요하다. 일반적으로, 모델의 크기가 너무 크면 모델의 자유도가 학습하고자 하는 샘플들 보다 더 많기 때문에 학습 패턴을 기억하는 경향이 생기게 될 것이고, 반면에 너무 작을 경우에는 초기치와 학습 패턴들에 너무 민감하여 문제를 해결하지 못하는 경우가 발생할 것이다.

기존의 역전파 알고리즘에 있어서, 모델의 크기는 주로 오랜 과정에 걸친 시행착오에 의해 결정되어졌다. 이것은 다양한 크기의 모델들을 학습시켜서 그 중에서 입력 패턴을 제대로 학습한 가장 작은 것을 선택하는 것이다. 이 방식은 최적의 모델 크기를 결정하기 위해서 상당량의 학습을 해야 하기 때문에 다소 비효율적이다. 우리는 새로운 알고리즘에서 학습 과정에서 네트워크의 크기를 결정하는 확장 은닉층 기술을 적용했다[5].

역전파 알고리즘은 steepest descent 기술을 기본으로 하기때문에, 자주 부분 최소점에 빠진다. 이경우, 은닉층 뉴런이 모델에 추가되면 여분의 단층 뉴런이 가중치 공간을 바꾸게 될 것이다. 이것이 네트워크가 부분 최소점에 빠지는 것을 방지한다[5].

III. 제안한 알고리즘

다층 신경회로망 모델은 역전파 알고리즘에 의해서 학습되며, 반복되는 모든 순간에 출력 뉴런에 대한 MSE가 계산되고, MSE의 변화율을 계산하여 이것이 일정한 값 이상일 경우 수렴중인 것으로 판단하여 학습을 계속하고, 일정값 이하일 경우에는 은닉층의 뉴런을 추가하여 학습을 계속하게 된다. .이 과정은 전체 MSE가 이전에 설정된 오차로 떨어질 때까지 계속된다.

역전파 알고리즘에서, 각각의 가중치는 다음과 같이 변한다.

$$\triangle w_{ij}(t) = \eta\delta_j x_i + \alpha\triangle w_{ij}(t-1) \quad (1)$$

여기서 x_i는 뉴런 j로의 입력 값이고 η는 학습률, α 는 모멘텀, δ_j는 오차 항이 된다:

이때 출력 뉴런에 대한 δ_j는 아래와 같이 주어진다.

$$\delta_j = o_j(1-o_j)(t_j - o_j) \quad (2)$$

여기서 o_j 는 뉴런 j의 출력이고 t_j는 j번째 목표값이다.

은닉층의 가중치에 대하여,

$$\delta_j = x_i(1-x_j)\sum_k \delta_k w_{jk} \quad (3)$$

여기서 k는 k번째 출력 뉴런을 의미한다.

학습도중에 MSE가 미리 설정한 오차 ε보다 작아지면 학습을 종료하게 되며 전체적인 MSE는 식(4)와 같다.

$$MSE = \frac{1}{P}\frac{1}{N}\sum_{p}\sum_{j}(t_j - o_j)^2 \qquad (4)$$

여기서 P는 입력패턴의 개수이고 N은 출력층 뉴런의 개수이다.

출력이 수렴하는가 하지 않는가를 판단하는 기준이 되는 ε의 정확한 값을 설정하는 것은 매우 중요하다. 다소 큰 ε을 선택하면 출력 뉴런들에 대해 완전하지 않은 은닉층 크기가 설정되고, 반면에 작은 ε 을 선택하게되면, 출력 값이 요구되어지는 목표값에 보다 근접하게 되나 다수의 불필요한 뉴런을 제거할 수 없게 되는 경우가 발생한다.

마지막으로 은닉층 뉴런의 증가 여부를 결정하는 k 번째 MSE의 변화율은 아래와 같다.

$$\rho_k = \frac{|MSE_{k-1} - MSE_k|}{MSE_k} \qquad (5)$$

여기서 k는 가장 최근에 학습이 끝난 구간이다. 이때 ρ는 한 구간에서의 전체 MSE의 변화율과 같다. 만약 ρ가 상대적으로 크다면 , 네트워크는 현재의 은닉층 뉴런으로 결과 값에 접근 해 가는 중임을 의미한다. 그렇지 않으면, 네트워크는 결과 값을 찾기 힘든 상태이거나 부분 최소점에 빠진 상태이고[1], 하나 혹은 더 이상의 은닉층 뉴런을 추가하여 요구된 결과를 얻을 때까지 재학습해야 한다.

Ⅳ 시뮬레이션

일반적인 다층신경회로망에서 은닉층이 1개인 경우를 사용하여 K. S. Narendra가 제안한 모델 4 가지 중 모델 Ⅰ과 모델 Ⅳ에 대하여 제안한 알고리즘을 시뮬레이션 한다[6].

Model Ⅰ:

식별하고자 하는 시스템은 식(6) 과 같다.

$$\begin{aligned} & y_p(k+1) \\ & = 0.3y_p(k) + 0.6y_p(k-1) + f[u(k)] \end{aligned} \qquad (6)$$

시스템을 동정하기 위하여 구성되는 직 병렬 모델은 식(7)과 같이 나타낼 수 있다.

$$\begin{aligned} & \widehat{y_p}(k+1) \\ & = 0.3y_p(k) + 0.6y_p(k-1) + N[u(k)] \end{aligned} \qquad (7)$$

시스템의 비 선형 부분인 f[u(k)]는 식 (8)과 같다.

$$\begin{aligned} f[u(k)] &= 0.6\sin(\pi u(k)) \\ &+ 0.3\sin(3\pi u(k)) + 0.1\sin(5\pi u(k)) \end{aligned} \qquad (8)$$

이때 시스템의 입력 신호식은 식 (9)와 같다.

$$u(k) = \sin\left(\frac{2\pi k}{250}\right) \qquad (9)$$

제안한 방식을 이용한 신경회로망의 동정특성은 그림 2이고 그림 3는 오차 곡선이다.

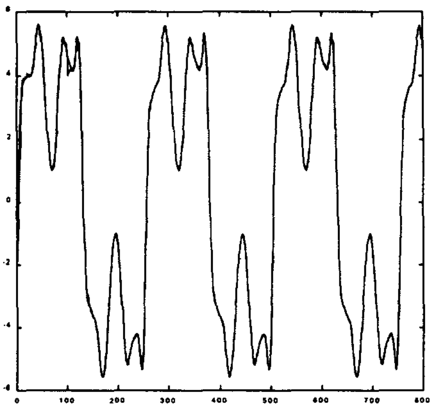

그림 2. Model Ⅰ의 동정특성

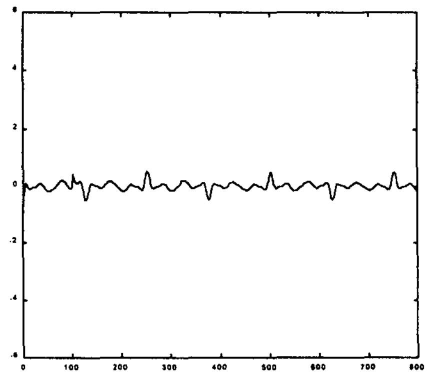

그림 3. Model Ⅰ의 오차곡선

Narendra의 시스템 식별에서는 1개의 은닉층과 30개의 뉴런으로 구성하여 식별하였으나 제안한 알고리즘으로는 15개의 뉴런만으로 식별이 가능하였다.

Model Ⅳ:

식별하고자 하는 시스템은 식(10) 과 같다.

$$y_p(k+1)=f[y_p(k), y_p(k-1), y_p(k-2), u(k), u(k-1)] \quad (10)$$

알려지지 않은 함수 f는 식 (11)과 같다.

$$f[x_1, x_2, x_3, x_4, x_5]=\frac{x_1x_2x_3x_4x_5}{1+x_3^2+x_2^2} \quad (11)$$

이때 시스템의 입력 신호식은 식 (12)와 같다.

$$u(k)=\sin\left(\frac{2\pi k}{250}\right) \quad (12)$$

제안한 방식을 이용한 신경회로망의 동정특성은 그림 4이고 그림 5는 오차 곡선이다.

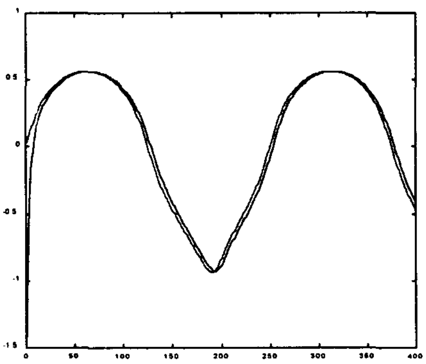

그림 4. Model Ⅳ의 동정 특성

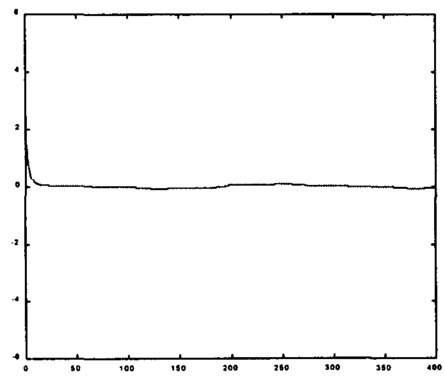

그림 5. Model Ⅳ의 오차곡선

Narendra의 시스템 식별에서는 1개의 은닉층과 30개의 뉴런으로 구성하여 식별하였으나 제안한 알고리즘으로는 17개의 뉴런만으로 식별이 가능하였다[6].

이상과 같이 제안한 방식은 Narendra의 신경회로망보다 은닉층의 뉴런수를 약 50% 감소시킴을 확인하였다.

V. 결론

본 논문에서는 다층 신경회로망 모델의 은닉층 뉴런의 개수를 최적화 하는 역전파 알고리즘을 제안하였다. 제안한 알고리즘은 모든 출력의 MSE의 변화율을 관찰하여 은닉층의 뉴런수를 유동적으로 조절한다.

제안된 알고리즘의 유용성을 확인하기 위하여 Narendra의 모델을 이용하여 시스템을 동정하여 보았다. 실제 Narendra가 제안한 모델에서 사용한 뉴런수의 약 50% 정도의 뉴런만을 가지고 동일한 수준으로 시스템을 동정할 수 있었음을 확인하였다.

[참고문헌]

[1] I,D.Landau, SYSTEM IDENTIFICATION AND CONTROL DESIGN, Prentice-Hall,Inc., 1990.

[2] K.S.Narendra and A.M.Annaswamy, Stable Adaptive Control, ADDISON- WESLEY PUBLISHING COMPANY, 1989.

[3] K.J.Astrom and B.Witenmark, ADAPTIVE CONTROL, ADDISON - WESLEY PUBLISHING COMPANY, 1989.

[4] Madan M. Gupta, Naresh K. Sinha, Intelligent Control Systems: Theory and Applications, IEEE PRESS., 1996

[5] Y.Hirose. K Yamashita. and S. Hyiya. "Back-Propagation Algorithm Which Varies the Number of Hidden Units". Neural Networks. Vol.4.

[6] K. S. Narendra, K. Parthadarathy, "Identification and Control of Dynamical Systems using Neural Networks, vol.1, mo.1, pp.4-27, 1990

[7] I. D. Landau, "SYSTEM IDENTIFICATION AND CONTROL DESIGN,", prentice Hall Inc, 1990.