

## 신경망 학습 변수의 시변 제어에 관한 연구

박 종 철, 원 상 철, 최 한 고  
금오공과대학교, 전자공학부

### A study on time-varying control of learning parameters in neural networks

Jong-chul Park, Sang-chul Won, Han-go Choi  
School of Electronic Engineering, Kumoh Nat. Univ. of Tech.  
hgchoi@knut.kumoh.ac.kr

#### Abstract

This paper describes a study on the time-varying control of parameters in learning of the neural network. Elman recurrent neural network (RNN) is used to implement the control of parameters. The parameters of learning and momentum rates in the error backpropagation algorithm are updated at every iteration using fuzzy rules based on performance index. In addition, the gain and slope of the neuron's activation function are also considered time-varying parameters. These function parameters are updated using the gradient descent algorithm. Simulation results show that the auto-tuned learning algorithm results in faster convergence and lower system error than regular backpropagation in the system identification.

#### I. 서 론

최근에 동적 신경망이 temporal processing이 요구되는 분야, 즉, 시계열 예측, 비선형 필터링, 제어 플랜트 모델링, 패턴인식, 그리고 다른 디지털 신호처리 등에서 실질적인 문제 해결을 위해 많이 사용되고 있다[1]. 이러한 분야에서 일반적으로 가장 많이 사용하고 있는 동적 신경망으로

서 리커런트 신경망(recurrent neural network, RNN)이 사용되고 있으며, 뉴런의 활성화함수로 미분가능 함수를 많이 사용하고 있다. 신경망의 학습과정에서 안정되고 빠른 학습이 요구되는데 기존 신경망에서는 학습과 관련된 변수, 즉, 학습율과 모멘텀 율은 사용자의 경험에 의해 상수 값을 사용하고 있다. 이러한 변수들을 일정한 상수 값을 설정하여 학습할 경우 학습이 느리거나 최종 에러가 발산하는 경우가 가끔씩 발생하고, 학습 도중에 학습이 정체되는 경우가 생긴다. 따라서 이러한 현상을 줄이고 신경망의 성능을 향상시키기 위하여, 즉, 이러한 변수를 시변(time varying) 변수로 가정하여 학습이 반복될 때마다 자동으로 변수 값을 제어함으로써 더 안정되고 빠른 수렴을 구현할 수 있다.

본 논문에서는 시변 변수를 제어하기 위해 널리 사용되고 있는 오차 역전파 (error backpropagation) 학습 알고리즘의 학습율과 모멘텀 율을 시변으로 가정하고 학습과정에서 수렴 속도를 향상시키고 최종 시스템 에러를 최소화하기 위해 fuzzy rule을 도입하여 자동으로 변수 값들을 시간에 따라서 갱신되도록 하였다. 보다 안정적인 fuzzy rule의 작성을 위해 모멘텀 율과 학습율에 원하는 성능지수에 근거로 하여 fuzzy rule을 도입하였다[2]. 또한, 뉴런의 활성화함수로 hyperbolic tangent 함수를 선정하였으며, 함수의 기울기와 이득 변수도 시변으로 가정하고

gradient descent 알고리즘을 사용하여 매 epoch마다 갱신되도록 하였다. 제안된 방법은 히든 층의 모든 뉴런이 입력층으로 궤환되어 fully connected되는 Elman 리커런트 신경망에 대해 적용하였다. 제안된 알고리즘에 대한 성능평가를 위해 선형과 비선형 시스템 identification을 수행하였다.

## II. 신경망 변수 갱신

그림 1은 이러한 학습 및 함수 변수들을 갱신하기 위해 사용한 Elman 리커런트 신경망 구조를 보여주고 있다.

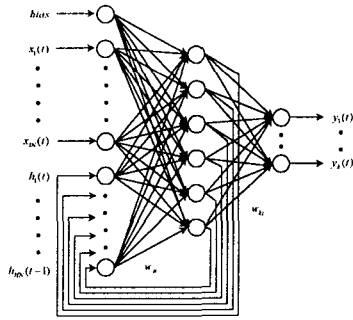


그림 1. Elman 리커런트 신경망 구조

Elman 구조는 히든 층의 모든 뉴런출력이 입력층으로 궤환되는 구조로서, 층간의 연결이 모두 연결되는 fully-connected 신경망이다. 이러한 궤환구조가 신경망의 동특성을 향상시킨다.

다층 신경망에서 사용되는 오차 역전파 알고리즘의 일반적인 연결계수  $w$ 에 대한 갱신 알고리즘은 다음과 같다.

$$\Delta w(t+1) = -\eta \Delta E(w(t)) + \alpha \Delta w(t) \quad (1)$$

여기서  $E$ 는 cost function,  $\eta$ 는 학습율, 그리고  $\alpha$ 를 포함하는 항을 모멘텀항이라 하며  $\alpha$ 를 모멘텀율이라 한다. 신경망의 수렴특성을 향상시키기 위해 상수로 설정하여 사용하고 있는 학습율과 모멘텀율을 시간함수로 가정하고 다음과 수정한 갱신 알고리즘을 사용한다.

$$\Delta w(t+1) = -\eta(t) \Delta E(w(t)) + \alpha(t) \Delta w(t)$$

학습 변수  $\eta(t)$ ,  $\alpha(t)$ 를 시스템 오차  $E$ 가 최소가 되도록 자동 동조시킴으로써 안정되고 더 빠른 수렴을 구현하고자 한다. 이를 위해 시변 학

습변수는 fuzzy 제어를 설계하여 매 epoch마다 갱신하도록 하였다.

학습 변수를 제어하기 위한 피드백 값으로 신경망의 MSE(표의 E), MSE의 변화분(표의 CE), 그리고 CE의 변화분(표의 CCE)을 사용하였다. 그리고 MSE의 절대치는 시스템에 따라 가변적이므로 학습 대상 시스템 출력의 평균 파워(power)로 정규화 한 후 피드백 값으로 사용하였다. 따라서 입력 퍼지 소속함수의 퍼지 분할 값은 출력 평균 파워에 대한 오차의 파워 비가 된다. 학습변수 변화분  $\Delta w$ 의 제어 규칙은 피드백 입력 E와 CE간의 관계로 나타낼 수 있는데, 표 1의 E-CE 규칙표와 같은 값으로 설정하였다.

표 1. 학습을 제어를 위한 퍼지 규칙표

E \ CE	ZE	PS	PB
NB	ZE	ZE	ZE
NS	ZE	ZE	PS
ZE	NS	ZE	PS
PS	NS	NS	ZE
PB	NS	NS	NS

습변수 변화분  $\Delta ce$ 의 제어 규칙은 피드백 입력 CE와 CCE간의 관계로 나타낼 수 있는데, 표 2의 CE-CCE 규칙표와 같은 값으로 설정하였다.

표 2. 모멘텀을 제어를 위한 퍼지 규칙표

CE \ CCE	NB	NS	ZE	PS	PB
NB	NS	NS			
NS	NS				
ZE		PS	PS	PS	
PS					NS
PB				NS	NS

위의 제어 규칙만 사용할 경우 학습 변수가 급격히 작아져 오차곡선의 local minimum에 도달하므로 더 이상의 학습이 진행되지 못한다. 그리고 퍼지 출력함수 스케일에 따라 학습 변수가 급격하게 변화하므로 학습효과가 저하될 뿐만 아니라 진동이 발생할 수 있으므로 이를 보완하기 위해 몇 가지 보조 규칙을 도입하였다.

퍼지 소속함수의 입력과 출력의 값은 그림 2와 3과 같다.

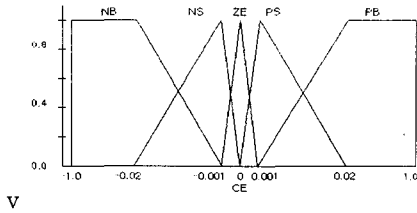


그림 2. 퍼지 입력 소속함수

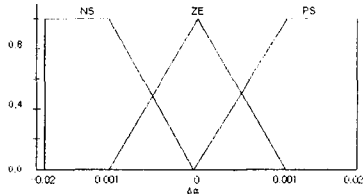


그림 3. 퍼지 출력 소속함수

그림에서 퍼지 분할은 시스템의 출력 평균파위에 대한 오차 파워의 비를 성능지수로 설정하여 값을 결정하였다.

학습계수 뿐만 아니라 본 연구에서는 다층 신경망 뉴런의 활성화함수에 포함된 변수, 즉, 함수변수도 시간함수로 가정하고 학습이 진행될 때 갱신하도록 하였다. 일반적으로 많이 사용되고 있는 hyperbolic tangent 함수에 이득과 기울기를 시간변수로 정의하면 다음과 같다.

$$f(x) = g(t) \tanh(s(t)x) \quad (3)$$

여기서  $g(t)$ 와  $s(t)$ 는 각각 함수 이득과 기울기를 나타내며 시변변수이다. 이들 변수에 대한 갱신은 gradient descent 알고리즘에 따라 변화되도록 하였으며, 갱신 알고리즘은 다음과 같다.

$$\Delta g_k(t) = -\eta_g \frac{\partial E(t)}{\partial g_k(t)} = \eta_g \delta_g(t) \Big|_{g(t)} \quad (4)$$

$$\Delta s_k(t) = -\eta_s \frac{\partial E(t)}{\partial s_k(t)} = \eta_s \delta_s(t) \Big|_{s(t)} \quad (5)$$

신경망의 히든층에서는 hyperbolic tangent 활성화함수를 사용하였으나, 출력층의 뉴런은 선형함수를 사용하였으며 히든층과 출력층 뉴런출력은 각각 다음과 같다.

$$y_j(t) = f(net_j) = g_j(t) \tanh(s_j(t) net_j) \quad (6)$$

$$y_k(t) = f(net_k) = a \cdot net_k \quad (7)$$

### III. 실험결과

그림 1은 제안된 RNN 신경망의 성능평가를

위해 학습 및 함수 변수가 갱신하도록 학습알고리즘이 포함된 RNN을 이용하여 시스템 identification을 수행하기 위한 구성도이다.

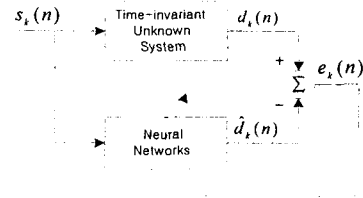


그림 4. 시스템 identification 구성도

입력으로는 가우시안 분포를 갖는 랜덤신호(평균=0, 표준편차=1)를 발생시켜 사용하였으며, 대상 시스템의 출력과 신경망 출력의 차이가 최소가 되도록 신경망의 연결계수, 학습 및 함수변수를 갱신하였다. 시험 대상 시스템으로 선형과 비선형 시스템에 대해서 각각 실험하였다. 신경망의 구조는 입력, 히든, 그리고 출력층의 뉴런수가 각각 4,8,1이며, 지연  $M=4$ 이다.

제안된 알고리즘에 대한 성능평가는 학습 및 함수변수를 상수로 사용한 RNN과 서로 비교하였는데, 서로 다른 입력과 초기치를 사용하여 10번 시스템 identification을 수행하여 평균한 결과를 상호 비교하였다. 초기 학습율=(0.5, 0.3, 0.1), 모멘텀율=0.01, 그리고 히든층의 함수 이득과 기울기는 모두 0.4를 사용하였으며, 2000번 학습을 수행하였다.

#### (1) 선형 시스템

선형 대상 시스템은 아래 식과 같이 정의된 식을 이용하였다.

$$y(n) = 0.921x(n) + 0.107x(n-1) + 0.336x(n-2) + x(n-3) + 0.336x(n-4) + 0.107 + 0.921x(n-6) \quad (8)$$

그림 5는 제안된 학습알고리즘과 기존의 단순 알고리즘과의 수렴과정을 보여주고 있으며, 표 3은 아래와 같은 식을 사용하여 시간영역과 주파수 영역에서의 오차값을 비교하고 있다.

$$t\_err = \frac{1}{N} \sum_{k=1}^N e_k^2 \quad (9)$$

$$f\_err = \sqrt{\frac{1}{N} \sum_{k=1}^N |fft(d(k)) - \hat{d}(k)|^2} \quad (10)$$

표 3에서의 FRNN#1은 학습변수만 갱신한 경우이고, FRNN#2는 학습 및 함수변수 모두 갱신을 의미한다.

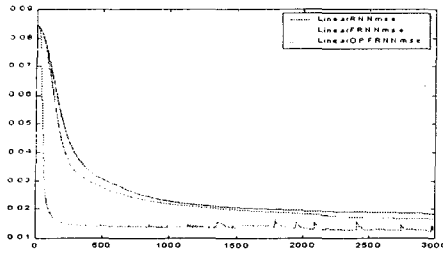


그림 5. 선형 시스템에 대한 수렴곡선(η = 0.1)

표 3. 선형시스템의 오차비교

η	Error	RNN	FRNN#1	FRNN#2
0.1	Time	0.0184	0.0168	0.0129
	Freq.	2.89	2.67	2.34
0.3	Time	0.0158	0.0160	0.0151
	Freq.	2.46	2.63	2.44
0.5	Time	0.0147	0.0155	0.0133
	Freq.	2.44	2.58	2.38

선형 시스템 identification에서 FRNN#2는 학습속도와 시스템 오차면에서 기존의 RNN보다 성능이 우수하였는데, 특히, 학습속도가 매우 향상되었다. FRNN#1은 RNN에 비해 수렴속도는 향상되었으나 시스템 오차는 거의 동일하였다. 기존 RNN의 경우 초기 학습율이 크면 학습이 진행됨에 따라서 진동이 발생하는데 퍼지 규칙을 적용할 경우 이러한 진동을 크게 감소시킨다.

(2) 비선형 시스템

비선형 대상 시스템으로는 다음과 같이 정의된 시스템에 대해 수행하였다.

$$y(n+1) = 0.15x(n) + 0.3x(n-1) + 0.6x^3(n) + 0.18x^2(n) - 0.24x(n) \quad (11)$$

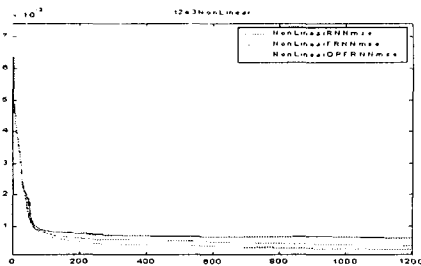


그림 6. 비선형 시스템에 대한 수렴과정(η = 0.5)

그림 6은 비선형 시스템에 대해 제안된 학습 알고리즘과 기존의 단순 학습 알고리즘과의 수렴 과정을 보여주고 있으며, 표 4는 식(13)과 식(14)

를 사용하여 시간영역과 주파수 영역에서 오차값을 상호 비교하고 있다.

표 4. 비선형 시스템의 오차비교

η	Error	RNN	FRNN#1	FRNN#2
0.1	Time	4.32e-4	4.32e-4	3.08e-4
	Freq.	0.452	0.459	0.386
0.3	Time	6.78e-4	4.12e-4	2.87e-4
	Freq.	0.711	0.427	0.348
0.5	Time	6.16e-4	4.01e-4	2.68e-4
	Freq.	0.671	0.412	0.322

비선형 시스템에서도 FRNN#2는 학습속도와 시스템 오차에서 기존의 RNN보다 성능이 우수하였으며, 초기 학습율이 클수록 FRNN#2의 오차가 RNN보다 더욱더 감소되었다. FRNN#1은 RNN에 비해 수렴속도는 향상되었으나, 초기 학습율이 작은 경우는 시스템 오차는 거의 동일하였으나 학습율이 클수록 성능이 우수하였다.

IV. 결론

본 논문은 신경망이 안정되고 수렴특성을 향상시키기 위해서 신경망의 학습에서 기존에 상수로 사용되었던 변수들을 시변변수로 가정하고 학습이 진행함에 따라서 자동 제어할 수 있는 학습 알고리즘을 제안하였다. 사용된 시변변수로 학습율과 모멘텀율과 같은 학습변수는 성능지수에 근거하여 fuzzy 규칙에 따라서 갱신되었고, 활성화 함수의 이득과 기울기와 같은 함수변수는 gradient descent 알고리즘에 의하여 갱신되었다. 제안된 학습 알고리즘은 Elman RNN에 적용하여 상수값을 사용하는 기존의 학습알고리즘과 비교하였다. 시험결과에 따르면 제안된 학습 알고리즘은 기존 방법에 비해 수렴 속도와 시스템 오차에서 더 우수한 성능을 보여주었다.

참고문헌

[1] G. Panda and T. Chatterjee, "Broadband noise cancellation using a functional link ANN based nonlinear filter," 1977 ICNN, vol. 4, pp. 2061-2066, U.S.A., Jun. 1997.  
 [2] C. T. Lin and C. S. G. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Practice Hall, NJ, 1996.