

論文97-34C-6-9

# 시계열 예측을 위한 1, 2차 미분 감소 기능의 적응 학습 알고리즘을 갖는 신경회로망

## (A Neural Network with Adaptive Learning Algorithm of Curvature Smoothing for Time-Series Prediction)

鄭 韶 永 \*, 李 敏 鎬 \*\*, 李 壽 永 \*

(So-Young Jeong, Minho Lee, and Soo-Young Lee)

### 요 약

이 논문에서는 우수한 일반화 성능을 갖는 함수 근사화기를 위한 새로운 신경회로망 학습 알고리즘을 소개하고, SantaFe 경쟁 데이터 Set을 이용한 시계열 예측 문제에 적용하여 그 성능을 검증한다. 신경회로망의 일반화 성능을 향상시키기 위해서 기존의 학습을 위한 오차 함수에 은닉층 뉴런의 활성값을 제한하는 부가적인 오차 함수를 추가한 새로운 성능 지수 함수를 이용한다. 부가적으로 고려한 은닉층 뉴런의 오차 함수에 의해 다층 구조 신경망은 함수 근사화 과정에서 보다 부드러운 함수 곡면 근사화 특성을 갖게 되어 보다 우수한 일반화 성능을 갖는다. 제안되는 학습 알고리즘은 새로운 성능 지수 함수에 최대 급경사 강하법을 적용함으로써 기존의 오차 역전달 학습 알고리즘과 Hebb 학습 법칙, 그리고 연결강도 감쇄가 혼합된 형태의 학습 알고리즘이다.

### Abstract

In this paper, a new neural network training algorithm will be devised for function approximator with good generalization characteristics and tested with the time series prediction problem using SantaFe competition data sets. To enhance the generalization ability a constraint term of hidden neuron activations is added to the conventional output error, which gives the curvature smoothing characteristics to multi-layer neural networks. A hybrid learning algorithm of the error-back propagation and Hebbian learning algorithm with weight decay constraint will be naturally developed by the steepest decent algorithm minimizing the proposed cost function without much increase of computational requirements.

### 1. 서 론

신경회로망은 대상 시스템에 대한 입출력 데이터만

을 가지고 시스템의 특성을 적절히 추출할 수 있기 때문에 적응 제어나 예측, 음성 인식, 패턴 인식 및 최적화 등의 수학적 해결에 의해 풀기 어려운 여러 공학 분야에 널리 이용되고 있다. 특히 패턴 분류나 함수 근사화와 같은 문제에 다층 구조 신경회로망을 적용함으로써 기존의 예측 및 진단 모델로 많이 쓰여온 단순 회귀, 다중 회귀 모델, 단일 다변량 모델 및 상태 공간 모델에 비해 탁월한 성능을 보이고 있다. 신경회로망 모델의 일반화 성능은 학습해야 할 자료의 양과 해결해야 할 문제의 복잡도, 그리고 신경회로망 자체의 자유도 등에 크게 영향을 받는다. 일단 신경회로망의 은

\* 正會員, 韓國科學技術院 電氣 및 電子工學科  
(Dept. of Electrical Engineering, KAIST)

\*\* 正會員, 韓國海洋大學校 電氣工學科  
(Dept. of Electrical Engineering, Korea Maritime University)

※ 본 연구는 시스템 공학연구소 위탁과제(핵심소프트웨어 개발사업)로 수행되었음.

接受日字:1996年10月31日, 수정완료일:1997年6月9日

닉층 뉴런의 수가 정해지면 일반화 성능을 고려할 때 필요한 데이터의 최소량에 대한 이론적 연구 결과들이 발표되고 있지만, 학습할 데이터에 대해 신경회로망의 크기와 구조를 일반적으로 구하는 방법은 알려져 있지 않다.<sup>[11]</sup> 크기가 작은 신경회로망으로는 실질적인 문제를 정확히 나타내는데 한계가 있지만, 크기가 너무 큰 신경회로망은 overfitting이 일어날 가능성이 있다. 따라서 이러한 문제들을 해결하기 위해 신경회로망의 복잡도를 줄이는 방법으로는 pruning algorithm,<sup>[2]</sup> synaptic weight elimination,<sup>[3]</sup> weight decay,<sup>[4-6]</sup> weight sharing,<sup>[7-9]</sup> 등과 같은 알고리즘들이 있다. 한편 학습시킬 데이터에 잡음을 첨가시키거나 오차 함수에 추가 항을 더해서 문제의 복잡도를 증가시켜 overfitting을 막으려는 방법들도 제안되고 있다.<sup>[10,11,12]</sup>

이 논문에서는 학습 과정에 소요되는 계산량의 큰 증가 없이 보다 우수한 일반화 성능을 가지는 알고리즘을 개발하고 예측 모델의 비교로 많이 쓰이는 SantaFe competition data set을 통해 알고리즘의 우수성을 보인다. 오차 함수에 입력에 대한 출력 측의 민감도와 관련된 항과 저주파 필터링에 관계되는 항을 추가해 신경회로망 내부에서 위의 기능들을 처리할 수 있는 보다 일반성을 가지는 알고리즘이 제안된다. 그리고 시계열(Time Series) 예측 문제는 single step mode와 iterative step mode로 나뉘어 있는데 single step prediction은 예측하여 나온 값을 이용하지 않고 실제 주어진 데이터만을 입력으로 그 다음 값을 예측해 가는 방법이고 iterative step prediction은 예측되어 나온 값을 다시 입력으로 해서 예측해 가는 방법으로서 초기의 예측 오차가 누적되어 예측 항이 많아지면 오차가 커질 위험이 있다.<sup>[12]</sup> 따라서 초기 오차에 민감하지 않는 신경회로망의 모델의 필요성이 제기되며 다음 장의 1절에서 다루고, 실제의 많은 예측 문제에서 나타나는 고주파 특성인 overfitting을 줄이기 위해 저주파 필터링을 하는 방법이 다음 장의 2절에서 소개된다. 제안된 알고리즘의 일반화 성능을 검증하기 위한 Bench Mark 자료로 많이 이용되는 SantaFe 경쟁 Data를 이용하여 시계열 예측 문제에 적용하고 기존의 연구 결과들과 비교한다.

## II. 본 론

신경회로망의 대부분의 지도 학습 알고리즘에서 오

차 함수는 출력층 오차들의 제곱의 합으로 정의되고 이 오차 함수를 최소화하여 주어진 데이터로부터 입·출력 관계를 학습하게 된다. 여기에선 입·출력 데이터 간의 일차 미분항(민감도와 관련됨)과 이차 미분항(주파수와 관련됨)을 오차 함수에 부가적인 항으로 첨가한다.

### 1. 입·출력간의 민감도를 낮추는 알고리즘

잡음이 섞여 들어온 입력 데이터나 신경회로망의 시냅스나 뉴런이 오동작을 할 때라도 정확하게 연상(association)이나 분류(classification)을 하기 위해서 입력 값에 대한 출력측의 민감도를 줄여야 할 필요가 있다.

L개의 층을 가진 전방향(feed-forward) 신경회로망에 연쇄법칙(chain rule)을 적용하면 다음과 같은 민감도에 관한 식을 얻을 수 있다.

$$\frac{\partial y_i}{\partial x_k} = \sum_{j_1, \dots, j_{L-1}} W_{j_{L-1}}^{(L)} W_{j_{L-2}}^{(L-1)} \dots W_{j_2}^{(2)} f_L(\hat{y}_i) \hat{f}_{L-1}(\hat{h}_{j_{L-1}}^{L-1}) \dots \hat{f}_1(\hat{h}_k^1) \quad (1)$$

여기서  $x_k$ 와  $y_i$ 는 각각 입력 벡터의 k번째 성분과 출력 벡터의 i번째 성분이고,  $W_{j_{L-1}}^{(L)}$ 은 L번째 층의 시냅스 연결을 나타낸다. 출력층이 L번째 층이라면 i는  $j_L$ 과 같게 된다.  $f_L(\cdot)$ 은 L번째 은닉층의 시그모이드(Sigmoid) 함수  $f_L(\cdot)$ 의 미분이고,  $\hat{h}_{j_L}^{L-1} = \sum_{j_{L-1}} W_{j_{L-1}}^{(L)} \hat{h}_{j_{L-1}}^{L-1}$ 은 L번째 은닉층의  $j_L$ 번째 뉴런의 시냅스 통과 후 활성화 신호이다. 그리고  $\hat{h}_{j_L}^L = f_L(\hat{h}_{j_L}^{L-1})$ 는 시냅스 통과 전 활성화 신호이다.

설명의 편의를 위해서 L번째 층의 모든 뉴런이 모두 같은 시그모이드 함수  $f_L(\cdot)$ 을 갖는다고 가정하였다. 함은 모든 은닉층의 모든 뉴런들에 대해 행해진다. 향상된 fault-tolerance 성능을 위한 추가 오차 함수는 모든 i, k 인자에 대한 민감도의 제곱의 합으로 정의되는데, 이때 계산의 효율과 해석의 편리를 위해 간단한 하나의 항을 쓰도록 한다. 식(1)로부터 은닉층의 활성화 신호를 포화(saturation)시켜서 민감도를 줄일 수 있다. 즉, 신경회로망의 학습 과정에서 출력측 오차를 줄이는 것뿐만 아니라, 동시에  $f_L(\hat{h}_{j_L}^{L-1})$ 도 아주 작은 값이 되도록 신경회로망의 연결 강도(Interconnection weight)를 구한다. 은닉층 뉴런들로 표현되는 hyperplane은 훈련 자료들로부터 멀어지게 되고, 시냅스

연결 강도(synaptic weight)와 자료의 작은 변화는 큰 차이를 만들지 않으므로 강인한 분류(robust classification) 성능을 가질 수 있을 것으로 기대된다.

일반적인 오차 함수 대신에 다음 식(2)의 새로운 오차 함수를 정의한다.

$$E = E_o + \gamma E_h = \frac{1}{M} \sum_{s=1}^M E_o^s + \frac{\gamma}{M} \sum_{s=1}^M E_h^s \quad (2)$$

여기서  $E_o^s = \frac{1}{2N_0} \sum_i (t_i^s - y_i^s)^2$  는 s번째 훈련 패턴의 정규화(normalized)된 출력 오차이다. 새로 추가된 은닉층의 벌칙(penalty)항은 다음과 같다.

$$E_h^s = \frac{1}{2} \prod_{l=1}^{L-1} \frac{2}{N_l} \left[ \sum_{j_l=1}^{N_l} \hat{f}(\hat{h}_{j_l}^{s,l}) \right] = \frac{2^{L-2}}{N_1 N_2 \dots N_{L-1}} \sum_{j_1, j_2, \dots, j_{L-1}} \hat{f}(\hat{h}_{j_1}^{s,1}) \hat{f}(\hat{h}_{j_2}^{s,2}) \dots \hat{f}(\hat{h}_{j_{L-1}}^{s,L-1}) \quad (3)$$

여기서,  $t_i^s$  와  $y_i^s$  는 각각 s번째 저장된 패턴에 대한 목표 출력값과 실제 출력값이다. 그리고  $\hat{h}_{j_l}^{s,l}$  는 1번째 은닉층의  $j_l$ 번째 뉴런의 시냅스 통과후 값이다. 여기서  $M$ ,  $N_0(=N_L)$ ,  $N_1$ 은 각각 저장된 패턴의 수, 출력 뉴런의 수, 1번째 은닉층의 뉴런의 수이다. 두 개의 오차 항은 이러한 수들로 정규화되어 있다. 만일 은닉층의 활성화 신호가 시그모이드(Sigmoid) 함수의 선형 구간에 있으면 시냅스 연결을 통과한 값에 매우 민감해지고 은닉층 오차  $E_h^s$  가 커진다. 출력 오차  $E_o$  와 은닉층 오차  $E_h$  는 모두 비슷한 값들로 정규화되어 있음을 주목할 필요가 있다. 즉, 매우 작은 초기 연결 세기를 가질 때, hyperbolic tangent sigmoid 함수에 대해 0.5 근처의 값으로 정규화된다.  $E_h$  를 최소화시켜서 신경회로망을 더 좋은 robustness를 갖는 비선형 구간에서 동작하게 할 수 있다.  $\gamma$  는 신경회로망의 학습 과정에서 은닉층 오차  $E_h$  의 출력 오차  $E_o$  에 대한 상대적 중요도이다.

신경회로망의 학습에 흔히 사용되는 최대 오차 급경사 강하법(steepest-descent error minimization) 알고리즘을 사용하여 신경회로망을 학습시킨다. 제일 마지막 층은 추가된 오차 항의 영향을 받지 않겠지만 다른 층에서는 전체 오차 함수의 각 weight에 대한 편미분 항이 오차 함수에 추가된다. 그리고 연결 강도(Interconnection weight)의 변경은 다음과 같다.

$$\Delta W_{j_l i_l}^{l,0} = -\eta_l \frac{\partial E}{\partial W_{j_l i_l}^{l,0}} = \eta_l h_{j_l}^{l-1} \delta_{j_l}^l \quad (4)$$

여기서  $\delta_{j_l}^l$  은 1번째 층에서의 시냅스 통과후 활성화 신호  $h_{j_l}^l$  에 대한 전체 오차 함수의 민감도이며 다음과 같이 역전파(backpropagation)으로 계산할 수 있다.

$$\delta_{j_l}^l = -\frac{\partial E}{\partial \hat{h}_{j_l}^l} = \hat{f}'(\hat{h}_{j_l}^l) \sum_{j_{l+1}=1}^{N_{l+1}} \delta_{j_{l+1}}^{l+1} W_{j_{l+1} j_l}^{l+1,0} + \gamma E_h \frac{\hat{f}'(\hat{h}_{j_l}^l)}{\sum_{k=1}^{N_l} \hat{f}'(\hat{h}_k^l)} \hat{h}_{j_l}^l \quad (5)$$

민감도  $\delta_{j_l}^l$  는 일반적인 경우처럼  $\delta_{j_l}^L = \frac{1}{MN_L} (t_{j_L} - j_{j_L}) \hat{f}'(\hat{y}_{j_L})$  로 출력층에서 정의된다.  $\eta_l$  은 1번째 층의 학습 계수이고 양극(bipolar) 시그모이드 함수  $f(x) = \frac{(1-e^{-x})}{(1+e^{-x})}$  에 대해  $f = \frac{(1-f^2)}{2}$  와  $f = -ff'$  가 성립한다.(윗참자 s는 생략하였다.) 식(4)와 식(5)로부터 연결 강도(Interconnection weight) 변경에 두 개의 성분, 즉 출력단 오차가 역전파된 항과 입력값의 변화에 대한 출력값의 민감도를 줄이기 위한 은닉 뉴런의 벌칙항으로 구성되어 있음을 알 수 있다. 추가된 항은  $h_{j_l}^{l-1} h_{j_l}^l$  에 비례하며, 제안한 알고리즘은 신경회로망의 학습 알고리즘으로 널리 사용되는 오차 역전파 알고리즘과 Hebbian 학습 규칙으로 이루어져 있다. 따라서, 시계열 예측 문제(Time series prediction problem)에서의 Iterative prediction 모드에서 예측 데이터의 차이에 의해 발생하는 예측 오차의 누적을 제안한 학습 알고리즘의 강건성(Robustness)에 의해 줄여줄 수 있을 것으로 기대된다.

## 2. 주파수 필터링을 하는 알고리즘

많은 예측 문제에 있어 출력 값의 급격한 변화량 - 높은 주파수 특성 - 에는 관심이 없기 때문에 curvature smoothing이나 저주파 필터링을 하기 위한 post-processing의 과정을 두는 경우가 많다.<sup>[13]</sup> 여기에선 이러한 과정을 신경회로망 자체의 부가적인 오차 함수에 포함시켜 주파수 특성을 개선하는 알고리즘을 제안한다.

신호  $y(x)$ 의 주파수는 입 · 출력 데이터의 이차미분 항 -  $(d^2 y / dx^2) / y$  - 과 관련이 있다고 볼 수 있으므로, 위 식(1)을 한 번 더 편미분해서 다음과 같은 식을 얻을 수 있다.

$$\frac{d^2 y_i}{dx_i^2} = - \sum_j W_{ij}^{(2)} f_j(\hat{h}_j) (W_{jk}^{(1)})^2 \quad (6)$$

여기서 논의를 간단하게 하기 위해 선형적인 출력을 갖고 하나의 은닉층으로 구성된 신경회로망을 가정하였다. 비선형 시그모이드 함수의 특징인  $f = -ff'$  를 이용하였다. 여기서도 이차 미분에 대한 정확한 식을 사용하는 대신에, 식(6)과  $y_i = \sum_j W_{ij}^{(2)} f_j(\hat{h}_i)$  를 비교하고 더 간단한 형태를 채택하여 다음과 같은 저주파 필터 특성을 갖는 추가 오차 항을 정의하도록 하자.

$$E_k^s = \frac{1}{2N_k} \sum_j \hat{f}(\hat{h}_j^s) \sum_k (W_{jk}^{(1)})^2 = \frac{1}{N_k} \sum_j \hat{f}(\hat{h}_j^s) e_j \quad (7)$$

여기서  $e_j = \frac{1}{2} \sum_k (W_{jk}^{(1)})^2$  는 첫 번째 층의 synaptic weight energy로 정의된다.  $f$  와  $e_j$  가 모두 양수임에 주목할 필요가 있다. 그리고, 출력 값의 차이에 의해서만 정의되는 오차 함수를 다음 식(8)과 같이 새롭게 정의한다.

$$E = E_o + \gamma E_h = \frac{1}{2N_o} \sum_i (t_i - y_i)^2 + \gamma E_h \quad (8)$$

$E_o$  는 출력층의 오차를 의미하고,  $E_h$  는 식(2.7)에서 나타낸 부가적인 제한 함수로 신경회로망의 입출력 데이터의 이차 미분 제한에 의해 신경회로망이 저주파수 특성을 갖도록 하기 위한 은닉층 벌칙(Penalty)을 위한 함수이다. 편의를 위해 첨자 s는 생략하였다. 이제 식 (8)에 최대 오차 급경사 강하법(Steepest descent error minimization method)과 Chain rule을 적용하면, 다음 식 (9)와 같은 학습 알고리즘을 얻게 된다.

$$\begin{aligned} \Delta W_{jk}^{(1)} &= \eta_1 \hat{f}(\hat{h}_j) [x_k \sum_i \delta_i W_{ij}^{(2)} + \gamma (x_k h_j e_j - W_{jk}^{(1)})] \\ W_{jk}^{(1)}(new) &= W_{jk}^{(1)}(old) + \Delta W_{jk}^{(1)} \end{aligned} \quad (9)$$

식 (9)에서 대괄호 안의 첫 번째 항은 일반적인 역전파 오차 항이며, 두 번째 항은 추가된 오차 항으로 Hebbian 학습규칙과 연결 강도 감소(Weight decay)로 구성되어 있다. 제한한 알고리즘은 표준 MLP모델에 계산량을 크게 증가시키지 않는 간단한 항을 추가함으로써 신경회로망의 주파수 특성을 좋게 할 수 있을 것으로 기대된다. 다음 장에서 SantaFe competition data set A를 이용한 컴퓨터 시뮬레이션으로 제안한 알고리즘들의 일반화 성능을 보인다.

### III. 시뮬레이션 결과 및 논의

그림 1은 시계열 예측 문제에서 많이 쓰이는 SantaFe competition data set의 특성을 가지는  $NH_3$  레이저 펄스의 모습을 나타낸 것이다. 이 데이터들 중에서 1000개를 신경회로망 학습에 이용하고 그 이후의 100개의 데이터를 예측하게 된다. 실험에 사용된 신경회로망은 2층의 구조로 되어 있고, 입력층 뉴런의 수가 25개, 은닉층 뉴런의 수는 40개 그리고, 1개의 출력층 뉴런으로 구성되어 있다. 다음 그림에서 두 가지 모드 - single step 과 iterative step - 로 실험된 결과를 보인다.

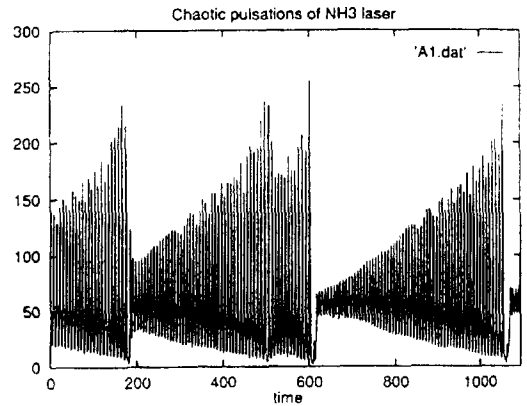


그림 1. 혼돈 특성을 갖는 레이저 데이터  
Fig. 1. Time points of chaotic laser data.

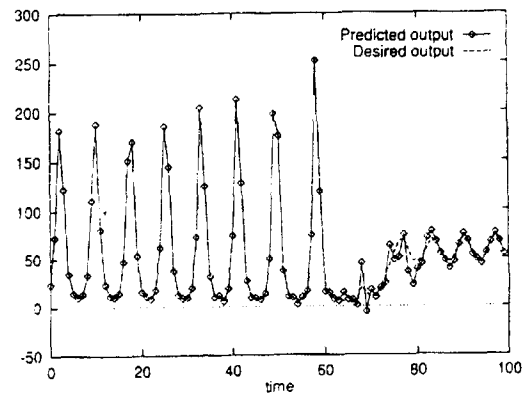


그림 2. 표준 MLP에서 Single step prediction 결과  
Fig. 2. Simulation result of single step prediction using standard multi-layer perceptron.

그림 2는 표준 MLP(multi-layer perceptron)모델

에서 Single step prediction을 한 결과를 나타낸 것이고, 그림 3은 제안된 1차 입출력 미분 감소를 추가적인 오차 함수로 갖는 신경회로망 모델에서  $\gamma$ 가  $10^{-2}$ 일 때의 Single step prediction을 보인 것이다. 그림 4는 표준 MLP 모델에서의 Iterative mode에 대한 실험 결과를 나타내고, 그림 5는 제안된 1차 미분 오차항을 갖는 모델에서의 Iterative mode에 대한 실험 결과를 나타낸다. 결과를 살펴보면 Iterative step에서 제안한 알고리즘의 우수성이 많이 보여짐을 알 수 있고, chaotic collapse가 일어나는 지점인 time point 60 근처까지 비교적 정확한 예측을 할 수 있다는 것을 볼 수 있다.

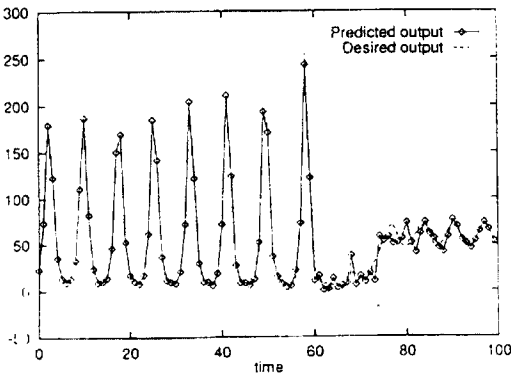


그림 3. 제안된 1차 입출력 미분 감소 신경회로망을 이용한 Single step prediction 결과  
Fig. 3. Simulation result of single step prediction using proposed model with 1st derivative constraint.

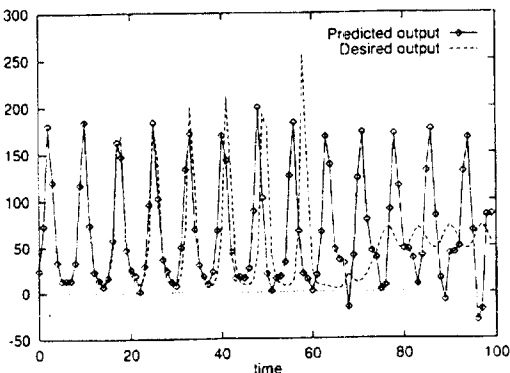


그림 4. 표준 MLP에서 Iterative mode prediction 결과  
Fig. 4. Simulation result of iterative mode prediction using standard multi-layer perceptron.

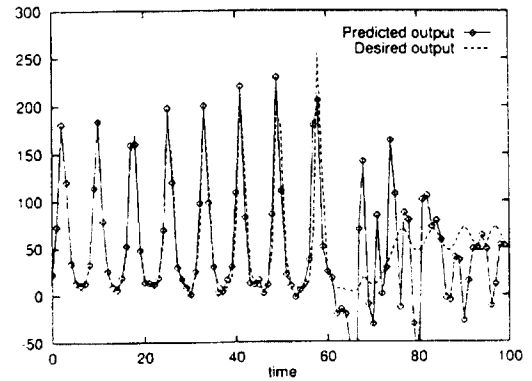


그림 5. 제안된 1차 입출력 미분 감소 신경회로망을 이용한 Iterative mode prediction 결과  
Fig. 5. Simulation result of iterative mode prediction using proposed model with 1st derivative constraint.

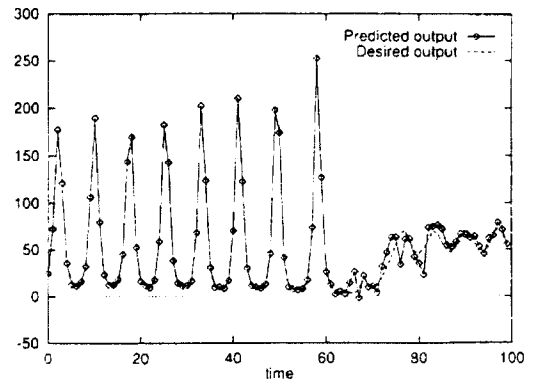


그림 6. 제안된 2차 입출력 미분 감소 신경회로망을 이용한 Single step prediction 결과  
Fig. 6. Simulation result of single step prediction using proposed model with 2nd derivative constraint.

그림 6은 추가적인 오차 함수에 저주파 필터링 기능을 할 수 있도록 제안한 모델에서  $\gamma$ 가  $10^{-2}$ 일 때 Single step prediction에 대한 실험 결과이고, 그림 7은 Iterative step prediction결과를 나타낸 것이다. 여기에서도 Iterative step일 때 표준 MLP모델보다 더 나은 성능을 보여주고 있음을 알 수 있다.

한편, 신경회로망을 주어진 데이터로 학습시킬 때 오차가 감소하는 그래프를 그림 8과 그림 9에 나타내었는데 제안한 알고리즘이 local minima에서 벗어나는 시간이 조금 빠름을 볼 수 있다. 저주파 필터링을 추가한 모델이 학습 시간이 가장 작게 걸리고 local mi-

nima에도 잘 빠지지 않고 오차가 감소함을 볼 수 있다. 제한한 알고리즘에서  $\gamma$ 는 출력층의 오차와 추가된 은닉층의 오차간의 상대적 중요성을 가리키는데  $\gamma$ 를 점점 증가시켜 가면 학습에 걸리는 시간이 많이 소요되고, 또 감소시켜 가면 표준 MLP 모델에 접근하기 때문에 학습 시간과 향상시킬 성능을 고려해 적절히 정해 줄 필요가 있다.

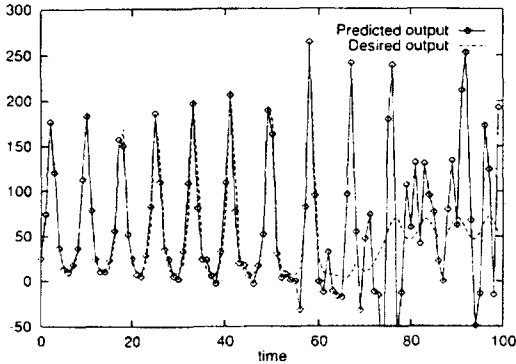


그림 7. 제안된 2차 입출력 미분 감소 신경회로망을 이용한 Iterative mode prediction 결과

Fig. 7. Simulation result of Iterative mode prediction using proposed model with 2nd derivative constraint.

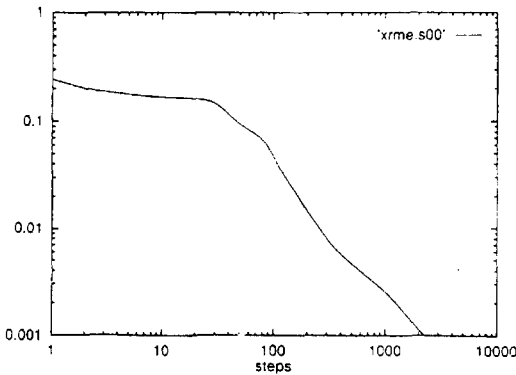


그림 8. 표준 MLP 모델로 학습시킬 때 오차 감소 곡선

Fig. 8. Error curve during training process of standard multi-layer perceptron.

다음 표 1과 표 2는 NMSE(normalized mean square error)를 통해 제안한 알고리즘과 표준 모델을 비교한 것이다. 표를 보면 Single step에선 오히려 제안된 모델이 약간 떨어지는 성능을 보이는 데 소수 셋째 자리까지는 학습할 때마다 달라질 수 있는 trun-

cation 오차로 볼 수 있고, Iterative step에선 제안한 모델이 더 우수함을 볼 수 있다. 그리고 예측한 100개의 데이터 중에서 60까지의 오차를 따로 보였는데 collapse가 일어나는 지점을 전후해 얼마나 잘 예측하는가를 보여주는 지표로 이용될 수 있다.

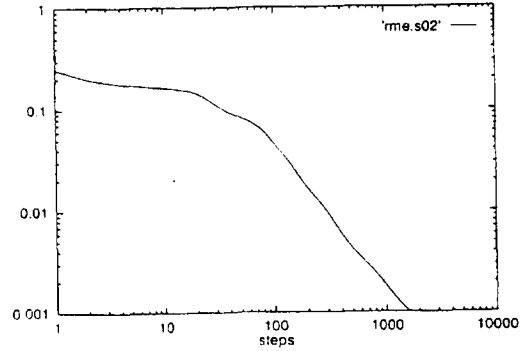


그림 9. 제안된 1차 미분 감소 신경회로망 학습시의 오차 곡선

Fig. 9. Error curve during training process of the proposed model using 1st derivative constraint.

표 1. 입출력 민감도를 줄이기 위해 제안한 1차 모델과 기존 모델에서의 NMSE 비교

Table 1. NMSE Comparison of the proposed 1st derivative model and the standard MLP model.

예측모드 \ 모델	표준 MLP 모델 (60스텝까지의 오차)	제안된 모델(1차미분항고려) (60스텝까지의 오차)
Single step prediction	0.013554 (0.001483)	0.013919 (0.002126)
Iterative step prediction	1.334526 (0.779717)	0.591314 (0.133684)

표 2. 저주파 필터링을 할 수 있도록 제안된 모델에서의 NMSE

Table 2. NMSE of the proposed 2nd derivative model.

예측 모드 \ 모델	제안된 모델 (2차 미분제한 항을 추가) (60스텝까지의 오차)
Single step prediction	0.016841 (0.001248)
Iterative step prediction	1.062002 (0.057070)

## IV. 결론 및 검토

이 논문에서는 우수한 예측 성능을 갖는 다층 구조 신경회로망 모델을 제안하였다. 입·출력 데이터의 급격한 변화나 외부의 잡음에도 강건한(Robust) 특성을 갖고 보다 일반화 성능을 향상시킬 수 있도록 오차 함수에 출력층 오차뿐만 아니라 은닉층의 오차들도 포함시켰다. 기존의 신경회로망에서 SantaFe competition data set A의 시계열 예측을 할 때 누적된 오차 때문에 Iterative step mode의 예측 결과가 별로 좋지 않았으나 제안된 알고리즘으로 보다 우수한 예측을 할 수 있음을 보였다.

제안된 알고리즘의 장점은 학습 알고리즘이 간단하고 설계하기가 쉽고 거기에 따른 추가적인 계산량은 별로 증가하지 않았다는 점에 있다. 또, 이 알고리즘은 신경회로망 학습의 대표적 알고리즘인 오차 역전파 알고리즘에 Hebbian의 학습 방법과 연결 강도 감소(weight decay) 알고리즘을 자연스럽게 포함하고 있고, 새로운 매개변수  $\gamma$ 에 의해 추가된 항의 중요도를 바꾸어 가면서 학습시킬 수 있다. 학습의 초기 상태에서는 매우 작은 초기 랜덤 연결값(weight value)을 이용하므로 Hebbian 항은 크게 영향을 미치지 못하고, 첫 번째 역전파 항에 의해 학습이 진행됨에 따라 은닉층의 활성화 신호가 커지고 Hebbian항은 은닉층의 활성화 신호를 포화 구간에 위치시키므로 weight decay와 경쟁하게 된다. 실험에서는  $\gamma$ 의 값을 적절히 설정해 주는 것이 쉽지 않았는데  $\gamma$ 를 체계적으로 결정해주는 방법에 대한 연구가 있어야 하겠고,  $\gamma$ 의 값을 바꾸어 가면서 적응적으로 신경회로망을 학습시키는 방법도 연구되어야 할 과제라고 볼 수 있다.

## 참 고 문 헌

[1] E.B. Baum and D. Hausser, "What size net gives valid generalization?," *Neural Information Processing Systems*, vol. 1, pp. 89-90, 1989.

[2] R. Reed, "Pruning algorithms - A survey", *IEEE Trans. on Neural networks*, vol. 4, no. 5, pp. 740-747, 1993.

[3] Y. LeCun, J.S. Denker, and S.A. Solla, "Optimal brain damage," *Advances in Neural Information Processing Systems*,

vol 2, pp. 598-605, 1990.

- [4] M. Ishikawa, "Structural learning in neural networks," *Proc. International Conf. on Fuzzy Logic, Neural Nets, and Soft Computing*, pp. 37-44, 1994.
- [5] A. Krogh and J.A. Hertz, "A simple weight decay can improve generalization," *Advances in Neural Information Processing Systems*, vol. 4, pp. 950-957, 1992.
- [6] A. Weigend, D. Rumelhart, and B. Huberman, "Generalization by weight-elimination with application to forecasting," *Neural Information Processing Systems*, vol 3, pp. 875-882, 1991.
- [7] K. Fukushima, "Improved generalization ability using constrained neural network architectures," *Proc. International Joint Conference on Neural Networks*, pp. 2049-2054, 1993.
- [8] Y. LeCun, B. Boser, J.S. denker, D. Henderson, R.E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, pp. 541-551, 1989.
- [9] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Trans. Acoustics, speech, Signal Processing*, vol. 37, pp. 328-339, 1989.
- [10] J. Sietsma, R.J.F. Dow, "Creating artificial neural networks that generalize," *Neural Networks*, vol. 4, pp. 67-79, 1991.
- [11] S.Y. Lee and D.G. Jeoung, "Hybrid Hebbian/Backpropagation learning rule for improved generalization of multilayer feedforward neural networks," *Proc. International Conf. on Neural Information Processing*, vol. 1, pp. 189-194, 1994.
- [12] A.S. Weigend, N. A. Gershenfeld, Time series prediction: Forecasting the future and understanding the past, 1994.
- [13] S.Y. Lee and M. Lee, "Curvature smoothing and improved generalization by hybrid back-propagation/Hebbian learning rule," *World Congress on Neural Networks*, vol. 2, pp. 485-488, 1995.

## — 저 자 소 개 —

李 敏 鎬(正會員) 第 32卷 B編 第 12號 參照



鄭 韶 永(正會員)

1975년 6월 8일생. 1996년 한국과학기술원 전기 및 전자공학과 졸업(학사). 1996년 ~ 현재 한국과학기술원 전기 및 전자공학과 석사과정. 주관심 분야는 신경회로망 모델링, 시계열 예측 등임

李 壽 永(正會員) KITE Journal of Electronics Engineering vol. 3 no. 1, pp. 77-78, 1992. 參照