하이브리드 학습알고리즘의 다층신경망을 이용한 시급수의 비선형예측

조 용 현, 김 지 영
대구효성가톨릭대학교 공과대학 전자정보공학부
yhcho@cuth.cataegu.ac.kr

Nonlinear Prediction of Time Series Using Multilayer Neural Networks of Hybrid Learning Algorithm

Yong-Hyun Cho, Ji-Young Kim
School of Electronics and Information Eng., Catholic Univ. of Taegu-Hyosung
yhcho@cuth.cataegu.ac.kr

< Abstract > This paper proposes an efficient time series prediction of the nonlinear dynamical discrete-time systems using multilayer neural networks of a hybrid learning algorithm. The proposed learning algorithm is a hybrid backpropagation algorithm based on the steepest descent for high-speed optimization and the dynamic tunneling for global optimization. The proposed algorithm has been applied to the 600 samples of 700 sequences to predict the next 100 samples. The simulation results shows that the proposed algorithm has better performances of the convergence and the prediction, in comparision with that using backpropagation algorithm based on the gradient descent for multilayer neural network.

1. 서 론

과거의 측정값으로 부터 지식을 추출하여 시급수의 미래값을 예측하는 것은 과학, 경제, 그리고 공학 등의 분야 등에서 매우 절실하게 요구되는 연구과제이다. 지금까지 이를 위한 많은 연구들이 활발하게 이루어지고 있다[1-4].

예측문제는 일반적으로 여러 가지 요소에 의하여 특징지워질 수 있다. 이러한 요소로는 이용되는 데이터의 수, 예측되는 시간단계, 그리고 과정의 랜덤정도(randomness) 등이 고려되고 있다. 예측문제를 위한 한 가지 강력한 접근은 주어진 동적 과정이나 현상을 기초로 하는 법칙을 찾는 것이다. 이때의 법칙이 상미 방정식으로 주어질 수 있다면 이를 해결함으로서 미래값을 예측할 수 있을 것이다. 그러나 동적과정에 대한 정보는 종종 부분적이며 불완전한 정보로 알려진 해석적인 모델로 예측이 불가능하다. 또 다른 접근으로 시급수에 내포된 어떤 경험적인 규칙을 찾는 것이다[1]. 이 방법에서도 실제로 주기성 등과 같은 규칙들은 잡음에 의해 감추어지며 심지어 어떤 동적과정은 데이터가 랜덤하게 주어지는 무질서한 시급수로 나타내질 수도 있다.

예측을 위한 기존의 방법들을 살펴보면 먼저, 시간에 대한 선형이나 지수형의 함수를 이용하여 데이터를 예측하는 선형회귀 모델과 좀 더 복잡한 과정를 위해 예측되는 변수와 또 다른 적당한 변수들의 가중된 조합을 이용하는 ARMA(auto-regressive moving average) 모델이 있다[2,3]. ARMA 모델은 이전의 변수들을 이용한다는 점에서 선형회귀 모델과 다르다. 그러나 이들 모델들은 과정이 비선형이고 이용되는 데이터가 많을 경우에 성능이 떨어지는 문제점이 있다. 한편, 또 다른 방법들은 주로 공학에 이용되는 방법으로 Kalman filtering과 Wiener-Volterra series에 기초를 둔 모델이다[3]. Kalman filtering에 기초를 둔 방법에서는 파라미터들과 공분산(covariance)이 시간에 따라 변한다 할지라도 이에 대해서 알려진 모델구조로 가정한다. 이 방법은 비교적 간단한 모델링에서는 잘 동작하나 잡음이 있는 과정에 대해서는 비선형 시스템의 동작을 정확하게 예측할 수 없는 불합리성이 존재한다. 또한 과거 데이터로부터 추출된 입출력관계의 확장다항식인 Wiener-Volterra series에 기초를 둔 방법은 임의의 함수를 근사화할 수 있다. 그러나 이 방법 역시 다중입력을 가진 모델에 대해서는 더욱 많은 입력 데이터를 요구한다. 따라서 기존의 예측방법들은 대규모의 비선형인 문제에서는 매우 비효율적이다.

최근, 대규모 분산 병렬처리와 학습 능력을 가진 신경망은 신호와 패턴처리 분야에서 특히 많은 관심이 되어왔다[1-5]. 이러한 속성을 가진 신경망에 기반을 둔 여러 가지 예측방법들이 기존의 방법들의 대안으로 제안되어 왔다. 신경망은 입출력 데이터 내에 포함된 어떤 비선형 연속함수를 근사화하거나 재구성할 수 있어 대단히 일반적 이면서도 융통성 있게 미래값을 예측할 수 있다. 이때 이용되는 신경망으로는 다층신경망, RBF(radial basis function)을 이용하는 신경망, 그리고 sigma-pi 신경망 등이 주로 이용된다[2]. 일반적으로 다층신경망은 미래값의 예측에 가장 널리 이용되고 있는 신경망으로 알려져 있다[1]. 이를 위한 학습알고리즘으로는 역전파(backpropagation : BP) 알고리즘이 가장 널리 이용되어 왔다[5]. 그러나 기울기하강(steepest descent)의 속성을 이용하는 역전파 알고리즘은 학습파라미터의 설정에 따라 수렴속도와 견실성 중 하나 이상의 문제를 가

지며, 전역최소점으로의 수렴이 보장되어 있지 않다. 이러한 문제를 해결하기 위한 여러 방법들이 많은 연구자들에 의해서 연구되어 왔다[5].

본 논문에서는 하이브리드 학습알고리즘의 3층 전향신경망을 이용하여 비선형 동적 불연속 시스템에서의 효율적인 시급수 예측에 대해서 제안하였다. 제안된 하이브리드 학습알고리즘은 기울기하강과 동적터널링(dynamic tunneling)의 속성을 조합한 역전파 알고리즘이다. 제안된 학습 알고리즘을 이용한 다층신경망을 700개의 순차적인 데이터 중에서 600 개를 학습시킨 다음 나머지 100 개의 데이터를 예측하는 시뮬레이션을 수행하여 기존의 기울기하강법에 기초를 둔 역전파 알고리즘을 이용한 신경망에 의한 결과와 비교 고찰하였다.

2. 신경망을 이용한 시급수 예측

한 개이상의 은닉층을 가지는 신경망은 다중 입력과 하나의 출력사이에 존재하는 비선형 특성을 사상시킬 수 있다. 이때 입출력 데이터 사이에 사상을 일반화시켜 시급수의 움직임을 지배하는 내재된 규칙을 찾아 향후 그것의 연속성을 예측하는 것은 신경망의 학습을 통하여 이루어 진다. 근사화 이론에서 보면, 신경망의 학습과정은 학습데이터에 가장 적합한 비선형 함수로 구성되는 다차원 공간을 형성하는 것으로 볼 수 있다. 신경망은 그러한 함수를 찾기 위해서 과거의 데이터들을 이용하여 학습된다. 충분한 과거의 데이터가 주어질 때 신경망의 일반화와 미래에 대한 예측은 근사화 이론으로 간단하게 설명될 수 있다. 일반화는 적합한 다차원의 근사화함수에 대해 시험데이터를 보간하는 것이며, 미래에 대한 예측은 다차원 함수의 추정으로 설명될 수 있다.

미지의 동적과정을 비선형 다중변수의 함수로 나타내면 다음과 같다. 즉,

$$y(k) = f[y(k-1),\ y(k-2),\ \cdots,\ y(k-n)] + e(k) \qquad (1)$$

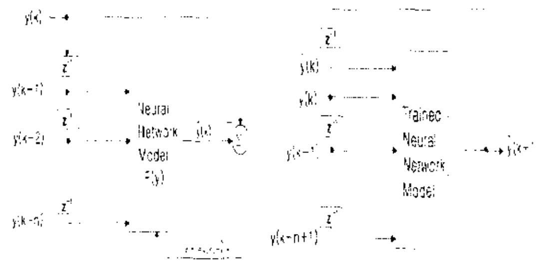
이다. 여기서 $y(k)$는 시급수의 표본들이며, $f(\cdot)$는 미지의 비선형함수이다. 이때 k는 $k = N, N-1, \cdots, n$로 $n < N$의 관계이고, N는 표본의 전체 개수이며 n는 예측차수이다. 또한 $e(k)$는 나머지 값으로 가우시안의 백색잡음으로 가정한다. 가우시안의 백색잡음은 잡음의 두 표본은 서로 무관하며, 이는 k 개의 통계적으로 독립인 잡음 표본이 있다는 것을 의미한다. 식 (1)에서 비선형함수는 다차원의 공간이며, 이는 시급수의 미래값은 n 개의 과거값의 비선형 함수로 표현됨을 의미한다. 결국 예측문제는 과거의 주어진 표본 $y(k)$로부터 미래의 표본 $y^\wedge(k)$를 구할 수 있는 비선형 함수 $f(\cdot)$로 표현되는 적당한 모델을 찾는 문제이다. 미래의 표본 $y^\wedge(k)$는 다음과 같이 나타낼 수 있다. 즉,

$$y^\wedge(k) = F[y(k-1),\ y(k-2),\ \cdots,\ y(k-n)] + e(k) \qquad (2)$$

이며, 여기서 $F(\cdot)$는 미지의 비선형 함수 $f(\cdot)$의 근사화 함수이며, 신경망으로 계산될 수 있다. 이때 원하는 표본값은 실제의 표본값 $y(k)$이다. 이때 신경망은 다음의 예측오차 $e(k)$를 최소로 하도록 학습되며, 오차함수는 다음과 같이 표현된다. 즉,

$$e(k) = y(k) - y^\wedge(k), \qquad n \le k \le N \qquad (3)$$

이다. 식 (3)의 자승 $e^2(k)$을 최소화하는 학습과정은 이용되는 방법에 무관하게 망의 실제 출력이 입력으로 궤환되지 않는 개루프 적응기획이다. 다음의 그림 1은 신경망을 이용한 예측시스템의 구성을 도시한 것이다.



(a) 학습을 위한 개루프 적응기획 (b) 시험 및 예측을 위한 폐루프 적응기획

그림 1. 신경망을 이용한 예측시스템의 구성도

그림 1(a)는 신경망을 학습시키는 과정으로 모델의 파라미터들을 추출하는 전향예측을 나타내며, 이는 과거의 입력 데이터들로부터 미지의 비선형 함수의 근사화 함수를 찾는 과정이다. 또한 그림 1(b)는 학습된 신경망을 시험하고 미래값을 예측하는 과정으로 시험은 $k < N$에서 이루어지며 예측은 $k \ge N$에서 얻어질 수 있다. 여기서 신경망의 일반화 성능은 반복적 예측을 수행함으로써 얻어지며, 예측성능은 폐루프 적응기획을 통하여 평가될 수 있다. 예측은 신경망에 의해서 계산된 단단계(one-step) 예측값만을 입력에 궤환시킴으로써 계산되는 단기예측과 예측된 출력들과 한 시간단위씩 이동된 실제 입력들을 모두 이용하는 장기예측으로 나눌 수 있다. 본 실험에서는 단기예측만을 고려하였다.

한편, 예측을 위한 시스템으로 사용되는 신경망으로는 다층신경망이 주로 이용되고 있다. 이러한 다층신경망을 학습하기 위한 학습알고리즘으로는 기울기하강의 속성을 이용하는 역전파 알고리즘이 널리 이용되고 있다. 그러나 역전파 알고리즘은 학습파라미터의 설정에 따른 수렴속도와 견실성 중 하나 이상의 문제를 가지며, 일반적으로 전역최소점으로의 수렴이 보장되어 있지 않다.

3. 기울기하강과 동적터널링 속성을 조합한 학습알고리즘

다층신경망은 입력층과 출력층 간에 하나 이상의 은닉층을 가진 구조이다. 은닉층은 단층 신경망이 가지는 여러 가지 한계점을 극복하기 위해서 제안되었다. 다층신경망의 학습알고리즘으로 역전파 알고리즘이 제안되어 최근에 널리 이용되고 있다.

제안된 학습알고리즘은 기울기하강과 동적터널링에 기초를 둔 역전파 알고리즘이다. 먼저 기존의 역전파 알고리즘에 대하여 간략하게 설명한다.

i_{jp}와 y_{kp}를 각각 p번째 입력 신호에 대한 은닉층과 출력층의 j번째와 k번째 뉴런 출력으로 하자. 이 때 신경망의 입출력 관계는 다음과 같다.

$$y_{kp} = f_k(\textstyle\sum_j w_{kj}\ i_{jp})$$
$$i_{jp} = f_j(\textstyle\sum_i w_{ji}\ x_{ip}) \qquad (4)$$

여기서 $f(\cdot)$와 w_{kj}는 뉴런의 활성화함수와 연결가중치를 나타낸다. 또한 전체 오차함수 $E(\mathbf{w})$와 p번째 입력에 대한 오차함수 $E_p(\mathbf{w})$는 다음과 같이 각각 정의된다.

$$E(\mathbf{w}) = \textstyle\sum_p E_p(\mathbf{w})$$
$$E_p(\mathbf{w}) = (1/2)\textstyle\sum_k (d_{kp} - y_{kp})^2 \qquad (5)$$

식에서 d_{kp}는 출력층의 k번째 뉴런에서의 원하는 출력이다. 연결가중치 $\mathbf{w}$에 대하여 $E(\mathbf{w})$를 최소화하기 위해, 기존의 역전파 알고리즘에서는 $(\partial E_p(\mathbf{w})/\ \partial \mathbf{w})$를 계산한 후 $E_p(\mathbf{w})$의 기울기 하강 방향으로 가중치를 경신시킨다. 이때 출력층에 대한 연결가중치는

$$\mathbf{w}_{kj}(t+1) = \mathbf{w}_{kj}(t) + \eta\ \delta_{kp}\ i_{jp} + \alpha\ \Delta_p\ \mathbf{w}_{ji}(t-1)$$
$$\delta_{kp} = (y_{kp} - d_{kp})\ f_k'(\textstyle\sum_j \mathbf{w}_{kj}\ i_{jp}) \qquad (6)$$

과 같이 경신되며, 또한 은닉층의 연결가중치는

$$\mathbf{w}_{ji}(t+1) = \mathbf{w}_{ji}(t) + \eta\ \delta_{jp}\ x_{ip} + \alpha\ \Delta_p\ \mathbf{w}_{ji}(t-1)$$
$$\delta_{jp} = f_j'(\textstyle\sum_i \mathbf{w}_{ji}\ x_{ip})\ \textstyle\sum_k \delta_{kp}\ \mathbf{w}_{kj} \qquad (7)$$

과 같이 경신된다. 여기서 η는 학습율이고 α는 모멘텀(momentum)이다. 즉, 연결가중치의 경신은 위의 식 (6)과 (7)에 따라 후향으로 이루어진다.

한편, 동적터널링 알고리즘은 Yao에 의해서 제안되었다[6]. 이는 동적최적화와 동적터널링으로 구성된다. 여기서 동적최적화는 초기 상태로부터 국소최소값 중의 한 값을 찾는 과정이며, 동적터널링은 찾은 국소최소값보다 하위유역에 위치한 새로운 초기 상태를 찾는 과정이다. 즉, 동적터널링은 동적최적화가 국소최소점에 빠지는 것을 막아준다.

따라서, 신경망에서 연결가중치 $\mathbf{w}$에 대한 오차함수 $E(\mathbf{w})$를 최소화하는 문제를 살펴보자. 이 문제는 $(\partial E_p(\mathbf{w})/\partial \mathbf{w})$를 이용하여 $E(\mathbf{w})$의 국소최소값 $\mathbf{w}^*$를 찾는 것으로 변형될 수 있다. 즉, $\Delta_p\mathbf{w} = -\alpha(\partial E_p(\mathbf{w})/\partial \mathbf{w})$는 동적최적화와 동일하다. 동적터널링은 동적최적화에서 구해진 $\mathbf{w}^*$에다 결정론적으로나 임의로 설정되는 교란 벡터 ε이 추가된 새로운 시작점에서 출발하는 것으로 다음과 같은 미분 방정식의 형태로 주어진다.

$$\Delta_p\mathbf{w} = -(\partial E_p(\mathbf{w})/\partial \mathbf{w})[1/[(\mathbf{w}-\mathbf{w}^*)^T\ (\mathbf{w}-\mathbf{w}^*)]^{\gamma}$$
$$-\ \theta\ E_p'(E_p\wedge(\mathbf{w}))] \qquad (8)$$

여기서, θ는 터널링 페널티(tunneling penalty)이며, γ의 값은 $(\partial E_p(\mathbf{w})/\partial \mathbf{w})$의 원점 $\mathbf{w}^*$의 차수보다 크거나 같은 값을 가진다. 또한 $E_p\wedge(\mathbf{w}) = E_p(\mathbf{w}) - E_p(\mathbf{w}^*)$로 정의되며, $E_p'(\cdot)$는 다음과 같이 두 개의 분리된 구분선형(piecewise linear) 함수로 정의된다.

$$E'(z) = \begin{cases} z & (z>0) \\ 0 & (z<0) \end{cases} \qquad (9)$$

식 (8)에서, 우편 첫 번째 항의 분모 $[(\mathbf{w}-\mathbf{w}^*)^T\ (\mathbf{w}-\mathbf{w}^*)]^{\gamma}$항은 동적최적화로부터 구해진 국소최소점인 평형상태 $\mathbf{w}^*$를 제거하기 위한 것이며, 두 번째 항은 하위유역에 존재하는 즉, $E_p(\mathbf{w}^{(0)}) \le E_p(\mathbf{w}^*)$인 새로운 초기 상태 $\mathbf{w}^{(0)}$를 찾기 위한 것이다.

그러므로 동적터널링에 기초한 역전파 알고리즘은 식 (8)처럼 주어지고, 이때 출력층과 은닉층 각각의 $(\partial E_p(\mathbf{w})/\partial \mathbf{w})$는 다음 식과 같이 정의될 수 있다.

$$\partial E_p(\mathbf{w})/\partial \mathbf{w}_{kj} = -\ \eta\ \delta_{kp}\ i_{jp}$$
$$\partial E_p(\mathbf{w})/\partial \mathbf{w}_{ji} = -\ \eta\ \delta_{jp}\ x_{ip} \qquad (10)$$

여기서 δ_{kp}와 δ_{jp}는 각각 식 (6)과 (7)로 주어진다.

따라서 기울기하강과 동적터널링을 조합한 역전파 알고리즘인 새로운 학습알고리즘의 흐름도를 그림으로 도시하면 그림 2와 같다. 그림에서, $E_g(\mathbf{w})$와 $E_d(\mathbf{w})$는 각각 기울기하강과 동적터널링에 기초한 역전파 알고리즘을 사용할 때 계산된 전체 오차함수값이다.

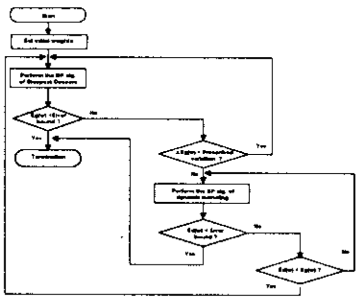


그림 2. 제안된 학습 알고리즘의 흐름도

4. 시뮬레이션 결과 및 분석

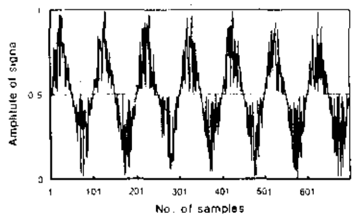
제안된 학습알고리즘을 이용한 신경망의 성능을 평가하기 위해서 3층 전향 신경망을 구성하였다. 다층신경망에서의 각 층간 뉴런 사이의 초기 연결가중치는 각각 -0.5에서 +0.5 사이의 임의의 값으로 설정하였다. 학습은 전체 반복회수가 10,000이상이거나 전체 오차값이 10^{-4}이하일 때, 또는 전체 오차함수 값의 변화가 10^{-6}이하일 때 종료되도록 하였다.

제안된 학습알고리즘의 다층신경망을 700 개의 순차적인 데이터 중에서 600 개의 데이터로 학습시킨 다음 나머지 100 개의 데이터를 예측하는 문제를 대상으로 시뮬레이션을 수행하여 그 타당성을 확인하였으며, 기존의 기울기하강법에 기초를 둔 역전파 알고리즘을 이용한 신경망에 의한 결과와 비교 고찰하였다. 여기서 신경망의 구조는 층사이에 완전한 연결을 가지는 3층 구조이며, 입력층과 은닉층 뉴런의 개수는 각각 8 개로 설정하였다.

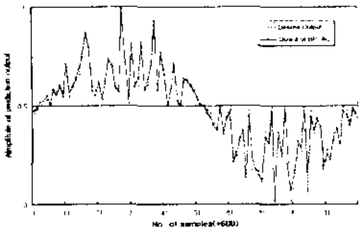
그림 3의 (a)는 700 개의 표본을 가진 입력 데이터 신호이며, 이들 중 600 개의 표본 데이터들을 신경망의 학습데이터로 이용하였다. 이때 입력 데이터들을 학습 전에 0.0에서 1.0의 범위내로 정규화시켰으며, 다층신경망의 학습률과 모멘트는 각각 0.01과 0.5로 설정하였다.

그림 3의 (b)와 (c)는 700 개의 데이터 중에서 처음 600 개의 데이터로 신경망을 학습시킨 다음 나머지 601 번째부터 700 번째까지 예측된 시뮬레이션 결과이다. 그림 3의 (b)는 기존의 기울기하강에 기초를 둔 역전파 알고리즘 이용하여 신경망을 학습시킨 결과이며, (c)는 제안된 하이브리드 알고리즘을 이용하여 신경망을 학습시킨 결과이다. 그림에서, 제안된 학습알고리즘을 이용

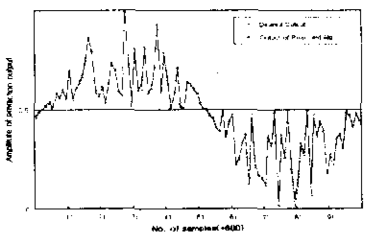
한 다층신경망에 의한 예측결과가 기존의 역전파 알고리즘에 의한 결과보다 더 우수한 성능을 가짐을 확인할 수 있다.



(a) 700 개의 표본 입력신호



(b) 기존 역전파 알고리즘에 의한 예측결과



(c) 제안된 알고리즘에 의한 예측결과

그림 3. 각 알고리즘을 이용한 예측결과

표 1은 초기 연결가중치를 100번 변화시켜가면서 학습시킨 결과, 최적해로 수렴된 경우들에 대한 반복회수 및 CPU 시간의 평균 m과 표준편차 σ를 나타낸 것이다. 이 결과는 600 개의 입력 표본들에 대해서 학습률과 모멘트를 각각 0.01과 0.5로 그리고 각 층의 뉴런수는 8로 설정하여 실험한 결과이다. 각 시도는 랜덤시드를 변화시켜 수행하였다. N_i와 t_i는 각각 기존의 역전파 알고리즘과 제안된 알고리즘에 의한 반복회수와 CPU 시간을 나타낸 것이다. 표에서 보는 것처럼, 제안된 알고리즘을 이용한 예측이 기존의 역전파 알고리즘보다 최적해로의 수렴속도와 수렴확률면에서 더욱 우수한 성능이 있음을 알 수 있다. 특히, 실험의 결과를 통해 제안된 알고리즘은 역전파 알고리즘보다 초기 가중치에 대한 의존도도 적음을 추정할 수 있다. 이는 일반적으로 입력 데이터의 무질서 정도가 높을수록 예측결과는 신경망의 초기 연결가중치에 더욱 민감한 것으로 제안된 알고리즘은 상대적으로 기존의 역전파 알고리즘보다도 무질서한 입력데이터의 예측에도 우수한 성능이 있음을 알 수 있다.

표 1. 600 개의 표본입력 데이터에 대한 100번 시도에 따른 실험 결과

	BP algorithm		Proposed algorithm	
	m	σ	m	σ
N_i	6042.9	1974.5	2745.4	1027.2
t_i	799.4	890.9	383.1	412.9
P_r	48 %		94 %	

m : mean, σ : standard deviation, N_i : No. of iterations, t_i : CPU time(sec), p_r : convergence ratio(%)

5. 결 론

본 논문에서는 하이브리드 학습알고리즘의 3층 전향신경망을 이용하여 비선형 동적 불연속 시스템에서의 효율적인 시급수 예측에 대해서 제안하였다. 제안된 하이브리드 학습알고리즘은 기울기하강과 동적터널링의 속성을 조합한 역전파 알고리즘이다. 여기서 기울기하강의 속성은 빠른 수렴을 위한 것이고, 동적터널링의 속성은 국소최적해로부터 탈출하여 전역최적해로의 수렴을 위한 것이다.

제안된 학습알고리즘을 이용한 다층신경망을 700 개의 순차적인 데이터 중에서 600 개를 학습시켜 다음의 나머지 100 개의 데이터를 예측한 결과, 기울기하강에 기초를 둔 기존의 역전파 알고리즘을 이용한 다층신경망에 의한 결과보다 우수한 수렴성능과 예측성능이 있음을 확인할 수 있었다. 또한 신경망의 초기 연결가중치에 대한 의존도도 더욱 낮아 상대적으로 무질서한 입력데이터의 예측에도 더 우수한 성능이 있음을 확인할 수 있었다.

참고 문헌

[1] A. Cichock and R. Unbehaun, *Neural Networks for Optimization and Signal Processing*, John Wiley & Sons., New York, 1993

[2] S. Haykin, *Neural Networks : A Comprehensive Foundation*, IEEE Press, New York, 1994

[3] L. H. Ungar, " Forecasting," *The Handbook of Brain Theory and Neural Networks*, MIT Press, Massachusetts, pp. 399-403, 1995

[4] T. Onoda, " Next Day Peak Load Forecasting Using an Artificial Neural Network," *International Joint Conference on Neural Networks*, Nagoya, vol. 2, pp. 2029-2032, Oct. 1993

[5] J. A. Freeman and D. M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques*, Addison Wesley, Lodon. 1991

[6] Y. Yao, "Dynamic Tunneling Algorithm for Global Optimization," *IEEE Trans. on Systems*, Man. and Cybernetics, vol. 19, no. 5, pp. 1222-1230, Sept/Oct. 1989