

## 순환신경망모형을 이용한 단기 시계열예측

윤 여 창<sup>1)</sup>

### 요 약

본 연구에서는 단순구조 순환신경망을 이용한 신경망예측과 전통적인 시계열예측 방법을 이용하여, 순환변동이 있는 시계열자료의 단기예측 오차를 비교한다. 순환신경망모형의 입력자료를 변화시키는 개선된 학습방법을 적용하여 시계열자료를 학습하고, 신경망예측의 결과는 선형 AR(9)모형, 비선형 SETAR모형 그리고 이들의 결합모형을 이용한 예측결과와 비교한다. 실증분석에 적용된 시계열자료는 1700년부터 1987년 까지의 태양흑점 자료이며 예측에 이용된 검정자료는 1980년부터 8년간의 자료이다.

### 1. 서 론

시계열자료 분석에서 중요시 되는 문제는 함수형태의 결정과 모수의 추정문제이다. 이는 신경망에서 각 층별 노드의 개수, 층의 개수, 변환함수의 형태등을 결정하는 문제와 신경망의 가중값을 결정하는 문제와 연결된다. 신경망모형을 이용한 분석 기법 및 모형화방법이 빠르게 발전하고 있지만 신경망모형은 대부분의 통계적인 예측모형들 보다 이해하기 어려운 구조로 되어 있으며, 추정해야 할 모수가 너무 많고 이로인한 과대적합 문제가 발생할 수 있다. 더우기 학습 과정에서 많은 계산시간을 요구하고 있다. 그러나 이러한 문제점에도 불구하고 컴퓨터 하드웨어의 급속한 발달로 신경망을 통계문제에 적용한 연구가 계속되고 있다. 특히 시계열예측과 신경망을 통한 예측의 비교 연구는 다음과 같다.

Sharda 와 Patil(1990)은 Makridakis 등(1982)의 75가지 시계열자료를 이용하여, 신경망모형을 이용한 분석이 Box-Jenkins만큼의 분석효율을 보이고 있음을 연구하였다. Foster 등(1991)은 연도별 자료인 경우에 신경망기법이 Holt의 지수평활법, Brown의 지수평활법 그리고 최소제곱법보다 우위임을 연구하였다(Hill 등, 1994). Sharda와 Patil(1992) 그리고 Tang 등(1991)은 추세나 계절성이 명확한 시계열자료와 정상시계열자료로 나누어 연구를 하였다. 이들은 정상시계열자료인 경우에 신경망 예측결과가 보다 효율적이며, 특히 불규칙한 시계열인 경우에 Tang 등(1991)은 신경망이 Box-Jenkins보다 예측효율이 높음을 밝혔다. Hill 등(1994)은 Makridakis 등(1982)의 111가지 시계열자료를 이용한 신경망분석과 전통적인 시계열분석을 통하여, Sharda와 Patil(1992)의 결과보다 더 신경망 예측의 효율이 높음을 보였다. 윤여창과 허문열(1997)은 단순신경망모형(simple neural network)을 이용할 때, 초기값과 은닉노드 개수의 변화를 실시간으로 적용시킬 수 있는 개선된 학습방법을 이용하여 신경망과 Box-Jenkins 예측결과를 비교하였다.

본 연구에서는 단순구조 순환신경망(simple recurrent model)을 이용하여 순환변동이 있는 시계열자료의 신경망분석을 통한 단기예측을 하고, 전통적인 Box-Jenkins방법을 이용한 시계열 예측결과와 비교하고자 한다. 특히 신경망의 여러 문제들 중에서 입력노드와 은닉노드 개수를 최대한

1) (565-701) 전북 완주군 삼례읍 후정리 우석대학교 전산통계학과 조교수.

제한시키면서, 추정해야 할 모수를 줄이고 계산시간을 절약할 수 있도록 개선시킨다. 또한 순환변동이 있는 시계열자료의 특성을 반영시키기 위하여 순환신경망모형의 입력자료를 변화시키는 학습방법을 적용하여 시계열자료를 학습하고, 이를 Tong(1990)의 비선형 시계열 예측결과와 비교한다. Tong(1990)은 1700년부터 1987년 까지의 태양흑점자료를 선형 AR(9)모형, 비선형 SETAR모형 그리고 이들의 결합모형을 이용하여 각각 예측오차를 비교하였다.

## 2. 순환신경망의 구조와 학습

### 2.1 순환신경망의 구조

은닉층과 출력층으로의 입력노드를 각각 2개,  $H$ 개 씩으로 제한한 단순 순환신경망  $2 \times H \times 1$  구조는 (그림 1)과 같다. (그림 1)에서 은닉층의 각 노드는  $y_l = g\left(\sum_{i=0}^2 a_{li}x_i\right)$ ,  $l=1, 2, \dots, H$ 이고, 출력노드는  $z = g\left(\sum_{j=0}^H b_j y_j\right)$ 이다. 여기서  $i=0$ ,  $j=0$ 인 경우에는 각각  $x_0=1$ ,  $y_0=1$ 이다. 그리고 변환함수는 다음 식(1)과 같다.

$$g(u) = [1 + \exp(-u)]^{-1}, \quad 0 < g(u) < 1. \quad (1)$$

### 2.2 학습과정

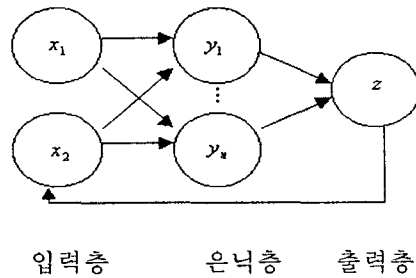
Smith(1993)에서 학습과정은 변환함수를 통과한 값과 목표값 사이에 가장 잘 적합되는 가중값을 찾는 과정이며, 올바른 가중값을 어떻게 찾아야 하는가가 학습의 중요한 문제라고 하였다. 이때  $n$ 번째 예제에 대하여 출력값 ( $z_n$ )과 목표값 ( $t_n$ )간의 학습오차는 다음과 같은 오차판단기준을 이용한다.

$$MSE = \frac{1}{2N} \sum_{n=0}^N (z_n - t_n)^2. \quad (2)$$

여기서  $N$ 은 예제들의 전체 개수이다.

본 연구에서 적용한 역전파 알고리즘을 이용한 학습과정의 개요는 <표 1>과 같다.

<표 1>에서 처리절차 `weight_initialization`은 제어상자를 이용한 윤여창과 허문열(1997)의 동적인 초기값 결정과정이다. 처리절차 `forward`는 각 예제에 대하여 변환함수를 현재 가중값에 적용하여 계산한 출력값을 구한다. 처리절차 `back`에서는 적응적 학습률(`adaptive learning rate`)을 이용하여 가중값의 변화량을 결정한다. 이때 오차의 판단기준은 식(2)를 이용하고 역전파 알고리즘을 이용한 학습과정은 Smith(1993)와 같다. 처리절차 `changeweight`에서는 각 예제에 대한 가중값들을 누적시킨 결과를 이용하여 현재 가중값들을 각각 갱신시킨다. 이와같은 과정을 반복하면서 오차가 최소화 되는 가중값들을 구하기 위하여, 각 노드에 대한 가중값의 변화가 한계값 이하가 될 때까지 반복 학습을 한다.



(그림 1) 단순 순환신경망 구조.

<표 1> 역전과 알고리즘을 이용한 개선된 학습과정.

```

weight_initialization()
repeat
{
    for(n in 1:examples)
    {
        forward()
        back()
    }
    changeweight()
}
    
```

본 연구에서 적용한 학습과정의 문제를 단순화 시키기 위하여 은닉노드에 대한 가중값  $a_{ij}$ 와 출력노드에 대한 가중값  $b_j$ 를 가중값  $w$ 라고 하면,  $m$ 번째 학습 후의 변화된 가중값은 다음 식(3)과 같다.

$$w_m = w_{m-1} + c_m \tag{3}$$

$c_m$ 은  $m$ 번째 학습 후의 가중값  $w$ 의 변화량이며 다음 식(4)와 같이 적응적 학습률을 이용하여 구한다.

$$c_m = \mu c_{m-1} - (1 - \mu) e_m d_m, \quad 0 \leq \mu < 1. \tag{4}$$

여기서  $e$ 는 가중값들이 변화됨에 따라 최소값으로 수렴되어 가는 비율을 조절하는 모수이며 다음 식(5)와 같다.

$$e_m = \begin{cases} e_{m-1} + x, & d_m f_m > 0 \\ e_{m-1} \times \phi, & d_m f_m \leq 0 \end{cases}, \quad 0 < \phi < 1. \tag{5}$$

$d_m$ 은 모든 예제들에 대하여 MSE를 각 학습량에서의 가중값으로 미분한 값들의 누적합이다. 따라서  $m$ 번째 학습 후의 누적 가중값들은 다음 식(6)과 같다.

$$d_m = \sum_{n=1}^N \left( \frac{\partial E}{\partial w_m} \right)_n \tag{6}$$

$f_m$ 은 누적합이 최근 변화에 더 큰 영향을 받도록 하기 위한 항으로서 다음 식(7)과 같다.

$$f_{m+1} = \theta f_m + (1 - \theta) d_m, \quad 0 \leq \theta < 1. \quad (7)$$

여기서  $x, \phi, \theta$  그리고  $\mu$ 는 학습률을 조정하기 위한 모수들으로써, Smith(1993)의 경험적인 값들인  $x=0.1, \phi=0.5, \theta=0.7$  그리고  $\mu=0.9$ 이다.

$m$  번째의 학습이 끝난 후에, 주어진 허용오차를 만족할 때까지 <표 1>의 역전파 알고리즘은 계속 반복된다.

### 3. 실증분석

이 장에서는 1700년부터 1987년 사이의 288년 동안 관측된 Wolf의 태양흑점자료를 다음 식(8)과 같이 변환 후에 분석한다.

$$Z = \frac{data - data_{(1)}}{data_{(t)} - data_{(1)}}, \quad t = 1, 2, \dots, 288. \quad 0 \leq Z \leq 1. \quad (8)$$

여기서  $data$ 는 태양흑점자료,  $data_{(1)}$ 은 최소값 그리고  $data_{(t)}$ 는 최대값이다.

태양흑점자료는 상승, 하강 추세나 확실한 계절성을 관찰할 수 없지만 9년에서 14년의 주기를 가지고 순환하며 평균주기는 11년인 순환변동을 가지고 있는 시계열이다. 이 자료에 대한 Tong(1990)의 결과는 8년간의 비교적 단기 예측결과이며 이를 순환신경망을 이용한 예측결과와 비교해 본다.

Tong(1990)은 태양흑점자료를 Subba와 Gabr(1984)의 AR(9) 모형, Tong과 Lim(1980)의 SETAR 모형 그리고 이들의 결합모형으로 1980년부터 1987년 까지의 예측을 비교하였다. 순환신경망을 이용한 학습은 1700년부터 1979년 까지의 280년 간의 자료를 학습자료로 이용하여 각각 200회 학습을 하였다. 예측을 위한 검정자료는 8년 간의 자료로써 1980년 부터 1987년 사이의 자료이다. 신경망 예측은 2개 씩으로 제한한  $Z_t$ 와  $Z_{t-k}$ 를 입력자료로 이용하여  $Z_{t+1}$ 을 예측한다. 여기서 태양흑점자료의 순환변동을 고려하기 위해 12가지의  $k$ 에 대하여 분석한다.

은닉노드의 개수를 결정하기 위하여 본 연구에서는 Baum과 Haussler(1988)의 규칙을 적용하였다. Tong(1990)의 결합모형 결과에서 MSE가 0.02보다 작지 않으므로 신경망 학습의 MSE 수렴 정도가 0.02, 0.03인 경우에는 주어진 조건에서 은닉노드의 개수가 각각 2와 3이다. 따라서  $2 \times H \times 1$ 이라는 간단한 구조 하에서 은닉노드의 개수는 2와 3의 변화만을 제시한다.

<표 2>는 Subba와 Gabr(1984)의 AR(9) 모형, Tong과 Lim(1980)의 SETAR 모형 그리고 이들의 결합모형으로 1980년부터 1987년 까지의 예측을 비교한 Tong(1990)의 결과를 신경망 결과와 비교하기 위하여 식(8)의 역변환으로 재 구성한 표이다. <표 3>은 은닉노드 개수가 2개인 순환신경망에서 200회 학습 결과를 이용한 예측오차로서,  $k$ 를 1, 2, ..., 12로 구분하였을 때의 신경망 예측오차를 각각 Tong(1990)의 결합모형과 비교한 결과이다. <표 4>는 은닉노드 개수가 3개인 순환신경망인 경우의 비교 결과이다.

신경망의 예측결과만 살펴볼 때,  $k$ 가 5, 6, 7 그리고 10, 11, 12인 경우는 단기예측에서  $k$ 가 1, 2, 3, 4, 8, 9인 결과보다 예측오차가 작다. 이 결과는 1/2, 1주기의 순환변동주기를 반영한 결과가 그렇지 않은 경우보다 예측오차가 더 작음을 보여주고 있다.  $k$ 가 1, 2, 3, 4인 경우에 200회의 학습으로는 충분한 결과를 보이고 있지 않고 예측결과가 발산되고 있다. 따라서 시계열자료에서 순환주기가 있는 경우에 신경망 입력자료는 의미가 있음을 알 수 있다. 은닉노드 개수가 3인 경우는

$k$ 가 4, 8, 9, 10, 12인 경우에서 은닉노드 개수가 2인 경우보다 예측오차가 작다. 신경망의 예측결과를 통계적 예측결과와 비교할 때, 신경망은  $k$ 가 5, 6, 7, 10, 11, 12인 경우에 Subba와 Gabr(1984)의 AR(9) 모형, Tong과 Lim(1980)의 SETAR 모형, Tong(1990)의 결합모형보다 단기 예측오차가 작다. 그러나  $k$ 가 1, 2, 3, 4, 8, 9인 경우에는 AR(9) 모형, SETAR 모형 그리고 결합 모형의 단기 예측오차가 더 작다.

#### 4. 결론

본 연구에서는 단순구조 순환신경망의 입력노드 개수를 2개로 제한시키면서 신경망 입력자료의 변화를 통한 다양한 학습결과와, Subba와 Gabr(1984)의 AR(9)모형, Tong과 Lim(1980)의 SETAR 모형, Tong(1990)의 결합모형을 이용한 태양흑점자료의 단기 예측결과를 비교하였다. 자료의 순환변동을 반영하기 위하여 신경망 입력자료는 12가지의  $k$ 에 대하여 각각 200회 학습한 결과를 이용하였다. 예측은 은닉노드 개수가 2개, 3개인 경우로 한정하였다.

순환신경망의 예측결과는 단기예측에서 1/2, 1주기의 순환변동주기를 반영한 결과가 그렇지 않은 경우보다 예측오차가 더 작았다. 이는 시계열자료에서 순환주기가 있는 경우에 단순 순환신경망의 입력자료에 따라 분석결과의 차이가 커짐을 의미한다. 순환변동주기를 반영한 순환신경망의 예측결과는 통계적 예측결과보다 단기 예측오차가 작았다. 그러나 순환변동주기를 반영하지 않으면 AR(9) 모형, SETAR 모형 그리고 결합모형의 단기 예측오차가 더 작았다.

제시된 결과는 Tong(1990)의 비교와 순환신경망모형을 이용한 단기 시계열예측 결과이다. 더우기 제한된 비선형모형과의 비교만이 제시되어 있다. 따라서 다양한 비선형모형들과 이들의 장기예측 결과와도 비교해 볼 필요가 있다.

<표 2> Tong(1990)의 결과 :  $k=11$ 인 경우의 신경망 예측오차와 비교하여 작은 오차를 갖는 경우에 †로 표시한다.

년도	AR(9) 모형	SETAR 모형	결합 모형
1980	.04257	.10515	.04257
1981	.03680	.03680	.03680
1982	.12829	.12355	.12829
1983	.06993	.01945 †	.06993 †
1984	.11987	.11462 †	.11462 †
1985	.06993 †	.02787 †	.02787 †
1986	.05994 †	.06940 †	.06940 †
1987	.04257 †	.09674 †	.09674 †

<표 3> 은닉노드 개수가 2인 순환신경망에서 200회 학습 결과를 이용한 예측오차 :  $k$ 를 1,2,...,12로 구분하였을 때의 신경망 예측오차를 각각 Tong(1990)의 결합모형과 비교하여 작은 오차를 갖는 경우에 †로 표시한다.

년도 $k$	1980	1981	1982	1983	1984	1985	1986	1987
1	.11293	.16735	.48878	.42434	.50229	.46248	.44783	.45695
2	.11697	.13022	.16409	.42273	.37817	.35880	.35360	.37815
3	.08691	.07340	.16049	.19512	.35622	.32530	.30117	.29129
4	.05487	.03908	.14018	.18022	.19168	.27731	.26120	.25227
5	.02778 †	.02476 †	.03597 †	.03528 †	.03161 †	.03320	.09933	.11649
6	.03049 †	.02227 †	.02211 †	.03735 †	.06581 †	.10470	.11008	.10302
7	.01421 †	.04645	.05160 †	.04771 †	.06637 †	.11192	.16907	.18228
8	.01489 †	.10375	.13722	.12276	.11073 †	.11832	.16294	.18117
9	.01640 †	.13206	.18954	.17919	.16134	.15056	.16690	.17365
10	.01005 †	.05941	.10603 †	.11213	.10673 †	.09789	.09262	.08778 †
11	.00279 †	.02189 †	.07035 †	.10673	.11544	.11473	.10685	.09994
12	.02382 †	.03033 †	.03489 †	.04289 †	.08884 †	.12068	.13111	.13239

<표 4> <표 3>에서 은닉노드 개수가 3인 경우.

년도 $k$	1980	1981	1982	1983	1984	1985	1986	1987
1	.11128	.16566	.48715	.42303	.50930	.46126	.44658	.45590
2	.10716	.11744	.13663	.41049	.36793	.33975	.33635	.35554
3	.10511	.08391	.20414	.23616	.37785	.34509	.31949	.30523
4	.04679	.03309 †	.13943	.18111	.19018	.27278	.25795	.24938
5	.04734	.08005	.10678 †	.11018	.10621 †	.10407	.13916	.13122
6	.02313 †	.03394 †	.02825 †	.03772 †	.06658 †	.11471	.16712	.15649
7	.01849 †	.05123	.05588 †	.06723 †	.10177 †	.15088	.20666	.21679
8	.02691 †	.10960	.12535 †	.10957	.11232 †	.12619	.15220	.16335
9	.02965 †	.10664	.15760	.14998	.13486	.12596	.13092	.12670
10	.01463 †	.05618	.09454 †	.09803	.09314 †	.08574	.08129	.07764 †
11	.00018 †	.02903 †	.07938 †	.11889	.12661	.12401	.11541	.10798
12	.00976 †	.08023	.08909 †	.13737	.11408 †	.11931	.11302	.10630

### 참 고 문 헌

- [1] 윤여창, 허문열. (1997). 신경망이론에 의한 시계열자료의 분석, 「한국통계학회논문집」, 제4권 1호, 91-99.
- [2] Baum, E.B. and Haussler, D. (1988). What Size Net gives Valid Generalization? *Neural Computation*, Vol. 1, 151-160.
- [3] Hill, T. et al. (1994) Artificial Neural Network Models for Forecasting and Decision

- Making, *International Journal of Forecasting*, Vol. 10, 5-15.
- [4] Makridakis, S. et al. (1982). The Accuracy of Extrapolation time series Methods : Results of a Forecasting Competition, *Journal of Forecasting*, Vol. 1, 111-153.
- [5] Sharda, R. and Patil, R. (1990) Neural Networks as Forecasting Experts : An Empirical Test, *Proceedings of the 1990 International Joint Conference on Neural Networks Meeting*, Vol. 2, 491-494.
- [6] Sharda, R. and Patil, R. (1992) Connectionist Approach to Time Series Prediction : An Empirical Test, *Journal of Intelligent Manufacturing*, Vol. 3, 317-323.
- [7] Smith, M. (1993). *Neural Networks for Statistical Modeling*, New York, Van Nostrand Reinhold.
- [8] Subba, R.T. and Gabr, M.M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics, 24. Springer.
- [9] Tang, Z., Almeida, C. and Fishwick, P.A. (1991). Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology, *Simulation*, Vol. 57(5), 303-310.
- [10] Tong, H. (1990) *Non-linear Time Series : A Dynamical System Approach*, Clarendon Press Oxford.
- [11] Tong, H. and Lim, K.S. (1980) Threshold autoregression, Limit Cycles and Cyclical Data, *Journal of Royal Statistical Society, B*, Vol. 42, 245-292.