

ARMA 데이터에 대한 Back-propagation 신경망의 구조 A Study on Construction of Back-propagation Architecture for ARMA data

김 나영¹⁾ · 김 희영²⁾

요 약

시계열 자료를 분석할 때 쉽게 접근하는 통계적 방법은 ARMA 모형이며 신경망 학습 방법 중에서는 다층 퍼셉트론에서의 Back-propagation 알고리즘이 일반적이다. Back-propagation을 비롯한 신경망 학습의 구조는 자료의 특성에 따라 경험적으로 결정하는 것으로 알려져 있다. 그러나 바로 이 점이 신경망 학습방법의 이용을 어렵게 하는 요인이기도 하다. 본 연구는 ARMA 모형 중 몇 개 유형의 자료에 대하여 Back-propagation 알고리즘을 적용함에 있어 어떠한 구조로 학습하는 것이 효율적인가를 입력층과 은닉층의 크기, 활성화 함수를 중심으로 검토하였다.

주요 용어 : 신경망, 시계열분석

1. 서론

인간의 두뇌에서는 뉴런들의 상호연결을 통하여 학습이 이루어지며 신경망(neural network)은 이러한 계산방식의 구조를 흉내내는 지능적인 시스템의 일종이다. 신경망 이론은 음성·영상·문자인식 등 패턴인식 분야에서 활발하게 응용되고 있는데, 이는 인간의 뇌가 과거의 경험을 학습함으로써 정확한 연상작용으로 패턴을 구별 또는 분류하기 때문이다. 시계열 자료에 대해서도 신경망 이론은 주로 분류(classification)와 군집화(clustering)에 중점을 두고 있다. 그러나 시계열 자료 분석에서의 관건은 미래에 대한 예측일 것이다. 주가나 금리, 환율, 매출액 등과 같은 단변량 시계열 자료에 대해서 Box-Jenkins의 ARMA(또는 ARIMA)모형은 중·장기 예측에 비효율적인 것으로 알려져 있어서 신경망 학습을 하나의 대안으로 생각해 볼 만 하다. 시계열 자료에 대한 신경망의 적용은 다층 퍼셉트론(multi layer perceptron)에 학습방법으로 Back-propagation(이하 BP라 함) 알고리즘을 이용하는 것이 일반적이다.

시계열 자료의 BP 알고리즘에 관한 기존의 연구는 대부분 실제 자료를 대상으로 Box-Jenkins 모형과 비교하거나 몇 가지 구조(architecture)들 중 어느 것이 주어진 자료를 잘 설명하는지를 평균제곱오차(MSE) 등을 기준으로 비교하는 연구로 구분할 수 있다. Sharda와 Patil(1990)은 75가지 시계열에 대하여 비교한 결과 BP 알고리즘이 Box-Jenkins 모형보다 효율적이었다고 한 바 있다. Tang 등(1991)은 자료가 많을 경우 두 방법 모두 좋지만 단기 예측에는 Box-Jenkins의 모형이, 장기예측 또는 데이터가 적을 때는 신경망 모형을 이용하는 편이 좋

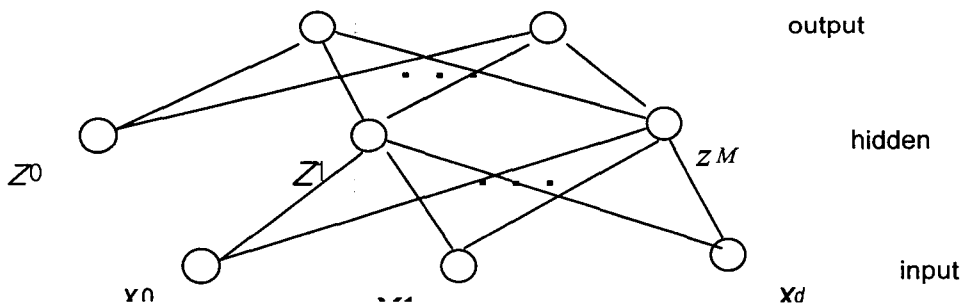
1), 2) (136-701) 서울특별시 성북구 안암동 5가 1번지 고려대학교 통계학과 박사과정

다고 하였다. Weigend와 Rumelhart(1991)는 1700년부터 1920년까지의 연간 흑점(sunspot) 자료를 training set으로 하여 학습시킨 후 1921년~1955년의 자료로 test한 결과 12개의 입력층 뉴런과 8개의 은닉층 뉴런, 1개의 출력층 뉴런으로 설명하는 12-8-1 구조가 가장 효율적이라고 하였다. Tang 등(1991)은 3가지의 월별 시계열 자료, 즉, 1949년~1960년의 항공기 승객 수, 1966년~1982년의 국내 및 국외에서의 자동차 판매량을 입력층과 은닉층, 출력층에 각각 1/6/12/24개씩의 뉴런이 존재하는 구조들로 비교 분석하였다. 이 때 입력층과 은닉층의 뉴런 수는 동일하게 정하였으며 출력층은 단기, 중기, 장기 예측의 상황에 맞도록 조절하게 된다. Kim(1999)은 주가, 금리, 환율 등의 재무예측에 신경망 모형을 이용하였다. 특히 1960년~1992년 미국의 10년 만기 국고채(US Treasury notes) 월별 수익률에 대하여 2-3-1, 3-3-1, 4-3-1 중에서 3-3-1 구조가 가장 효율적임을 보였다.

그러나 위의 연구들은 주어진 자료에 관한 것이었으며 신경망의 보편적인 구조에 대해서는 언급하지 않고 있다. 신경망의 구조는 자료의 성격에 맞게 경험적으로 결정한다고 알려져 있으나 자료의 성격을 아는 경우는 많지 않으며, 모든 가능한 구조를 고려하여 비교한다는 것은 사실상 불가능하다. 본 연구는 일반적으로 고려할만한 신경망 구조에 관한 대략적인 형태를 제시하는데 목적이 있다. 2장은 BP 알고리즘과 고려해야 할 사항들을 소개한다. 3장에서 몇 가지 시계열 모형에서 생성된 자료를 대상으로 한 모의실험과 그 결과를 제시한다. 4장에서는 KOSPI-200 지수로 3장의 결과를 검토한다.

2. Back-propagation 알고리즘

본 연구에서는 외부 입력을 받아들이는 입력층(input layer), 처리된 결과가 출력되는 출력층(output layer), 입력층과 출력층 사이에 위치하여 외부로 나타나지 않는 은닉층(hidden layer)의 3계층으로 구성된 신경망을 고려한다. 그 구조는 다음 그림과 같다.



<그림 2.1> 다층 신경망의 구조

오류 역전파 학습 알고리즘이라고도 불리는 BP 알고리즘은 <그림 2.1>과 같은 순방향 다층 신경망의 학습에 효과적이며 크게 세 부분으로 이루어진다. 먼저 입력 패턴을 신경망에 입력하여 출력을 구한 후 출력과 목표치의 차이 즉, 오차를 구한다. 마지막으로 오차 값을 역방향으로 전파시키면서 출력층의 연결강도 및 은닉층의 연결강도를 변경한다. 이를 구체적으로 나타내면 다음과 같다.

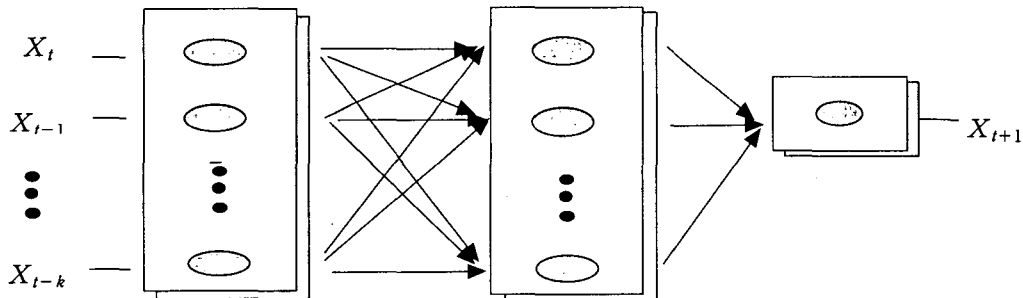
- 단계1: 연결강도와 counter를 초기화한다.
 - $v, w \leftarrow$ 임의의 작은 값
 - $p \leftarrow$ 학습 패턴 쌍의 수 $(x_1, y_1), \dots, (x_p, y_p)$
 - $k \leftarrow 1$ counter, $E \leftarrow 0$ 오차
- 단계2: 학습률($\alpha > 0$)과 오차범위(E_{max})를 결정한다.
- 단계3: 각각의 학습 쌍 (x, y) 에 대하여
단계4에서 단계8까지를 $k=p$ 가 될 때까지 반복한다.
- 단계4:
 - $NET_z = x_k v_k^t \leftarrow$ 은닉층의 가중합
 - $z = f(NEZ_z)$ ← 은닉층의 출력값
 - $$= \begin{cases} \frac{1}{1 + \exp(-NET_z)} : \text{unipolar sigmoid} \\ \frac{1 - \exp(-NET_z)}{1 + \exp(-NET_z)} : \text{bipolar sigmoid} \end{cases}$$
 - $NET_y = z w^t \leftarrow$ 출력층의 가중합
 - $y = f(NEZ_y)$ ← 최종 출력값
 - $$= \begin{cases} \frac{1}{1 + \exp(-NET_y)} : \text{unipolar sigmoid} \\ \frac{1 - \exp(-NET_y)}{1 + \exp(-NET_y)} : \text{bipolar sigmoid} \end{cases}$$
- 단계5: 제곱오차 E 를 계산한다.
 - $$E \leftarrow \frac{1}{2} (d_k - y_k)^2 + E$$
- 단계6: 오차신호를 계산한다.
 - $$\delta_y = \begin{cases} (d - y)y(1 - y) : \text{unipolar sigmoid} \\ \frac{1}{2}(d - y)(1 - y^2) : \text{bipolar sigmoid} \end{cases}$$
 - $$\delta_z = \begin{cases} z(1 - z) \sum_{i=1}^m \delta_y w_{i,z} : \text{unipolar sigmoid} \\ \frac{1}{2}(1 - z^2) \sum_{i=1}^m \delta_y w_{i,z} : \text{bipolar sigmoid} \end{cases}$$
- 단계7: 연결강도를 update한다.
 - $$w^{k+1} = w^k + \Delta w^k$$
 - $$= w^k + \alpha \delta_y z^k$$
 - $$v^{k+1} = v^k + \Delta v^k$$
 - $$= v^k + \alpha \delta_z x^k$$
- 단계8: $k \leftarrow k + 1$ 한 후, 단계3으로 간다.
- 단계9: $E < E_{max}$ 이면 종료하고, 그렇지 않으면 $E \leftarrow 0$ 을 대입한 후 단계3으로 간다.

BP 학습의 구조를 결정할 때 고려할 사항으로는 은닉층의 수, 입력층 · 은닉층 · 출력층 뉴런의 수, 학습률(learning rate), 활성화 함수(activation function) 등이 있다. 은닉층의 수가 많아지면 연결강도가 많아져서 신경망 학습이 매우 느려지게 되는데, 특수한 경우를 제외하고는 은닉층을 1개 또는 2개로 하여도 대부분의 응용에 적합한 것으로 알려져 있다. 또한, 은닉층을 2

개로 하면 1개로 할 때에 비하여 뉴런 수가 많아지므로 신경망의 성능이 개선될 것처럼 여겨지지만 실제로는 오히려 학습이 되지 않는 경우도 있으며 많은 연산이 소요되므로 처리 또는 반응시간이 느려지는 단점이 있다(Zurada(1992)). 따라서 본 연구에서는 은닉층이 1개인 경우만을 고려하였다.

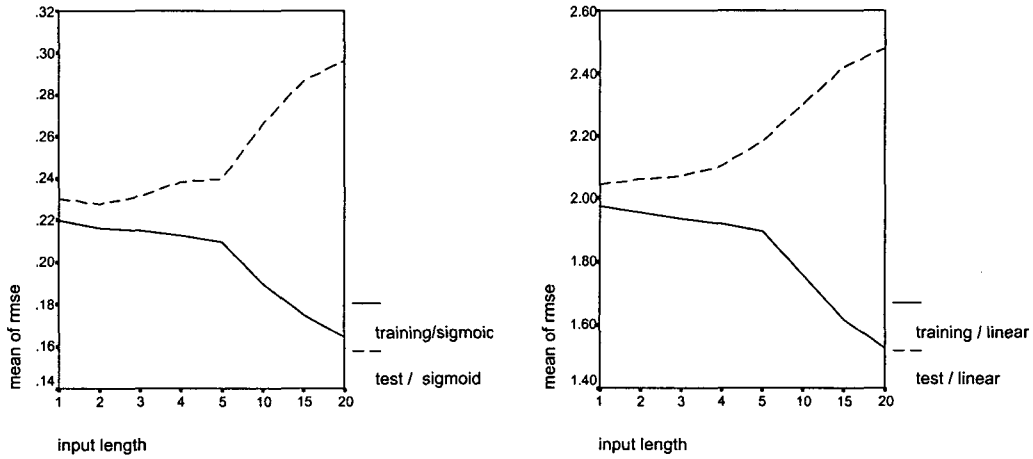
3. 모의실험 및 결과

본 연구에서는 일반적인 시계열 자료에 적용할 수 있는 신경망 구조를 알아보기 위하여 몇 가지 ARMA 모형에서 데이터를 생성하여 모의실험에 이용하였다. AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,1), ARMA(2,2)로부터 각각 1000개의 데이터를 생성하여 앞의 80%를 training set으로, 20%를 test set으로 사용하였다. 다음 그림에서와 같이 과거 몇 개 시점의 자료 값이 입력되어 바로 다음 시점의 자료 값이 출력되도록 한다. 이 때 모의실험을 통하여 입력층 뉴런의 수(input length: k)와 은닉층의 뉴런 수를 결정하고자 한다. 학습률은 각각 0.1, 300번 반복(epoch)하였으며 활성화 함수는 입력층→은닉층에는 양극성 시그모이드를, 은닉층→출력층은 양극성 시그모이드 함수와 선형 항등함수(identity function)를 각각 고려하였다. 활성화 함수로 양극성 시그모이드 함수를 사용하는 것이 일반적이지만 시계열 자료의 경우 입력층과 출력층이 동일한 scale을 갖는 연속형 값으로 이루어져 있기 때문에 은닉층→출력층에 선형 항등함수를 적용할 수 있다. 또한, 은닉층→출력층에 시그모이드 함수를 고려할 때는 전체 자료를 (-1, 1)로 scaling하여 학습시켰다. 효율성 비교의 기준으로는 평균제곱오차의 제곱근(Root Mean Squared Error : RMSE)을 이용하였다. 입력층 뉴런의 수 $k=1, 2, 3, 4, 5, 10, 15, 20$ 으로 하고 각각의 경우에 대하여 은닉층 뉴런의 수는 1~10, 15, 20을 고려하였다.

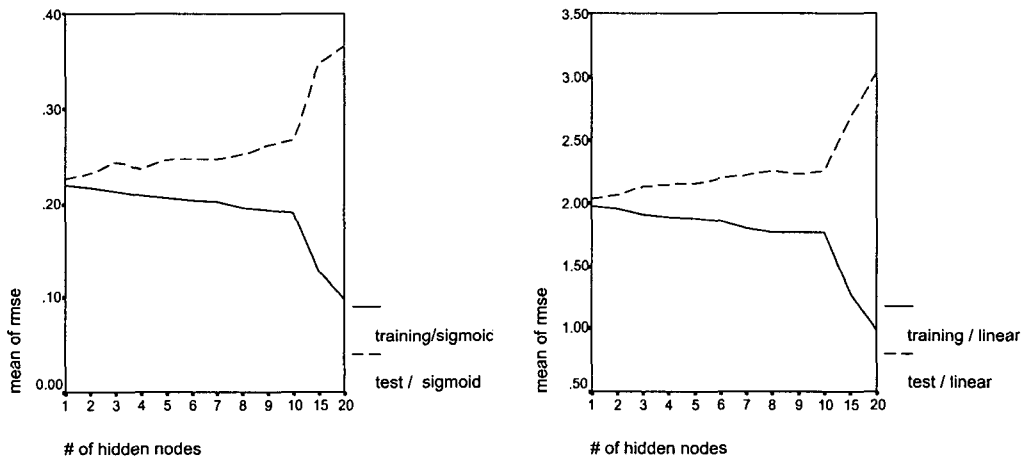


AR(1)에서 생성된 자료의 학습 결과는 다음 그림과 같다. <그림 3.1>은 입력층의 뉴런 수와 RMSE의 관계를, <그림 3.2>는 은닉층의 뉴런 수와 RMSE의 관계를 나타낸다. 활성화 함수에 관계없이 입력층 뉴런의 수가 많아질수록, 그리고 은닉층 뉴런의 수가 많아질수록 training set에 대한 RMSE는 감소하는 반면 test set에 대한 RMSE는 증가하는 경향을 보이고 있다. 입력층과 은닉층의 뉴런이 10개 이상일 때 test data에 대한 RMSE가 급격히 증가하는 것으로 보아 training data를 과대적합(overfitting)하고 있음을 알 수 있다. 또한 은닉층 뉴런의 수를 고정시켰을 때 입력층 뉴런의 수와 RMSE의 관계, 그리고 입력층 뉴런 수를 고정시켰을 때 은닉층 뉴런 수와 RMSE와의 관계도 동일한 형태를 따르고 있음을 알 수 있었다. 나머지 7개 모형에

대한 모의실험 결과도 AR(1)과 크게 다르지 않았다.



<그림 3.1> 입력층 뉴런 수에 따른 평균 RMSE

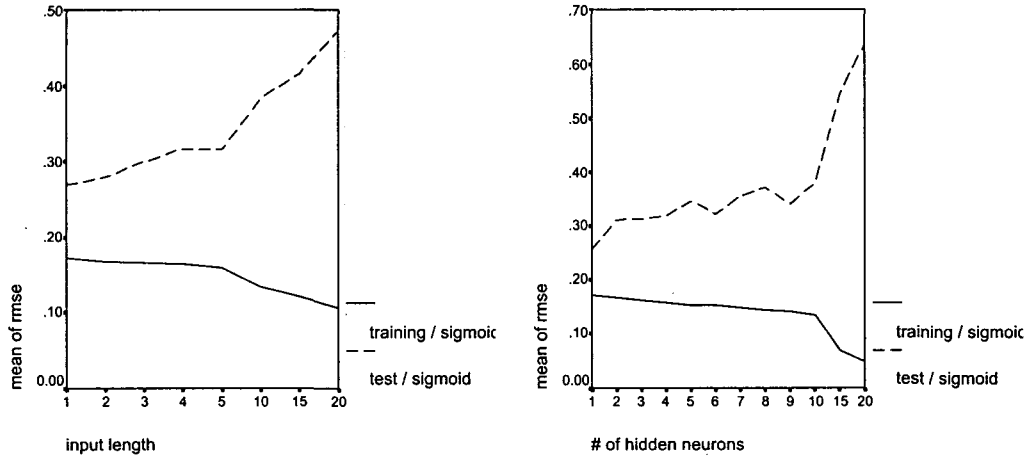


<그림 3.2> 은닉층 뉴런 수에 따른 평균 RMSE

4. KOSPI-200 지수에 대한 검토와 결론

3장의 모의실험을 통하여 BP 신경망의 구조를 결정할 때 입력층이나 은닉층에 뉴런을 반드시 많이 배치할 필요는 없다는 것을 알 수 있었다. 실제 자료의 분석에 있어서 자료의 양과 특성을 고려하여 뉴런의 수를 정하는 것이 좋겠지만 분석의 초기 단계에서 본 연구의 결과를 참고할 수 있을 것이다. 실제 자료에 대한 적용 가능성을 검토하기 위하여 1998년 1월 3일부터 2000년 8월 20일까지 총 697개의 일별 KOSPI-200 지수에 대하여 3장과 동일한 학습을 진행하였다. 이 자료는 비정상성(nonstationary) 계열이므로 1차 차분한 결과를 입·출력 자료로 사용하였다. 이러한 변수변환은 신경망 응용분야에서 전처리(preprocessing)과정에 해당하며 학습효과를 높이는 데 도움이 된다고 알려져 있다. 다음 그림은 양극성 시그모이드 함수를 사용하여

KOSPI-200 지수 자료에서 실험한 결과이다. 3장에서와 비슷한 형태를 보임을 알 수 있었다.



<그림 4.1> KOSPI-200 지수의 학습 결과

따라서 주어진 자료가 전처리 과정을 거친 시계열이라면 BP 구조를 결정할 때 입력층과 은닉층의 뉴런을 10개 이하로 잡아도 큰 무리는 없을 것으로 보인다.

참고문헌

- [1] 김대수 (1999). 『신경망 이론과 응용(1)』, 하이테크정보.
- [2] Steven H. Kim (1999). 『Data Mining in Finance』, Sigma Consulting Group.
- [3] Jacek M. Zurada (1992). 『Introduction to Artificial Neural Systems』, Info Access Distribution Pte Ltd.
- [4] Charytoniuk, W. and Chen, M. S. (2000). *Very Short-Term Load Forecasting Using Artificial Neural Networks*, IEEE, 15(1).
- [5] Choueki, M. H., Campbell, C. A. and Ahalt, S. C. (1997). *Building A Quasi Optimal Neural Network to solve the short-term load forecasting problem*, IEEE, 12(4).
- [6] Sharda, R. and Patil, R. B. (1990). *Neural Networks as Forecasting Experts: An Empirical Test*, Proceedings of the IJCNN Meeting, Washington, 491-494.
- [7] Tang, Z. Almeida, C. and Fishwick, P. A. (1991). *Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology*, Simulation, 57(5), 303-310.
- [8] Weigned, A. S. and Rumelhart D. E. (1991). *The Effective Dimension of the Space of Hidden Units*. IEEE(1991).