신경망을 이용한 비선형 시계열 자료의 예측

김인규*

Prediction for Nonlinear Time Series Data using Neural Network

Inkyu Kim*

요 약 본 논문에서는 분산이 각각 다른 이분산성을 갖는 비선형 시계열 자료를 가지고, 비선형 시계열 모형중 1차 일반화 확률계수 자기회귀모형(GRCA(1))과 자료의 형태에 상관없이 적용할 수 있는 신경망 모형을 이용하여 예측을 해서 어느 모형이 최소 평균예측오차제곱의 기준에서 비선형 시계열 자료의 예측에 적합한지를 비교 분석 하는 것이다. 조건부 이분산 모형에 따르는 자료로 확인된 종합주가지수 변동율에 대한 사례 분석 결과를 보면 신경망 모형은 단기 예측에서 좋은 예측 결과를 보였고, 비선형 모형인 GRCA(1) 모형은 장기 예측에서 좋은 예측 결과를 보여 주었다.

주제어 : GRCA(1) 모형, 신경망 모형, 비선형 시계열 자료

Abstract We have compared and predicted for non-linear time series data which are real data having different variences using GRCA(1) model and neural network method. In particular, using Korea Composite Stock Price Index rate, mean square errors of prediction are obtained in genaralized random coefficient autoregressive model and neural network method. Neural network method prove to be better in short-term forecasting, however GRCA(1) model perform well in long-term forecasting.

Key Words : GRCA(1) model, Neural network, non-linear time series

1. 서론

시계열 자료란 시간이 흐름에 따라 변하는 현상을 관찰함으로써 얻어지는 일련의 자료를 말한다. 선형 시계열 모형들은 실제 시계열 자료들에 아주 잘 적합 될 뿐만 아니라 분석을 하는 데 있어서 많은 연구가 이루어져 왔고 비선형 시계열 모형들에 관한 연구도 활발히 이루어져 왔으나 비선형 모형을 이용한 예측에 관한 연구는 아주 미흡한 편이다. 신경망 모형은 수많은 뉴런(neuron)으로 연결되어 있는 뇌의 구조와 같이 서로 다른 노드와 연결된 구조를 가지고 다른 뉴런으로부터 입력을 받으면 뉴런은 활성화되고 다른 뉴런들에게 신호를 통해서 전달된다. 신경망 모형은 이러한 신호의 전달을 수학적으로 모형화한 것이며 이는 여러 응용 분야에 폭 넓게 적용되고 있다.

Nicholls와 Quinn(1982)는 처음으로 RCA 모형(random coefficient autoregressive model)을 소개하였는데 여기서 1차 RCA 모형의 정상해(stationary solution)가 존재한다는 것을 증명하였으며, 최소제곱 추정량의 강한 일치성(strong consistency)과 점근적인 정규성(asymptotic normality)을 조사하였다[6]. Hwang과 Basawa(1994)는 추세 요소를 가지고 있는 1차 RCA 모형에 대해 LAN(local asymptotic normality)의 성질을 연구하였으며, 또한 확실한 정규 조건들 하에서 RCA 모형의 일치성과 조건적인 최소제곱 모수들의 점근적인 정규성에 대하여 연구하였다[2]. 그리고 Hwang과 Basawa(1998)의 연구에서 처음으로 RCA모형의 일반화된 모형, GRCA모형(generalized random coefficient autoregressive)을 소개하였고, 확률계수 벡터의 평균에 대한 조건부 최소제곱 추정량과 가중치가 부여된 조건부

*우송정보대학 컴퓨터정보과 부교수
논문접수: 2012년 9월 27일, 1차 수정을 거쳐, 심사완료: 2012년 10월 20일

최소제곱 추정량을 구하였고 그들의 극한분포에 대하여 연구하였다[3].

신경망을 시계열분석에 처음 이용한 학자는 Lepedes와 Faber(1987)로서 시계열 자료에 대해 다계층 구조를 갖는 퍼셉트론을 역전파 알고리즘을 이용하여 학습시킨 결과 관측치는 물론 예측치에 있어서도 기존의 계량적 예측기법보다 우수함을 입증하였다[4]. Hill, O'Conner 그리고 Remus(1996)는 시계열 자료의 예측에 있어서 신경망 모형을 이용한 예측방법과[1] Makridakis(1978)등이 제안한 6가지 통계적인 시계열 분석방법을 이용한 예측을 비교하였다[5]. White(1988)는 미국 IBM의 일별 주식값 변동자료에 대하여 신경망 이론과 Box-Jenkins 방법을 이용하여 예측을 하였는데, 한 시점 앞의 예측에서는 신경망을 이용한 방법이 Box-Jenkins 방법보다 평균제곱오차(mean square error)가 상대적으로 작게 나타나는 것을 보였다[8].

본 논문에서는 이분산성을 따르는 비선형 시계열 자료를 가지고 비선형 모형중 GRCA(1)모형과 자료의 형태에 상관없이 적용할 수 있는 신경망(neural networks)모형을 이용하여 예측을 해서 어느 모형이 시계열 자료에 적합한지를 비교 분석을 해보는 것이다. 예측 결과를 비교 분석해 본 결과 단기예측에서는 신경망을 이용한 예측값이 비선형모형을 이용한 예측값보다 좋았으며, 장기예측에서는 GRCA모형을 이용한 예측값이 더 좋은 것으로 나타났다.

2. 본론

2.1 GRCA(1) 모형에 의한 예측

Hwang과 Basawa(1994)는 추세 요소(regression trend)를 가지고 있는 1차 RCA 모형에 대해 연구하였으며, 정규조건들(regularity conditions)하에서 모형의 일치성과 최소제곱 추정량의 점근적인 정규성에 대하여 연구하였다.

비선형 시계열 모형인 p차 확률계수 모형(p-order random coefficient autoregressive model : RCA(p))은 아래와 같다.

$$X_t = (\theta_1 + Z_{t1})X_{t-1} + \cdots + (\theta_p + Z_{tp})X_{t-p} + \epsilon_t \quad 식(1)$$

여기서 { Z_{tj}, j=1,⋯,p}는 평균이 0이고 분산이 유한한, 독립이고 동일한 분포를 갖는 확률 변수이며, { ε_t }와 독립이다. 또한 { ε_t }는 평균이 0이고 분산이 σ^2인 독립이며 동일한 분포를 갖는 확률변수의 수열이고, 그리고 $\Theta_1, \cdots, \Theta_p$ 는 모수들이다.

Hwang과 Basawa(1998)는 일반화된 p차 확률계수 모형(generalized p-order random coefficient autoregressive precess : GRCA(p))을 소개하였는데 그 모형은 다음과 같다.

$$X_t = \Phi'_t X(t-1) + \epsilon_t \quad 식(2)$$

여기서 $\Phi_t = (\Phi_{t1}, \cdots, \Phi_{tp})'$는 (p×1)인 확률계수의 벡터, $X(t-1) = (X_{t-1}, X_{t-2}, \cdots, X_{t-p})'$는 시점 t에서의 (p×1)인 과거 관측치의 벡터이고, $\{\epsilon_t\}$는 확률 오차들의 수열이다.

$\begin{pmatrix} \Phi_t \\ \epsilon_t \end{pmatrix}$, $t = 0, \pm 1, \pm 2, \cdots$,는 독립이며 동일한 분포를 갖는 $((p+1)\times 1)$인 확률 벡터들의 수열이며 평균과 분산은 다음과 같다.

$$E\begin{pmatrix} \Phi_t \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} \phi \\ 0 \end{pmatrix} \text{ 이고, } Var\begin{pmatrix} \Phi_t \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} \Sigma_\phi & \sigma_{\phi\epsilon} \\ & \sigma'_{\phi\epsilon}\sigma^2_\epsilon \\ & \end{pmatrix}.$$

여기서 $\phi = (\phi_1, \cdots, \phi_p)'$는 $(p\times 1)$ 모수 벡터이고,

$\Sigma_\phi = Var(\Phi_t)$는 $(p\times p)$ 행렬이고,

$\sigma_{\phi\varepsilon} = Cov(\Phi_t, \varepsilon_t)$는 $(p\times 1)$벡터이고,

$\sigma^2_\varepsilon = Var(\varepsilon_t)$이다.

Hwang과 Basawa(1998)에 의하면 ϕ의 조건부 최소제곱 추정량 $\hat{\phi}$과 $\hat{\phi}$의 극한분포는 다음과 같이 얻을 수 있다.

$$\underset{(p\times 1)}{\hat{\phi}} = \left[\sum_{t=1}^{n} X(t-1)X'(t-1)\right]^{-1}\left[\sum_{t=1}^{n} X_t X(t-1)\right],$$

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \tau_1). \quad 식(3)$$

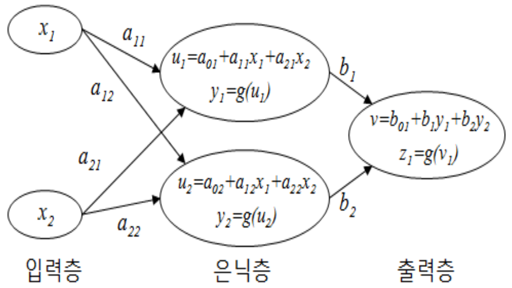
그리고 확률 계수 자기회귀 과정, $\{X(t)\}$의 예측치

는 다음과 같다.

$$\widehat{X(t)} = E(X(t) \mid F_{t-1}) = \beta' Y(t-1). \qquad 식(4)$$

2.2 신경망(Neural Network)에 의한 예측

신경망의 구조는 입력층, 은닉층, 출력층으로 구성되어 있으며 일반적으로 신경망의 구조는 입력노드의 개수 × 은닉노드의 개수 × 출력노드의 개수로 나타내는데 우리는 입력노드가 2개, 은닉노드가 2개 그리고 출력 노드가 1개인 2 × 2 × 1 신경망의 구조를 생각해 보자. 2 × 2 × 1 신경망의 구조는 <그림 1>과 같다.



[그림 1] 2×2×1 신경망의 구조

은닉층과 출력층의 각 노드는 [그림 2]에서와 같이 해당 노드로의 입력값들과 가중값(weights)들의 선형결합으로 계산한 후 변환 함수(mapping function)를 이용하여 노드의 출력값들을 구한다. 본 논문에서는 아래의 변환 함수를 사용하였다.

$$g(u) = \frac{1}{1+e^{-u}} \quad , \quad 0 < g(u) < 1 \qquad 식(5)$$

학습과정은 변환함수를 이용하여 계산된 함수값과 목표값 사이에서 가장 잘 적합되는 가중값을 찾는 과정을 말하며, 출력값과 목표값간의 학습오차를 최소화시킬 수 있는 오차판단 기준은 다음을 이용한다.

$$MSE = \frac{\frac{1}{2}\sum_{n=1}^{N}(z_n - t_n)^2}{N} \qquad 식(6)$$

여기서 N은 예제들의 전체 개수이고, z_n과 t_n은 각각 출력값과 목표값이다. MSE는 특이값에 따라 변이가 심하므로 안정적이지 못하지만 통계분석이나 신경망에서 다루고 있는 오차의 판단기준으로 주로 사용된다. 학습절차에서 C언어를 이용한 역전파 프로그램의 개요는 다음과 같다.

```
for(i=0; i<examples; i++)
    {
    forward( );
    back( );
    }
  changeWeight( );
```

학습규칙은 학습과정에서 가중값의 변화량을 적절하게 결정하는 방법으로서 w_m을 가중값 w의 m번째 학습 후의 변화된 가중값이라 하면 w_m은 다음과 같다.

$$w_m = w_{m-1} + c_m$$

여기서 c_m은 m번째 학습량이 끝났을 때 가중값 w의 변화량이다. 즉,

$$c_m = \mu c_{m-1} - (1-\mu) e_m d_m \quad 0 \le \mu < 1. \qquad 식(7)$$

여기서 e_m은 가중값이 변화하므로써 최소값으로 수렴되어가는 가중값에 대한 학습 비율로서 다음과 같다.

$$e_m = \begin{cases} e_{m-1} + \chi, & d_m f_m > 0 \\ e_{m-1} \times \phi, & d_m f_m \le 0, \ 0 < \phi < 1. \end{cases} \qquad 식(8)$$

여기서 χ와 ϕ는 모수이다. 그리고 d_m은 오차를 가중값에 대하여 미분한 모든 예제들의 누적합으로 m번째 학습후의 누적합이며, f_m은 현재와 과거의 미분값들의 가중 평균이다. 예측에 있어서 신경망의 모형은 1개의 입력층, 1개의 은닉층 그리고 1개의 출력층을 갖는 다계층 퍼셉트론을 사용하였으며, 신경망의 학습을 위하여 역전파 알고리즘을 사용하였고, 각 층의 변환함수는 시그모이드(sigmoid)함수를 사용하였다. 실제 시계열 자료에 포함되어 있는 관측치들은 신경망 모형에 선형 변환을

통하여 입력을 하여, 학습과정을 마친 후 출력되는 출력 값은 변환의 역과정을 거쳐 재변환된 값으로 출력하여 분석하였다. 관측치의 선형변환은 다음과 같은 식을 이용하였다.

$$\text{변환값} = \frac{\text{관측치} - \text{관측치의최소값}}{\text{관측치의최대값} - \text{관측치의최소값}} \quad \text{식}(9)$$

여기서 변환값은 0과 1사이 값을 갖는다.

학습과정을 정지시키는 규칙은 최대학습 수를 10,000으로 하여 10,000 epochs에 도달하거나 오차값이 0.001에 도달하는 경우에 멈추게 하였다. 따라서 10,000 epochs이 전에 학습이 수렴할 경우 오차값은 0.001 이하가 된다.

학습 후에 t+1시점 이후의 k개 예측값 $\widehat{Z_{t+1}}, \widehat{Z_{t+2}}, \cdots, \widehat{Z_{t+k}}$ 은 다음과 같이 구하는데 이 때의 시차는 1인 경우이다.

$$\begin{array}{ccc} <\text{ 예측값 }> & & <\text{ 입력값 }> \\ \widehat{Z_{t+1}} & \leftarrow & Z_t \text{ 와 } Z_{t-1} \\ \widehat{Z_{t+2}} & \leftarrow & \widehat{Z_{t+1}} \text{ 와 } Z_t \\ \vdots & & \vdots \\ \widehat{Z_{t+k}} & \leftarrow & \widehat{Z_{t+k-1}} \text{ 와 } \widehat{Z_{t+k-2}} \end{array} \quad \text{식}(10)$$

2.3. 사례분석

비선형 시계열 자료의 예측에 대하여 실제 자료를 이용하여 분석을 하고자 한다. 특히 이분산성을 갖는 자료들에 대하여 GRCA(1)모형과 신경망(neural networks) 모형을 이용하여 자료를 예측하여 각각의 MSE를 구하여 어느 모형이 어떤 자료에 잘 예측을 하는지 비교하려 한다. 비교값으로는 아래와 같은 MSE를 이용하였다.

$$\text{MSE} = \frac{1}{N}\sum_{l=1}^{N} e_l^2 \quad \text{식}(11)$$

여기서 e_l 은 l 시점후의 예측오차이며 아래와 같다.

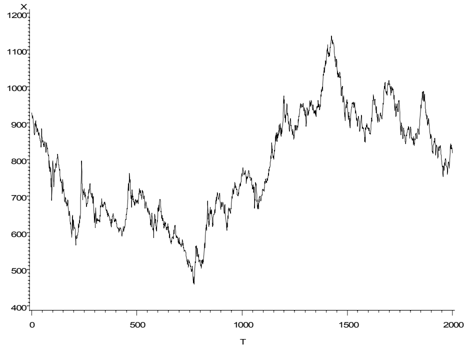
$$e_l = Z_{t+l} - \widehat{Z}_n(l)$$

〈표 1〉 신경망구조의 비교

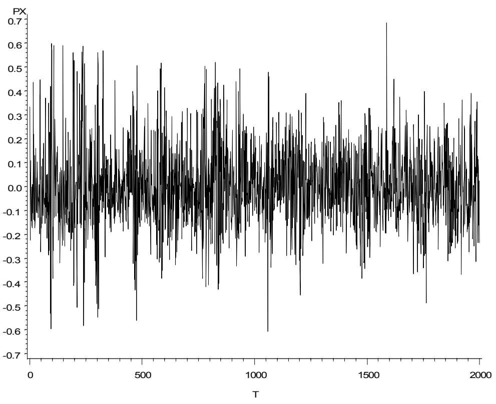
κ=0.1 ϕ=0.5 Θ=0.7 μ=0.9 epoch=200			
		RMSE	
Input	Hidden	학습오차	예측오차
1	1	0.021008	0.016444
1	2	0.019610	0.016030
1	3	0.019253	0.016074
1	4	0.019354	0.016269
2	1	0.018602	0.015159
<u>2</u>	<u>2</u>	<u>0.016294</u>	<u>0.013733</u>
2	3	0.016552	0.014886
2	4	0.016409	0.014767

신경망구조는 <표 1>과 같이 RMSE(Root Mean Square Error)를 이용하여 학습오차와 예측오차를 여러 신경망 구조에 비교하여 RMSE가 가장 작은 값을 갖는 2×2×1 신경망구조를 사용하였다.

본 연구에서 사용할 자료는 IMF 자금지원신청 이후 기간은 주가가 불안정하여 IMF 자금지원신청 이전 기간인 1990년 1월 3일부터 1996년 10월 31일까지의 한국증권거래소의 종가기준 일일 종합주가지수를 이용하여 이 후 12일간의 종합주가지수를 예측하여 실제 종합주가지수와 비교 분석하였다. 실제 자료들을 분산 안정화 변환, 즉 제곱근을 씌어 1차 차분을 한 변환을 통하여 분석에 적절한 정상(stationary) 시계열 자료로 만들어 예측을 한 후 신경망 결과와 비교하기 위하여 다시 역변환하여 분석을 하였다.



(X축:시간, Y축:종합주가지수)
[그림 2] 종합주가지수의 원자료



(X축:시간, Y축:변화율)

[그림 3] 종합주가지수의 변환($\sqrt{X_t}$ 대 $\sqrt{X_{t-1}}$)

위 [그림 3]에서 보듯이 1990년 1월 3일부터 1996년 10월 31일까지의 우리나라 일별 종합주가지수에 제곱근 변환을 한 자료는 정상성을 만족하고 이분산성이 있다는 것을 우리는 알 수 있다.

〈표 2〉 종합주가지수의 예측력 비교

	GRCA(1)	Nerual-Network
1	2.2650E-04	* 1.3009E-05
2	1.8398E-04	* 8.8925E-05
3	1.9812E-04	* 8.1471E-05
4	4.6021E-04	* 1.1389E-04
5	4.1461E-04	* 1.2647E-04
6	4.3656E-04	* 1.3081E-04
7	* 4.8045E-04	2.0464E-03
8	* 6.1513E-04	2.2927E-03
9	* 7.3692E-04	3.0129E-03
10	* 6.6888E-04	3.2690E-03
11	* 6.2906E-04	7.7645E-03
12	* 6.1517E-04	5.9617E-02

(* 표시를 한 것은 MSE가 더 작은 값이다.)

<표 2>로부터 우리나라의 일별 종합주가지수를 이용한 예측에서는 비선형 모형인 GRCA모형은 좋은 예측값을 보여주었고, 신경망 방법에 의한 예측값은 비선형 모형과는 조금 차이가 있는 예측값을 보여 주었다. 초기 6일간의 예측에서는 신경망을 이용한 예측값이 비선형모형을 이용한 예측값보다 좋았으며, 7일 이후의 예측에서는 GRCA모형을 이용한 예측값이 더 좋았다.

3. 결론

선형 Box-Jenkins모형과 신경망 방법을 비교분석하는데 있어서 단기예측에서는 선형 Box-Jenkins모형의 예측값이 좀 더 좋은 예측값을 가졌고, 장기예측에서는 신경망 방법의 예측값이 더 좋은 예측값을 갖는 것을 알았지만(Tang 등(1991))[7] 위의 결과로부터 Tang의 결과와는 다른 비선형모형을 이용한 예측에서는 단기 예측에서는 신경망 방법의 예측값이 더 좋았고, 장기예측에서는 비선형모형을 이용한 예측값이 더 좋은 것으로 나타났다. 이러한 결과를 통해서 보면 신경망 방법은 관찰 가능한 자료의 끝부분에 영향을 받아 초기예측값에 조금 더 적절한 예측값을 제공하는 것으로 볼수 있고, 비선형 모형은 관찰된 자료의 전체적인 추세와 경향을 그대로 반영하므로 초기의 예측값보다는 중·장기의 예측값에 좀더 적절한 예측값을 가짐을 알수 있었다.

참 고 문 헌

[1] Hill, T., O'Connor, M. and Remus, W., “Neural Network Models for Time Series Forecasts”, Management Science, 42, 1082-1092, 1996.

[2] Hwang, S. Y. and Basawa, I. V., “Large sample inference for conditional exponential families with applications to nonlinear time series models”, Journal of Statistical Planning and Inference 38, 141-158, 1994.

[3] Hwang, S. Y. and Basawa, I. V., “Parameter estimation for generalized random coefficient autoregressive processes”, Journal of Statistical Planning and Inference 68, 323-337, 1998.

[4] Lepedes, A. and Faber, R., “Nonlinear Signal Processing Using Neural Networks : Prediction and System Modeling”, Los Almos National Laboratory Report, LA-UR-87-2662, 1987.

[5] Makridakis, S. and Wheelwright, S., “Forecasting : Method & Applications”, Santa Barbara, John Wiley & Sons, 1978.

[6] Nicholls, D. F. and Quinn, B. G., “Random Coefficient Autoregressive Models: An Introduction. Lecture Notes in Statistics”, Vol. 11,

Springer-Verlag, New York, 1982.
[7] Tang, Z., Almeida, C., Fishwick, P.A., “Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology”, Simulation, 57(5), 303-310, 1991.
[8] White, H., “Economic Prediction Using Neural Networks : The Case of IBM Stock Prices”, IJCNN, 2, 451-458, 1988.

김 인 규

- 1986년 5월: University of Akron 수학과(이학사)
- 1987년 12월: University of Georgia 통계학과(이학석사)
- 1999년 2월: 충북대학교 전자계산학과(이학박사)
- 1992년 9~현재: 우송정보대학 컴퓨터정보과 교수
- 관심분야: 시계열분석, 신경망, 프로그램 언어
- E-Mail: ikkim0056@wsi.ac.kr