

신경망이론에 의한 시계열자료의 분석1)

윤여창²⁾, 허문열³⁾

요 약

본 연구에서는 신경망이론을 이용하여 시계열자료를 분석할 때 문제가 되고 있는 초기 가중값을 선정하는 방법을 제시하고자 한다. 기존의 연구에서 학습을 위한 초기 가중값의 결정은 난수에 의존하고 있다. 본 연구에서는 신경망학습의 효율적인 초기값을 선택하기 위하여 제어상자를 이용한다. 그리고 학습과정에서 가중값의 변화를 추적하고 적절한 가중값의 범위를 탐색하면서 새로운 초기값을 제어상자를 통하여 실시간으로 재설정하는 방법을 제시한다.

1. 서 론

신경망이론을 이용하여 시계열자료를 학습할 때에 직면하는 문제들 중에서, 학습의 초기값을 어떻게 효과적으로 선정하게 하는가가 주요 관심사이다. 학습할 시계열자료에 대하여 신경망의 가중값에 대한 사전 정보를 알 수 있다면, 신경망학습이 효율적으로 이루어질 수 있다. 그러나 그렇지 못한 경우가 대부분이기 때문에 기존의 연구에서 신경망학습을 위한 초기 가중값은 난수를 이용하고 있다.

Ripley(1994)는 신경망의 초기 가중값으로 난수를 적용한 학습을 수없이 반복한 후에, 보다 잘 수렴되는 경우의 난수를 학습의 초기값으로 택하고 있다. Smith(1993)는 랜덤하게 선택한 초기값으로 수많은 실행을 반복하는 시행착오 후에 채택된 경험적 모수들을 적응적 학습률(adaptive learning rate)에 적용하여 신경망학습을 하고 있다. 윤여창(1995)은 시계열자료의 신경망학습을 위한 초기값 선택의 새로운 방법으로써 Easton(1994)의 제어상자(controller box)를 이용하고 있다.

본 연구에서는 신경망이론을 이용하여 시계열자료를 분석하기 위한 학습방법으로 적응적 학습률을 이용한다. 학습의 초기값은 제어상자를 이용하여 동적(dynamic)으로 선택한다. 신경망학습과 예측에 이용할 자료는 시계열분석에서 자주 이용되는 고전적인 자료인 Wolf의 태양흑점자료(sunspot data)이다.

2. 신경망의 구조와 학습

2.1 신경망의 구조

입력노드가 2개, 출력노드가 1개 그리고 은닉층이 1개인 신경망을 고려해 보자. 은닉층과 출력

1) 이 논문은 1995년도 성균관대학교 성균학술연구비에 의하여 연구되었음.
2) (565-701) 전북 완주군 삼례읍 후정리 우석대학교 계산통계학과 전임강사.
3) (110-745) 서울시 종로구 명륜동 3가 성균관대학교 통계학과 교수.

층의 각 노드가 2개인 입력노드를 갖는다고 할 경우에, 입력노드의 개수×은닉노드의 개수×출력노드의 개수, 즉 2×2×1 신경망의 구조는 (그림 1)과 같다. 여기서 은닉층과 출력층의 각 노드는 해당 노드로의 입력값들과 가중값들의 선형결합으로 계산한 후, 다음과 같은 변환함수를 이용하여 노드의 출력값들을 계산한다.

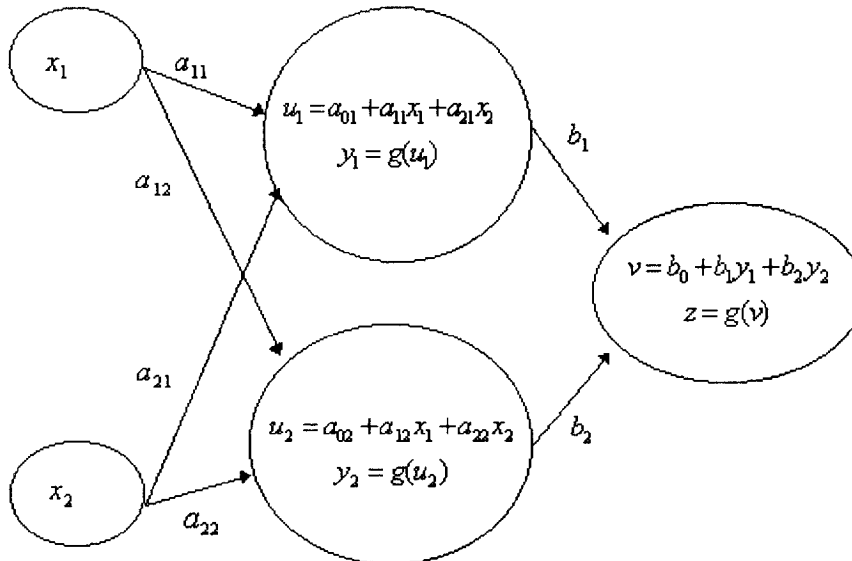
$$g(u) = \frac{1}{1 + e^{-u}}, \quad 0 < g(u) < 1.$$

2.2 신경망의 학습

신경망이론에서 학습과정은 변환함수를 통과한 값과 목표값 사이에 가장 잘 적합되는 가중값을 찾는 것이며, 올바른 가중값을 어떻게 찾아야 하는가가 학습의 중요한 문제다. n 번째 예제에 대하여 변환함수를 통과한 값, 즉 출력값 (z_n)과 목표값 (t_n)간의 학습오차는 식 (1)과 같은 오차판단기준을 이용한다. 여기서 N 은 예제들의 전체 개수이다.

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

학습절차는 역전파(backpropagation) 알고리즘을 이용한다. 각각의 처리절차는 먼저 변환함수를 현재 가중값에 적용하여 계산하고 각 예제에 대한 가중값의 변화량을 결정한다. 그리고 각 예제에 대한 가중값들을 누적시킨 결과를 이용하여 현재 가중값들을 바꿔준다. 이와같은 과정을 반복하면서 MSE가 최소화 되는 가중값들을 구하기 위하여, 각 노드에 대한 가중값의 변화가 매우 적어질 때까지 가중값들을 계속적으로 변화시키면서 학습을 진행시킨다.



<그림 1> 2×2×1 신경망의 구조

각 가중값들에 대한 순수 변화량은 전체 예제들에 대해 미분한 값들을 누적시키면서 구하게 되는데, 이는 다음과 같이 단순화 시킨 식으로 표현될 수 있다. 즉 은닉노드에 대한 가중값 a_{ij} 와 출력노드에 대한 가중값 b_j 를 가중값 w 라고 하면, m 번째 학습한 후의 가중값은 식 (2)와 같다.

$$w_m = w_{m-1} + c_m, \quad (2)$$

여기서 c_m 은 m 번째 학습 후의 가중값 w 의 변화량이다.

$$c_m = \mu c_{m-1} - (1 - \mu) e_m d_m, \quad 0 \leq \mu < 1. \quad (3)$$

식 (3)에서 음의 부호는 MSE가 감소되는 변화를 나타내 준다. 모수 μ 는 가중값의 변화량 c_m 이 d_m 에 영향을 받는 정도를 조절하여 준다. 그리고 e 는 가중값들이 변화됨에 따라 최소값으로 수렴되어 가는 비율을 조절하는 모수 x 와 ϕ 를 이용하여 수렴되는 정도를 조절하고 있으며 다음 식 (4)와 같다.

$$e_m = \begin{cases} e_{m-1} + x, & d_m f_m > 0 \\ e_{m-1} \times \phi, & d_m f_m \leq 0 \end{cases}, \quad 0 < \phi < 1. \quad (4)$$

d_m 은 모든 예제들에 대하여 MSE를 각 학습량에서의 가중값으로 미분한 값들의 누적합이다. m 번째 학습 후의 누적 가중값들은 다음 식 (5)와 같다.

$$d_m = \sum_{n=1}^N \left(\frac{\partial E}{\partial w_m} \right)_n. \quad (5)$$

누적합이 최근 변화에 더 크게 또는 더 작게 영향을 받도록 하기 위해서는 다음 식 (6)을 적용할 수 있다.

$$f_{m+1} = \theta f_m + (1 - \theta) d_m, \quad 0 \leq \theta < 1. \quad (6)$$

따라서 m 번째의 학습이 끝난 후에 가중값을 계속적으로 변화시킨 다음, 새로운 가중값을 이용하여 다시 학습과정을 반복하면서 MSE의 변화가 한계값 이하이거나 주어진 학습 시간에 도달되면 학습과정이 끝난다. 본 연구에서 적용한 학습방법은 식 (3)과 같은 적응적 학습률을 이용한다. 시계열자료의 학습예제에는 변동주기를 고려한 입력값들을 이용한다.

3. 제어상자를 통하여 초기값을 결정하는 방법

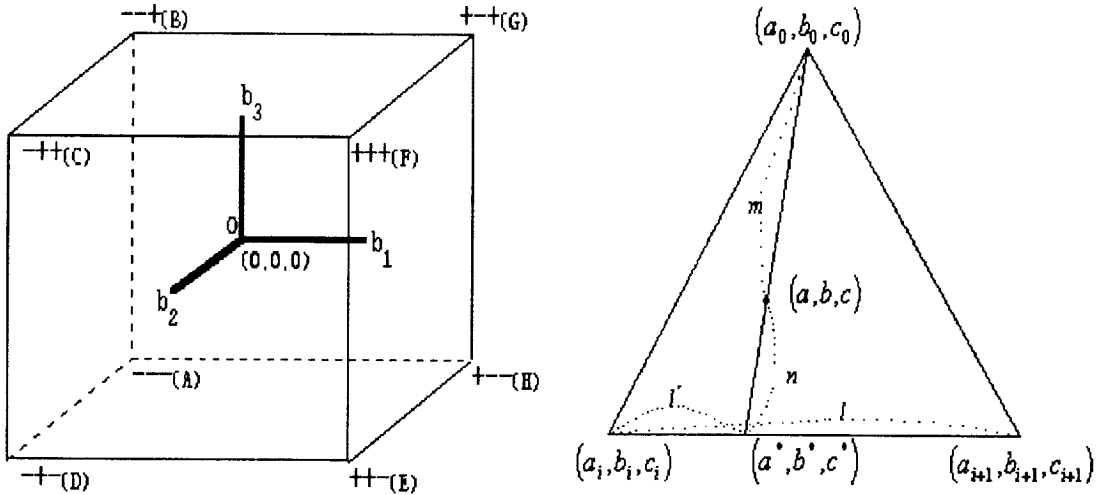
Lewbel(1994)은 전통적인 시계열자료의 분석이나 신경망을 통한 분석에서 중요시 되고 있는 두 가지 공통사항으로써, 함수형태의 결정과 모수의 추정문제가 중요하다고 하였다. 여기서 함수형태의 결정이라고 하면 전통적인 Box-Jenkins 방법에서의 모형식별문제이며, 신경망인 경우에는 각 층별 노드의 개수, 층의 개수, 변환함수의 형태등을 결정하는 문제이다. 그리고 모수의 추정문제라고 하면 Box-Jenkins 방법에서의 모형추정문제이며, 신경망인 경우에는 가중값을 결정하는 문제를 의미한다.

시계열자료의 신경망학습을 위해 본 논문에서 적용한 적응적 학습률의 모수들에 따라 학습속도가 달라질 수 있다. 모수들에는 학습과정에서 가중값들이 변화됨에 따라 최소값으로 수렴되어 가는 비율을 조절하는 모수인 x , ϕ 가 있고, 학습과정에서 가중값의 최근 변화에 영향을 받는 정도

를 조절하는 모수인 θ, μ 가 있다. 이와같은 모수들에 적절한 값을 주어 신경망학습의 시행착오를 겪는 실험을 반복하면서, 경험적으로 가장 학습효과가 높고 수렴 속도가 빠른 모수들을 선택하게 된다. Smith는 경험적으로 가장 학습효과가 높도록 학습률을 조정하기 위한 모수들로서 $x=0.1, \phi=0.5, \theta=0.7$ 그리고 $\mu=0.9$ 를 택하였다.

신경망학습에 시계열자료를 적용할 때, 초기값은 중요한 역할을 한다. 일반적으로 신경망학습의 초기값은 일정한 구간에서 발생시킨 난수를 이용한다. 본 논문에서 이용되는 초기값은 컴퓨터와의 상호작용(interactive)을 통하여 실시간으로 선정된 값을 이용하고자 한다. 이 방법을 적용하기 위하여 Easton(1994)이 제시한 제어상자(controller box)를 이용하였다. 유종영(1994)은 제어상자의 개념을 회귀진단에 이용하고 있으며, 윤여창(1995)은 시각적인 접근을 이용하여 동적으로 제어상자 위에서 초기값을 찾고 이 값을 시계열자료의 신경망학습에 이용하고 있다. 제어상자를 통하여 선택된 값을 신경망의 초기값으로 적용하기 위한 방법은 다음과 같다.

다차원 공간의 모든 점을 2차원으로 축소하여 표시한다는 것은 불가능하다. 따라서 보다 효율적인 표현방법이 필요하다. 예를들어 $2 \times 2 \times 1$ 신경망의 구조에서 가중값 a_1, a_2 그리고 b 는 각각 3차원 값을 갖는다. 여기서 가중값 b 가 가질 수 있는 값 (b_1, b_2, b_3) 을 (그림 2a)의 3차원 공간에서 (그림 2b)와 같이 2차원 공간으로 단순화 시킨 평면으로 나타내는 것은 가능하다. 선택될 수 있다. 즉 6면체의 각 모서리 (A, B, C, D, E, F, G, H)를 따라가면서, 중점 O와 연결시킨 2^3 개의 이등변삼각형들은 $\triangle OAB, \triangle OBC, \triangle OCD, \triangle ODE, \triangle OEF, \triangle OFG, \triangle OGH, \triangle OHA$ 이다. 6면체의 중점 $(0,0,0)$ 으로 부터 각 모서리의 위치가 양의 방향에 위치하면 (+)기호를, 음의 방향에 위치하면 (-)기호를 부여한다. 이와같이 평면으로 단순화 시킨 이등변삼각형들 중에서 $\triangle OAB$ 의 내부점 (a, b, c) 를 신경망의 다차원 가중값으로 인식하는 방법은 다음과 같다.



(a) 3차원 공간

(b) $\triangle OAB$

<그림 2> 3차원 공간과 이등변삼각형

중심점 $(0,0,0)$ 에서 각 꼭지점까지의 상대거리를 k 라고 할 때 시작 꼭지점의 좌표 $(-k, -k, -k)$ 에서부터 차례로 $(-k, -k, k)$, $(-k, k, k)$, $(-k, k, -k)$, $(k, k, -k)$, (k, k, k) , $(k, -k, k)$, $(k, -k, -k)$ 에서 다시 시작 꼭지점으로 되돌아 오는 순환경로를 삼각형의 각 밑변으로 한다. 그러면 양변의 길이가 k 이고 밑변의 길이가 $2k$ 인 이등변삼각형이 2^3 개 생기며, 그 첫 번째 삼각형은 (그림 2b)와 같다. 여기서 삼각형의 각 꼭지점 (a_0, b_0, c_0) , (a_i, b_i, c_i) , $(a_{i+1}, b_{i+1}, c_{i+1})$ 은 컴퓨터화면의 절대좌표로 주어지고, 선택된 점의 절대위치 (a, b, c) 도 알고 있으므로, m, n, l', l 의 길이는 각각

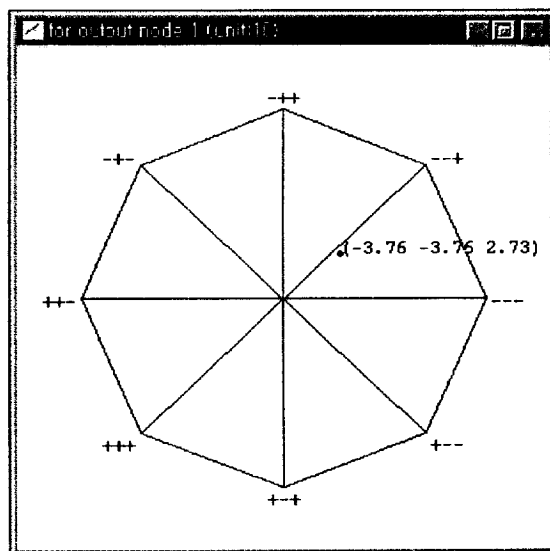
$$\|m\| = \sqrt{(a-a_0)^2 + (b-b_0)^2 + (c-c_0)^2}, \quad \|n\| = \sqrt{(a-a^*)^2 + (b-b^*)^2 + (c-c^*)^2},$$

$$\|l'\| = \sqrt{(a_i-a^*)^2 + (b_i-b^*)^2 + (c_i-c^*)^2}, \quad \|l\| = \sqrt{(a_{i+1}-a^*)^2 + (b_{i+1}-b^*)^2 + (c_{i+1}-c^*)^2}$$

이다. 따라서 시작 꼭지점 (a_i, b_i, c_i) 와 절대좌표 (a^*, b^*, c^*) 의 상대길이를 c 라고 하면 절대좌표에 대한 상대좌표를 구할 수 있다. 즉 $l:l' = 2:c$, $c = \frac{2 \cdot l'}{l}$ 이며 $t = -k + c$ 이다. 그러므로 절대좌표 (a^*, b^*, c^*) 의 상대좌표는 $(-k, -k, t)$ 이다. 따라서 선택된 점의 절대위치 (a, b, c) 의 상대좌표 (x, y, z) 는 이등변삼각형의 꼭지점과 밑변을 이은 사선을 내분한 점이므로 다음 식(7)과 같다.

$$(x, y, z) = \left(\frac{-m \cdot k}{m+n}, \frac{-m \cdot k}{m+n}, \frac{m \cdot k \cdot t}{m+n} \right). \tag{7}$$

본 논문에서는 (그림 3)과 같은 제어상자를 통하여 선택될 수 있는 가중값들의 범위를 $(-10 -10 -10)$ 에서 $(+10 +10 +10)$ 까지 설정하였다. 따라서 (그림 3)의 제어상자에서 선택된 가중값 $b = (-3.76 -3.76 2.73)$ 은 (그림 2b)의 $\triangle OAB$ 평면 위의 한 점 (a, b, c) 에 대한 상대좌표이다.



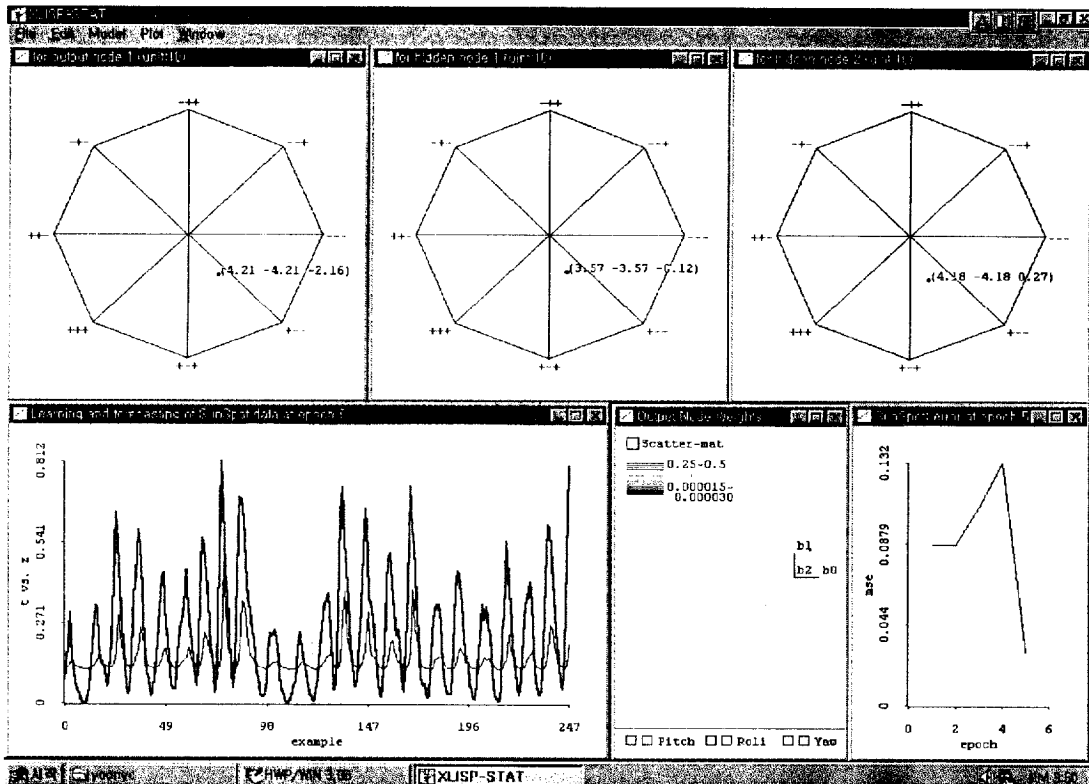
<그림 3> 제어상자

이렇게 선택된 값들은 신경망 학습의 초기값으로 사용되는데, 제어상자를 통하여 동적으로 새로운 가중값을 다시 적용시켜 학습할 수 있다.

(그림 4)는 태양흑점자료를 이용하여 신경망학습을 하기 위한 초기 과정이다. 본 논문에서 고려한 신경망은 $2 \times 2 \times 1$ 구조이므로 입력노드에 대한 3차원 가중값의 제어상자가 각각 2개, 은닉노드에 대한 제어상자가 1개 있다. 각 연결 가중값에 대한 초기값들은 출력값과 목표값 사이에 보다 시각적으로 잘 적합되는 값으로 랜덤하게 제어상자를 통하여 선택되며, 이를 신경망학습의 초기값으로 설정한다. (그림 4)에서 선택된 초기값은 다음과 같다.

$$A = \begin{bmatrix} 3.57 & 4.18 \\ -3.57 & -4.18 \\ -0.12 & 0.27 \end{bmatrix}, \quad b = \begin{bmatrix} 4.12 \\ -4.12 \\ -2.16 \end{bmatrix}.$$

학습의 전 단계에 적용될 수 있는 이 방법은 MSE를 작게 해주는 임의의 초기값들을 제어상자 위에서 동적으로 변화시킬 수가 있다. 이렇게 하면 시행착오를 줄이고 효율적으로 초기값을 찾을 수 있다. 만약에 초기값이 잘못 선택되어 가중값이 더 이상 변화되지 않으면 학습과정 중이라도 제어상자를 통하여 새로운 초기값을 실시간으로 재설정할 수 있다. 이와같은 방법을 반복하면서 MSE를 줄여 나간다.



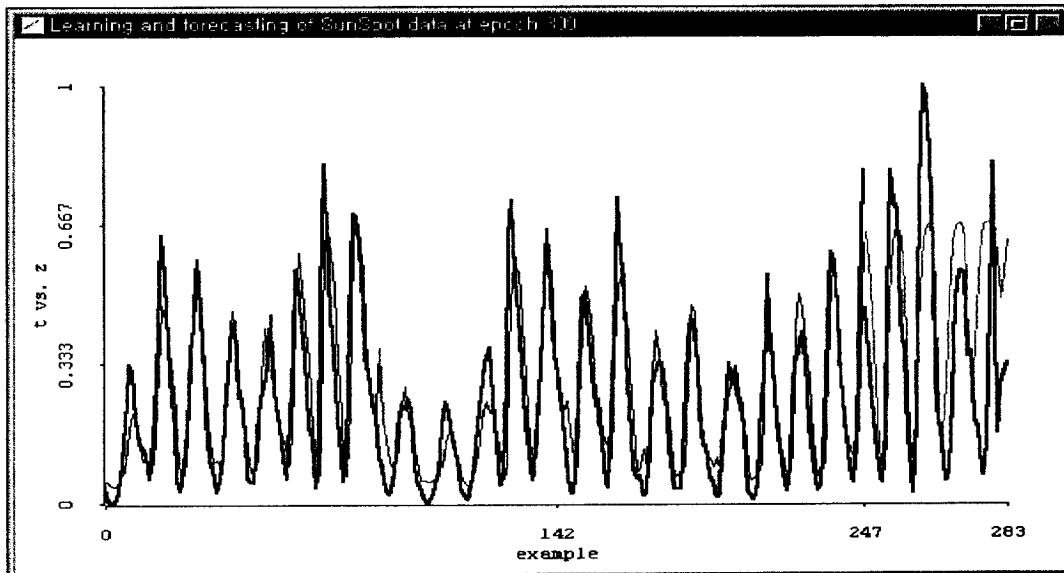
<그림 4> 초기 가중값의 선택과정 : 시계열자료를 분석하기 위한 초기 화면은 3개의 제어상자 화면, 학습과정과 예측결과 화면 그리고 가중값과 오차의 변화화면등으로 구성된다. 학습과정 화면에서 가는 실선은 초기값을 선택하였을 때의 적합 정도를 나타내 주고 있다.

4. 실증분석

이 장에서는 앞에서 제안된 제어상자를 통한 학습방법을 이용하여, 1700년부터 1983년 사이의 284년 동안 관측된 Wolf의 태양흑점자료를 변환함수 $g((Z_t - \bar{Z})/s_z)$ 로 변환시킨 후에 분석한다. 태양흑점자료는 상승 또는 하강 추세나 확실한 계절성을 관찰할 수 없지만 9년에서 14년의 주기를 가지고 순환하며 평균주기는 11년인 순환변동을 가지고 있는 시계열이다. 이 자료를 이용하여 신경망학습을 하고 그 예측결과를 Box-Jenkins 예측결과와 비교해 본다. 신경망의 학습을 위해서 1700년부터 1947년 까지의 248년 간의 자료를 학습자료로 이용하고, 예측을 위한 검정자료는 36년 간의 자료로써 1948년 부터 1983년 사이의 자료다. $2 \times 2 \times 1$ 신경망구조에서는 2개 썩의 순차적인 입력자료들을 이용하여 신경망학습과 예측을 한다. 따라서 입력자료는 t 시점과 $t-1$ 시점, t 시점과 $t-5$ 시점 그리고 11년의 순환변동을 반영하기 위하여 t 시점과 $t-10$ 시점의 세가지 예제로 구분하였고, 각 경우의 신경망학습의 결과를 SAS분석 결과와 비교한다.

(그림 5)는 태양흑점자료를 이용한 신경망학습과 36년 예측을 수행한 결과이다. (그림 5)에서 학습의 결과로 변화된 최종 가중값들인 은닉노드에 대한 가중값 a_{ij} 와 출력노드에 대한 가중값 b_j 는 다음과 같다.

$$A = \begin{bmatrix} 1.119646452747359 & -0.767010329695036 \\ -3.979707374168427 & -6.914862506268461 \\ -0.803948198200696 & -3.411689934169255 \end{bmatrix}, \quad b = \begin{bmatrix} 0.688736970867564 \\ -3.363312242762144 \\ -4.023137334662552 \end{bmatrix}$$



<그림 5> 신경망학습을 300번 수행한 후에 36년 예측한 값 : 태양흑점자료는 굵은 실선으로 표시 되어 있고 학습자료(0~247)와 검정자료(248~283)로 분리된다. 이 그림에서 신경망학습과 예측을 위한 변동주기는 10인 경우이다. 가는 실선은 학습한 결과(0~247)와 예측결과(248~283)를 나타내고 있다.

<표 1> 태양흑점자료에서의 예측오차(MSE) : 변동주기를 1, 5, 10으로 구분하였을 때의 MSE를 각각 통계적 분석(SAS)결과와 비교한다. 변동주기가 1과 5, 10인 경우의 MSE가 통계적 분석결과보다 작은 경우에는 각각 †와 ‡로 표시한다.

예측기간	통계적 방법	2×2×1신경망		
		lag=1	lag=5	lag=10
1년	0.00021	0.000155 †	1.50265e-6 ‡	1.00265e-7 ‡
3년	0.00182	0.003002	0.001401 ‡	0.001117 ‡
6년	0.00117	0.004093	0.001083 ‡	0.001047 ‡
12년	0.00119	0.004669	0.001277	0.007172
24년	0.00784	0.006852 †	0.001368 ‡	0.005191 ‡
36년	0.01302	0.012741 †	0.002825 ‡	0.045021

이 가중값을 이용하여 36년 까지의 신경망 예측을 하고, 그 예측결과를 Box-Jenkins 예측결과와 비교해 보면 (표 1)과 같다.

(표 1)의 통계적 방법은 SAS를 이용한 모형식별과 모수추정의 결과이며 (식 8)과 같다.

$$AR(2) ; Z_t = 0.14628654 + 1.4145Z_{t-1} - 0.71994Z_{t-2} + a_t. \quad (8)$$

태양흑점자료는 추세나 계절성이 없고 평균주기가 11년인 순환변동을 가지고 있는 시계열이므로 본 연구에서는 편의상 세가지의 변동주기에 대하여 학습과 예측을 하였다. 신경망의 예측결과만 살펴볼 때, 변동주기가 5인 경우는 모든 예측기간에서 변동주기가 1인 결과보다 MSE가 작다. 평균주기가 11년인 순환변동주기를 반영하기 위한 변동주기가 10인 경우는, 12년과 36년 예측기간을 제외한 기간에서 변동주기가 1인 결과보다 MSE가 작다. 변동주기가 5와 10인 경우를 비교하면 12년과 24년 예측을 제외하면 변동주기가 10인 경우에 MSE가 작다. 신경망의 예측결과를 통계적 예측결과와 비교할 때, 신경망은 변동주기가 10인 경우에 12년과 36년 예측을 제외한 다른 예측기간에서 통계적 예측결과보다 MSE가 작다. 변동주기가 5인 경우에는 12년 예측을 제외한 다른 예측기간에서 통계적 예측결과보다 MSE가 작다. 변동주기가 1인 경우에는 단기예측과 장기예측에서 MSE가 더 작지만 중기예측에는 전통적인 방법이 보다 효율적이다.

5. 결론

본 연구에서는 시계열자료를 이용하여 신경망 학습과 예측을 할 때 문제가 될 수 있는 초기값을 선정하는 방법을 제시하였다. 신경망학습의 다차원 초기값은 기존의 난수보다는 Easton(1994)의 제어상자를 이용하였다. 학습과정 중이라도 제어상자를 이용한 시각적인 접근을 이용하여 실시간으로 가중값을 재설정할 수 있도록 하였다. 이러한 처리는 학습과정에서 수많은 시행착오를 줄임으로써 학습시간을 줄일 수 있다. 제어상자는 다차원 공간을 2차원으로 축소하여 보다 효율적으로 다차원 가중값들을 표현시킬 수 있지만 모든 가중값 공간을 나타내지는 못한다. 또한 Easton(1994)은 경험적으로 6차원이 넘는 경우에는 제한된 화면 크기에서 최소 64개의 개별 삼각형 평면에 대한 가중값의 변화를 고려해야 하기 때문에 시각적으로 효율성이 떨어진다고 하였다. 이와같은 제어상자의 한계 때문에 본 연구에서는 보다 복잡한 신경망 구조 보다는 입력노드와 은닉노드의 개수를 각각 2개 씩으로 제한한 2×2×1 구조로 한정하였다. 그러나 경험적으로 보면

$2 \times 2 \times 1$ 구조에서도 학습과 예측이 가능하였다.

참 고 문 헌

- [1] 유종영 (1994). 동적그래픽스에 의한 회귀진단, 성균관대학교 박사학위논문.
- [2] 윤여창 (1995). 시계열자료에서 신경망이론의 적용, 성균관대학교 박사학위논문.
- [3] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis : Forecasting and Control*, Holden-Day.
- [4] Easton, G.S. (1994). A Simple Dynamic Graphical Diagnostic Method for Almost Any Model, *Journal of the American Statistical Association*, Vol. 89, 201-207.
- [5] Kuan, C.M. and White, H. (1994). Artificial Neural Networks : An Econometric Perspective, *Econometric Reviews*, Vol. 13(1), 1-91.
- [6] Lewbel, A. (1994). Comments on Artificial Neural Networks : An Econometric Perspective, *Econometric Reviews*, Vol. 13(1), 99-103.
- [7] Riply, B.D. (1994). Neural Networks and Related Methods for Classification, *Journal of Royal Statistical Society, B*, Vol. 56(3), 409-456.
- [8] Smith, M. (1993). *Neural Networks for Statistical Modeling*, New York, Van Nostrand Reinhold.
- [9] Tang, Z., Almeida, C., Fishwick, P.A. (1991). Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology, *Simulation*, Vol. 57(5), 303-310.
- [10] White, H. (1988). Learning in Artificial Neural Networks : A Statistical Perspective, *Neural Computation*, Vol. 1, 425-464.