

적응적 학습방법과 초기값의 개선에 의한 신경망 모형을 이용한 시계열 예측

윤 여 창[†] · 이 성 덕^{††}

요 약

본 연구에서는 신경망 모형을 이용한 시계열 예측에 있어서 분석할 시계열의 특성에 맞는 적응적 학습률을 구하고 초기값의 동적인 적용을 통한 개선된 학습방법을 이용하여 신경망 예측을 하고 통계적인 Box-Jenkins 예측 결과와 비교해 봄으로써 두 방법간의 시계열 예측 효율성을 비교한다. 신경망 모형에 맞는 적응적 학습률은 표준 직교 배열표에 의해 실험계획을 한 25가지의 모수 조합으로부터 구하고, 신경망 학습의 초기값은 Easton의 제어상자를 동적으로 적용하여 실시간으로 선택할 수 있도록 한다. 실증분석에 적용된 시계열자료는 1700년부터 1988년 까지의 태양 흑점 자료이다.

A Time Series Forecasting Using Neural Network by Modified Adaptive Learning Rates and Initial Values

Yeo-Chang Yoon[†] · Sung-Duck Lee^{††}

ABSTRACT

In this work, we consider the forecasting performance between neural network and Box-Jenkins method for time series data. A modified learning process is developed for neural network approach at time series data, i.e. properly adaptive learning rates selecting by orthogonal arrays and dynamic selecting of initial values using Easton's controller box. We can obtain good starting points with dynamic graphics approach. We use real data sets for this study : the Wolf yearly sunspot numbers between 1700 and 1988.

1. 서 론

신경망 모형을 이용하여 시계열자료를 학습할 때에 나타날 수 있는 문제들로는, 초기 가중값과 초기 포화점 그리고 학습률 등을 들 수 있다. 시계열자료에 대하여 신경망의 초기 가중값에 대한 사전 정보를 알 수 있거나 적절한 학습률을 구하고 포화점을 쉽게 벗어나

게 할 수 있다면, 시행착오를 줄이면서 더 적절한 신경망 학습 결과를 얻을 수 있다. 그러나 그렇지 못한 경우가 대부분이므로, 초기 가중값 선택의 기존 연구는 난수에 의존하고 있으며 학습률은 경험적인 값들을 이용하고 있다.

Russo[1]는 초기값을 $-2.4/\text{입력노드의 개수}$ 와 $+2.4/\text{입력노드의 개수의 범위 내에서 발생시킨 난수}$ 를 취하고 있다. Smith[2]는 전역최소값으로 수렴시키기 위해서 수많은 실행을 반복하는 시행착오를 하게 되는데 경험적으로 $0.2 \times (X - 0.5)$, $X \sim U(0,1)$ 의 난수를 이용한다. 그러나 신경망 모형에서는 분석할 자료에

* 이 논문은 1998년도 우석대학교 학술연구조성비에 의하여 연구되었음.

† 정 회 원 : 우석대학교 전산통계학과 교수

†† 정 회 원 : 충북대학교 통계학과 교수

논문접수 : 1998년 3월 24일, 심사완료 : 1998년 8월 13일

따라서 잘못 선택된 초기값으로 인하여 조기 포화점에 수렴되는 경우를 학습 결과로 택하는 경우가 많다. 이와 같은 학습은 전역 최소값으로 수렴시키지 못하므로써 오차가 많이 포함될 수 있는 비정상적의 학습 과정이다. Ripley[3,4]는 랜덤하게 선택한 몇 개의 초기값들을 각각 신경망의 초기 가중값으로 사용하면서 보다 잘 적합되는 초기값을 최종적으로 사용하여 학습을 하고, Kushner[5], White[6]등의 전역최소값을 구하는 처리절차들을 이용하여 초기값의 문제를 해결할 수 있다고 하였다. 윤여창[7,8]은 초기값을 시각적으로 적절히 변화시키는 방법을 이용하여 조기 포화점을 효과적으로 벗어나게 할 수 있는 신경망 학습을 유도하고 있다.

적용적 학습률을 구하는데 있어서 Smith[2]는 수많은 실행을 반복하는 시행착오 후에 선택된 적용적 학습률의 경험적 모수들을 이용하여 신경망 학습을 하고 있다. 윤여창[7,8]은 시계열자료의 신경망 학습을 위하여 Smith[2]의 경험적 모수들을 이용하면서, 초기값을 결정하기 위한 새로운 방법으로써 Easton[9]의 제어상자를 이용하여 동적으로 초기값을 선택하고 있다.

본 연구에서는 시계열분석에서 자주 이용되는 고전적인 자료인 Wolf의 태양흑점자료를 이용하여 신경망 예측결과와 전통적인 시계열예측의 결과를 비교한다. 신경망의 개선된 학습 방안으로써 먼저 태양흑점 시계열자료에 잘 적합되는 적용적 학습률의 모수들을 선택하는 실험을 제시한다. 학습에 이용되는 적용적 학습률의 모수들은 표준 직교 배열표에 의해 설정된 25가지 경우에 대하여 학습오차가 작고 빠르게 수렴되는 경우의 모수를 선정하여 신경망 학습과 예측을 한다. 선택된 모수들을 이용한 학습과정 중이라도 조기 포화점에 빠지는 경우가 나타나면 이를 실시간으로 벗어나게 하면서 학습을 진행시킨다. 또한 본 연구에서는 입력노드가 2개이고 은닉층이 1층인 신경망 모형을 적용하고 있기 때문에, 입력자료에 대해서도 시차에 변화를 주면서 학습하고 그 결과를 비교한다. 개선된 학습 과정을 이용한 신경망 예측 결과는 Box-Jenkins 예측결과[10,11]와 비교한다.

2. 적용적 학습률을 이용한 학습과정

학습과정은 변환함수를 통과한 값과 목표값 사이에 가장 잘 적합되는 가중값을 찾는 과정이며, 올바른 가중값을 어떻게 찾아야 하는가가 학습의 중요한 문제다.

n 번째 예제에 대하여 출력값 (z_n)과 목표값 (t_n)간의 학습오차는 식(1)과 같은 오차판단기준을 이용한다.

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

여기서 N 은 학습 예제들의 전체 개수이고 변환함수는 다음 식(2)와 같다.

$$g(u) = [1 + \exp(-u)]^{-1}, 0 < g(u) < 1 \tag{2}$$

학습절차는 역전파 알고리즘을 이용한다. 각각의 처리절차는 최초에 변환함수를 현재 가중값에 적용하여 계산하고 각 예제에 대한 가중값의 변화량을 결정한다. 그리고 각 예제에 대한 가중값들을 누적시킨 결과를 이용하여 현재 가중값들을 바꿔준다. 이와 같은 과정을 반복하면서 (z_n)과 (t_n)간의 오차가 최소화 되는 가중값들을 구하기 위하여, 각 노드에 대한 가중값의 변화가 없을 때까지 가중값들을 계속적으로 변화시키면서 학습한다.

전체 예제들에 대하여 미분한 값들을 누적시켜서 각 가중값들에 대한 순수 변화량을 구하기 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.

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

여기서 c_m 은 m 번째 학습 후의 가중값 w 의 변화량이고 다음 식(4)와 같다.

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

여기서 음의 부호는 MSE가 감소되는 변화를 나타내준다. e 는 가중값이 변화됨에 따라 최소값으로 수렴되어 가는 비율을 조절하는 모수이며 다음 식(5)와 같다.

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

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

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

누적합이 최근 변화에 더 큰 영향을 받도록 하기 위해서는 다음 식(7)을 적용한다.

$$f_{m+1} = \theta f + (1 - \theta)d_m, 0 \leq \theta < 1 \tag{7}$$

따라서 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(4)와 같은 적응적 학습률을 이용한다.

3. 개선된 학습과정

Lewbel[12]은 시계열자료의 분석이나 신경망을 통한 분석에서 중요시 되고 있는 두 가지 공통 사항으로써, 함수형태의 결정(신경망인 경우에 각 층별 노드의 개수, 층의 개수, 변환함수의 형태등을 결정하는 문제)과 모수의 추정문제(신경망의 가중값)가 중요한 사항이라고 지적 하였다. 이 장에서의 개선된 신경망 학습과정에는 분석할 시계열의 특성에 맞는 적응적 학습률을 구하는 방법을 제시하고 초기값의 동적인 적용을 통한 학습방법을 개선하여 시계열 자료의 예측을 한다.

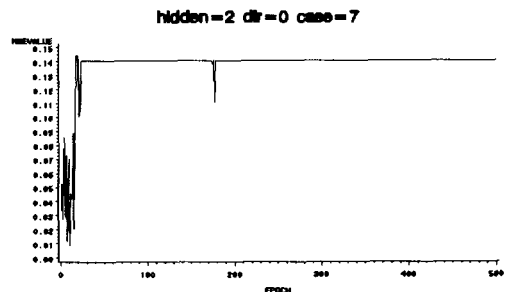
3.1 적응적 학습률

시계열자료의 학습을 위해 적용한 적응적 학습률의 모수를 결정하는 문제는 중요하다. 일반적으로 학습률의 모수를 설정하기 위해서는 수많은 모수들의 조합을 이용한 시행착오를 겪는 학습을 반복 실행하면서 가장 학습효과가 높고 수렴 속도가 빠른 모수들을 선택하게 된다. Smith[2]는 경험적으로 가장 학습효과가 높도록 학습률을 조정하기 위한 모수로 $k=0.1$, $\theta=0.5$, $\theta=0.7$ 그리고 $\mu=0.9$ 를 주고 있다. 그러나 이와 같은 모수들은 출력의 형태가 두 가지 패턴인 신경망의 학습에 적용하면 효과적이지만, 출력의 형태가 다양한 시계열자료의 학습과 예측에는 많은 경우에서 조기 포화점에 빠지게 된다. 따라서 시계열자료에 입력 노드와 은닉노드의 개수를 제한시킨 신경망 모형을 적용할 때에는 자료에 가장 잘 적합되는 모수를 선택하여야 한다.

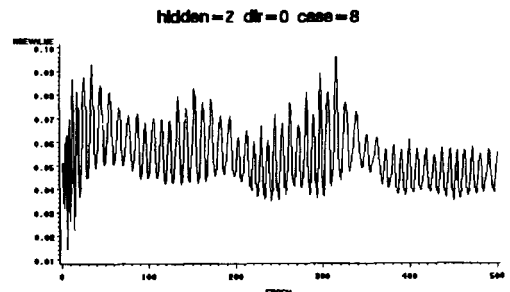
시계열자료를 학습시킬 때에 모든 가능한 모수들의 조합을 이용하여 신경망 학습을 수행하기는 불가능한 일이다. 따라서 적응적 학습률의 모수 k, ϕ, θ, μ 를 각각 제한적인 값들인 0.1, 0.3, 0.5, 0.7, 0.9로 변화시키면서 최적의 모수 조합을 구한다.



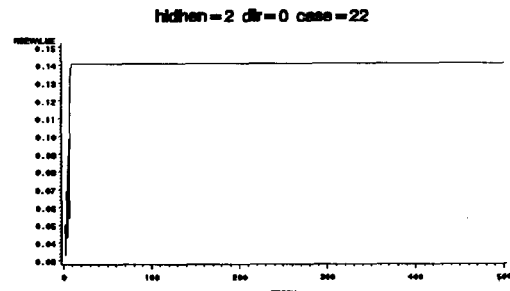
(a) $k=0.3, \phi=0.7, \theta=0.7, \mu=0.7$



(b) $k=0.3, \phi=0.3, \theta=0.5, \mu=0.7$



(c) $k=0.9, \phi=0.1, \theta=0.9, \mu=0.7$



(d) $k=0.9, \phi=0.3, \theta=0.1, \mu=0.9$

(그림 1) 태양흑점자료에 대한 적응적 학습률의 모수를 구하는 과정.

(Fig. 1) Learning process at adaptive learning rates.

본 연구에서 태양흑점 자료에 대한 적응적 학습률의 모수를 결정하기 위한 실험계획으로, MSE가 작고 빠르게 수렴되는 모수조합을 찾기 위하여 표준 직교 배열표에 의해 분류된 25가지의 경우를 각각 500회 학습하여 MSE의 변화를 각각 살펴보면 (그림 1)과 같은 4가지의 주요 패턴으로 분류된다.

(그림 1) (a)는 MSE가 빠르게 감소하면서 수렴될 수 있는 모수이다. 이와 같은 경우는 학습률이 적절한 경우이다. (그림 1) (b)는 수렴되는 초기 학습과정에서 최소값으로 수렴되어 가는 비율을 조절하는 모수 k, ϕ , 의 부적절한 값 적용으로 인한 발산이 나타나고 있다. (그림 1) (c)는 최소값으로의 진행 폭이 너무 커서 최소값 주변에서의 거미집 현상이 나타나고 있다. (그림 1) (d)는 급격한 발산이 나타나고 있다. 따라서 최적 학습률은 각 모수 조합을 이용한 학습 결과에서 MSE가 작고, 빠르게 수렴되는 경우인 $k=0.3, \phi=0.7, \theta=0.7, \mu=0.7$ 를 취한다.

3.2 초기값

시계열자료를 이용하여 신경망 학습을 할 때, 초기값은 중요한 역할을 한다. 일반적으로 신경망 학습에서의 초기값은 난수로부터 생성시킨 특정 초기값을 사용하고 있다.

본 연구에서는 초기값을 결정하기 위한 방법으로써, Easton[9]이 제시한 제어상자(controller box)를 신경망에 도입하여 다차원의 적절한 초기값들을 동적으로 적용시킨다. Easton[9]은 동적 회귀진단기법에 제어상자를 이용하여 회귀모형을 탐색하고 있다. 이 연구에서는 최소제곱법에 의해 구한 p 차 다중회귀계수의 추정값을 오차의 한계 범위 안에서 변화시키면서 적절한 회귀계수를 찾고 있다. 예를들어 $2p$ 의 요인실험인 경우에는 p 개의 회귀계수에 대한 변화를 살펴보기 위해서 p 차원의 도형을 고려한다. 여기서 p 차원의 도형안에 있는 수많은 점들을 효과적으로 2차원 평면에 표현시켜서 평면 내의 좌표를 회귀계수로 취한다.

p 차원 도형의 모서리들과 중점을 연결한 $2p$ 개의 삼각형들로 새로 구성된 2차원의 제어상자는 신경망의 가중값 a_{ij} 와 b_j 에 대한 초기값 설정에 적용시킬 수 있다. 만약에 $2 \times 2 \times 1$ 신경망구조에서는 b_j 가 3차원 도형의 좌표로 표현되고, 여기서 2^3 개의 삼각형들을

통하여 가중값 좌표를 구하면 초기값 설정이 효과적으로 이루어질 수 있게 된다. 학습의 전 단계에 적용될 수 있는 이 방법은 z_n 과 t_n 과의 오차를 작게 해주는 임의의 초기값들을 제어상자 위에서 동적으로 변화시킬 수가 있다. 이렇게 하면 시행착오를 최소한으로 줄일 수 있는 최적의 초기값을 찾을 수 있게 된다. 선택된 초기값을 이용하여 신경망 학습을 하면, 전체 학습시간이 줄어들 뿐만 아니라 효과적으로 시계열자료를 적합시킬 수 있다. 만약에 초기값이 잘못 선택되어 초기 포화점에 수렴되는 경우에는 학습과정 중이라도 제어상자를 통하여 새로운 가중값을 실시간으로 설정하여, 전역최소값으로 수렴될 수 있는 학습을 한다.

4. 실증분석

이 장에서는 앞에서 제안된 신경망의 개선된 학습방법을 이용하여 289년 동안 관측된 Wolf의 태양흑점 자료를 분석한다. 이 자료는 1700년부터 1988년 사이의 태양흑점의 개수를 나타내고 있으며, 시계열 특성은 상승 또는 하강 추세나 확실한 계절성을 관찰할 수 없다. 그러나 순환변동이 있는 시계열이다.

신경망 학습을 위해서 1700년부터 1951년 까지의 252년 간의 자료를 학습자료로 이용하고, 예측을 위한 검정자료는 36년 간의 자료로써 1953년부터 1988년 사이의 자료다. 신경망 학습과 예측에 이용되는 구조는 $2 \times 2 \times 1$ 신경망구조이며, 완전연결구조를 고려한다.

통계적인 방법과 신경망 방법에서의 자료 변환식은 다음 식(8)과 같이 변환된 자료를 이용한다.

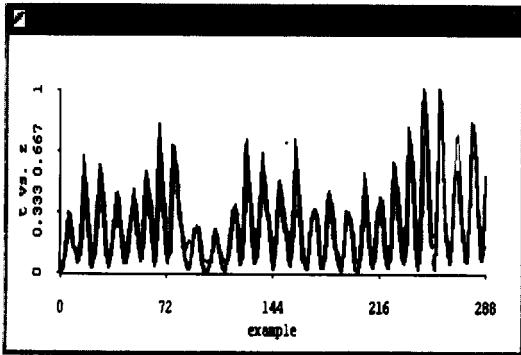
$$z_n = (t_n - t_{(1)}) / (t_n - t_{(1)}), \quad 0 < z_n \leq 1 \quad (8)$$

여기서 t_n 은 n 번째 태양흑점자료, $t_{(1)}$ 은 최소값 그리고 $t_{(n)}$ 은 최대값이다.

$t+1$ 시점의 신경망 예측은 2개 색으로 제한한 z_t 와 z_{t-k} 를 입력자료로 이용하여 z_{t+1} 을 예측한다. 여기서 태양흑점 자료의 순환변동을 고려하기 위해 시차 k 를 변화시킬 수 있다. $2 \times 2 \times 1$ 신경망 구조하에서 입력 시계열의 변동구조를 살펴 보면서 순환변동 주기를 고려할 수 있는데, 시계열이 상승 또는 하강 추세가 확실하면 k 를 1로 주면서 학습하고 확실한 계절성이나 순환주기가 있으면 k 에 반영시켜서 학습할 수 있다. 여기서는 11년의 순환변동 주기를 고려한 경우

와 그렇지 않은 두 경우로 나누어 예측한다.

(그림 2)는 태양흑점자료를 이용하여 300회의 신경망 학습을 하고, 36년 예측을 한 결과이다. 여기서 태양흑점자료는 굵은 실선으로 표시되어 있고 학습자료(0~252)와 검정자료(253~288)로 분리된다. 가는 실선은 학습한 결과와 예측결과를 나타내고 있다.



(그림 2) 신경망 학습을 300번 수행한 후에 36년 예측한 그림

(Fig. 2) 36 years forecast after 300 epochs

(그림 2)와 같은 학습 결과로 변화된 최종 가중값인 은닉노드에 대한 가중값 a_{ij} 와 출력노드에 대한 가중값 b_j 는 다음과 같다.

$$A = \begin{bmatrix} 3.7274 & 0.5627 \\ -3.3242 & -7.1459 \\ -0.4862 & -2.0249 \end{bmatrix}, b = \begin{bmatrix} 4.4907 \\ -5.3067 \\ -3.5308 \end{bmatrix}$$

이 가중값을 다시 순환 신경망 모델에 대입하여 1년, 3년, 6년, 12년, 24년 그리고 36년간의 신경망 예측을 하고, 그 예측결과를 Box-Jenkins 예측결과와 비교해 보면 <표 1>과 같다. 여기서는 $2 \times 2 \times 1$ 신경망의 시차를 1과 11로 구분하였을 때의 MSE를 각각 통계적 분석결과와 비교하였다. MSE가 상대적으로 작은 경우에는 각각 †와 ‡로 표시한다.

통계적 방법은 SAS를 이용 하였고 모형 식별 및 모수 추정 결과는 다음 식 (8)과 같은 AR(2) 모형이다.

$$z_t = 0.146 + 1.415z_{t-1} - 0.720z_{t-2} + a_t \quad (9)$$

신경망의 예측결과만 살펴볼 때 <표 1>에서 순환변동 주기를 고려한 시차가 11인 경우는 모든 예측기간에서 시차가 1인 결과보다 MSE가 작다. 신경망의 예측결과를 통계적 예측결과와 비교할 때, 신경망은 시

차가 11인 경우에 1년과 6년을 제외한 다른 예측기간에서 통계적 예측결과보다 MSE가 작다. 시차가 1인 경우에는 순환변동을 제대로 반영하지 못하여 모든 기간에서 통계적인 방법보다 MSE가 더 크다.

<표 1> 태양흑점자료에서의 예측오차(MSE)
(Table 1) MSE at sunspot data

예측 기간	통계적 방법	$2 \times 2 \times 1$ 신경망	
		lag=1	lag=11
1년	0.00021†	0.00822	0.00078
3년	0.00182	0.10005	0.00108†
6년	0.00117†	0.20559	0.01038
12년	0.00573	0.21324	0.00567‡
24년	0.00992	0.23997	0.00784‡
36년	0.01424	0.22248	0.01302‡

5. 결 론

본 연구에서는 신경망 모델을 이용하여 태양흑점 시계열자료를 학습할 때 자주 발생될 수 있는 조기 포화점 문제와 초기값 문제의 개선 방안을 논의하였다. 먼저 효과적인 적응적 학습률을 구하기 위하여 표준적교 배열표에 의해 분류한 25가지 경우에 대한 개별 학습을 통하여 가장 적절한 적응적 학습률을 구했다. 초기 가중값은 단순히 난수에 의존하지 않고 시각적으로 동적인 선택과정을 통하여 실시간으로 선택하면서 학습하였다. 이와 같은 과정을 통하여 조기 포화점에 빠지는 수많은 시행착오를 줄이면서 학습시간을 단축시킬 수 있었다. 11년의 순환변동 주기를 갖고 있는 태양흑점 자료의 특성을 반영하고, 노드의 개수를 제한한 신경망 $2 \times 2 \times 1$ 구조의 효율을 높이기 위해서는, $t+1$ 시점을 예측하기 위하여 t 시점과 $t-1$ 시점뿐만 아니라 t 시점과 $t-k$ 시점으로 입력 노드값을 변화시키면서 학습을 하였고 그 결과를 통계적인 결과와 비교하였다.

본 연구에서 고려한 예측기간에서는 신경망이 Box-Jenkins 예측결과보다 효율적이었다. 그러나 변환함수의 특성으로 인하여 장기예측에서는 직선으로 수렴하게 되는데, 다양한 변환함수를 적용할 필요가 있다. 그리고 본 연구에서는 $2 \times 2 \times 1$ 신경망구조만을 고려하였

는데, 시계열자료의 학습을 위한 적절한 신경망 구조를 결정할 수 있는 연구가 필요할 것으로 보이고, 태양흑점 자료에 대한 비선형 시계열 분석 결과와 비교해 볼 필요가 있다.

참 고 문 헌

[1] A.P.Russo, "Neural Networks for Sonar Signal Processing," IEEE Conference on Neural Networks for Ocean Engineering, Tutorial Nn. 8, 1991.
 [2] M.Smith, "Neural Networks for Statistical Modeling", New York, Van Nostrand Reinhold, 1993.
 [3] B.D.Riply, "Neural Networks and Related Methods for Classification," Journal of Royal Statistical Society, B, 56(3), pp.409-456, 1994.
 [4] B.D.Riply, "Pattern Recognition and Neural Networks", Cambridge University Press, 1996.
 [5] H.Kushner, "Asymptotic Global Behaviour for Stochastic Approximation and Diffusions with Slowly Decreasing Noise Effects: Global Minimization via Monte Carlo," SIAM Journal of Control Optimization, 47, pp.169-185, 1987
 [6] H.White, "Learning in Artificial Neural Networks : A Statistical Perspective," Neural Computation, 1, pp.425-464, 1988.
 [7] 윤여창, "시계열자료에서 신경망이론의 적용", 성균관대학교 박사학위논문, 1995.
 [8] 윤여창, 허문열, "Neural Network을 응용한 Time Series", 통계와 신경망, 한국과학기술원 인공지능연구센터/한국정보과학회 뉴로컴퓨팅연구회/한국통계학회 통계계산연구회주최 공동워크샵, 연세대학교 소프트웨어응용연구소 주관, pp49-69, 1996.
 [9] G.S.Easton, "A Simple Dynamic Graphical Diag-

nostic Method for Almost Any Model," Journal of the American Statistical Association, 89, pp.201-207, 1994.

[10] G.E.P.Box, G.M.Jenkins, "Time Series Analysis: Forecasting and Control", Holden-Day, 1976.
 [11] H.Tong, "Non-linear Time Series: A Dynamical System Approach", Oxford, Clarendon press, 1990.
 [12] A.Lewbel, "Comments on Artificial Neural Networks : An Econometric Perspective," Econometric Reviews, 13(1), pp.99-103, 1994.



윤 여 창

1986년 성균관대학교 통계학과 졸업(학사)
 1990년 성균관대학교 대학원 통계학과 졸업(석사)
 1995년 성균관대학교 대학원 통계학과 졸업(박사)

1995년~1997년 우석대학교 전산통계학과 전임강사
 1997년~현재 우석대학교 전산통계학과 조교수
 관심분야 : 신경망이론, 시계열분석, 컴퓨터그래픽스



이 성 덕

1982년 성균관대학교 통계학과 졸업(학사)
 1989년 (미)조지아대학교 통계학과 졸업(석, 박사)
 1991년~1993년 충북대학교 통계학과 전임강사

1993년~1997년 충북대학교 통계학과 조교수
 1997년~현재 충북대학교 통계학과 부교수
 관심분야 : 시계열분석, 통계자료분석, 실험계획법