

# 신경망을 이용한 시계열 분석 : M1-Competition Data에 대한 예측성과 분석

지원철\*

## Time Series Analysis Using Neural Networks : Forecasting Performance Analysis with M1-Competition Data

요 약

Neural Networks have been advocated as an alternative to statistical forecasting methods. However, the empirical evidences are not consistent. In the present experiments, multi-layered perceptron (MLP) are adopted as approximator to the time series generating processes. To prevent the learned MLP from being overfitted to the given time series, the information obtained from ARMA modeling is used to determine the architecture of MLP.

The proposed approach was tested empirically using the subsamples of the 111 time series used in the first Markridakis Competition. The forecasting results were analyzed to find out the factors that affect the performance of MLP. The experimental results show that the proposed approach outperforms ARMA models in terms of fitting and forecasting accuracy. In addition, it is found that the use of deseasonalized data improves the forecasting accuracy of MLP.

주제어 : 시계열분석, 신경망, ARMA모형, 과잉적합(overfitting)

### I. 서 론

신경망이 시계열분석에 이용될 수 있음이 알려진 후 (Lapedes 와 Farber, 1987), 신경망은 기존의 통

계적 예측방법론들을 대체할 수 있는 유력한 수단으로 주목받아 왔음에도 불구하고 사회경제 시스템으로부터 얻어진 시계열 자료들에 대한 실증적 분석들이 모두 일치된 연구결과를 보이지는 않았다. Werbos (1988), Weigend 등 (1990), Tang 등 (1991), Sha-

\* 홍익대학교

rda와 Patil (1992), Foster 등 (1992) 및 Hopcroft (1993) 등의 연구결과는 신경망을 이용해 좋은 결과를 얻었지만, White (1988), Fishwick (1989) 및 Ripley (1993) 등의 연구결과에서는 신경망이 기존의 통계적 방법들에 비해 더 좋은 예측결과를 보이지 않았다. 이와 같은 실증적 연구결과들의 불일치에 대해 Chatfield (1993)는 신경망이 새로운 유력한 예측수단임에 틀림없지만 사회경제 시스템의 예측에 이용되기 위해서는 아직도 체계적인 입증노력이 더 필요하다고 하였다.

기존 연구에 사용된 신경망 모형은 대부분 다계층 퍼셉트론(Multi-layered Perceptron, MLP)이었는데, MLP가 여러 응용분야에 적용되어 좋은 성과를 보였고 또 가장 폭넓게 이용되고 있는 신경망 모형이라는 점을 반영하는 것이다. 하지만 같은 신경망 모형을 사용했음에도 MLP의 예측성가에 대해 일치된 연구결과를 얻지 못한 이유로는 우선 분석대상이 되었던 사회경제 시스템들이 계속적으로 변화하고 외부로부터 받은 영향을 받고 있어 시계열 생성과정 자체가 변화되거나 비정상적인 관측치들이 시계열내에 많이 존재하기 때문으로 볼 수 있다. 즉 사회경제 시스템으로부터 얻은 대부분의 시계열들은 (1) 상당히 제한된 수의 관측치만을 가지며, (2) 시계열내의 잡음이 심하여 이를 단순한 임의변동이라고 간주하기 어렵고, (3) 기존의 시계열 분석방법들이 채택하고 있는 안정적 시계열이라는 가정이 성립되지 않는 경우가 많아서 단일변수 시계열 분석(univariate time series analysis)을 사용하여 좋은 예측치를 얻는 것을 어렵게 한다.

사회경제 시스템의 예측에서는 필요에 따라 원하는 만큼의 관측치를 얻기 위한 추가적인 실험자체가 불가능한 경우가 많고 또 오래된 관측치들은 시계열 분석에 도움이 되지 않아 사용가능한 관측치의 수는 많아야 200개 내외이고 50개 미만의 관측치만을 갖는 시계열들도 적지 않다. 따라서 상대적으로 매우 적은

관측치들을 가지고 심한 잡음과 추세나 계절성 등의 비안정적인 요소들이 내재된 시계열을 예측하기 위해서는 사전 자료분석의 중요성이 강조되고 있으며 복잡한 모형보다는 간단한 선형모형 또는 여러 개의 간단한 모형들로 부터 얻은 예측치들을 결합해 최종 예측치를 얻는 방법이 유효하다고 주장되기도 한다 (Markridakis, 1983, 1992).

하지만 최근 사회경제 시스템의 예측에 있어서도 비선형 모형에 의해 좋은 예측성가를 얻을 수 있다는 연구결과들 (Tong, 1990; De Gooijer 와 Kumar, 1992)을 감안할 때 신경망은 유력한 예측방법의 하나임에 틀림없지만 신경망이 성격상 추정해야할 파라메타의 수가 너무 많다는 점을 고려해야 한다. MLP의 응용 성과는 선택된 MLP의 구조, 즉 입력층 및 은닉층의 처리요소들의 수에 의해 많은 영향을 받음에도 불구하고 기존의 연구에서는 주어진 시계열에 대해 적합한 MLP 구조를 찾으려는 체계적인 노력이 보이지 않았다. MLP의 크기가 커지면 추정해야할 파라메타들인 연결가중치(connection weights)들이 너무 많아져 주어진 시계열내의 무의미한 잡음까지도 학습하여 MLP의 예측력이 손상될 수 있으며, MLP의 크기가 너무 작으면 학습된 MLP가 주어진 시계열을 제대로 설명하지 못하게 된다.

따라서 적절한 크기의 신경망을 사용하는 것이 중요하다. 특히, 신경망의 크기를 결정하는 첫번째 요인이 입력층 처리요소들의 수라는 점과 단일변수 시계열 예측의 성공여부는 채택된 방법론이 주어진 시계열이 갖는 특징들을 얼마나 잘 반영하는가에 달려 있다는 점을 고려할 때, 적절한 크기의 입력공간을 선택하는 문제는 MLP 응용의 성공여부를 좌우한다고 볼 수 있다. 즉 1년 단위의 계절성을 갖는 시계열 자료에 대해 6개월 동안의 자료만을 입력으로 사용하는 예측모형을 구성했다면 이 예측모형으로부터 좋은 예측치를 얻기는 힘들 것이다. 이와 같이 적절한 크기의 입력공간을 반영하는 신경망을 사용하기 위

해서는 사전에 주어진 시계열을 철저히 분석할 필요가 있음에도 불구하고 Weigend 등 (1990)을 제외한 기존의 연구들에서는 입력공간에 대한 충분한 고려없이 MLP의 크기를 임의로 결정하여 사용하였기 때문에 신경망의 예측성과에 대해 일치되지 않은 연구결과들을 가져왔다고 생각된다.

이상의 논의를 바탕으로 단일변수 시계열 분석에 있어 아직도 검증이 필요한 사항들을 정리해 보면 다음과 같다.

- (1) 시계열 내의 관측치 수가 적을 경우에도 MLP를 이용한 예측이 유효한가?
- (2) 추세 및 계절성 등 시계열이 갖는 특성이 MLP의 예측성과에 영향을 미치는가?
- (3) 좋은 예측성과를 얻기 위한 MLP의 크기 및 구조는 어떻게 결정되어야 하는가?
- (4) MLP의 학습절차는 예측성과에 어떤 영향을 주는가?
- (5) 사회경제 시스템의 예측에 있어서도 비선형 모형이 선형모형에 비해 우월한 예측성과를 보일 수 있는가?

본 논문의 목적은 이러한 사항들을 가능하면 많은 실세계의 시계열을 대상으로 검증하는 것으로 제2장에서는 연구에 사용된 시계열 자료인 M1-Competition 자료 및 연구방법에 대해 설명하고, 제3장에서는 MLP의 예측성과를 비교분석하는데 사용된 선형모형인 자기회귀 이동평균 모형 (Autoregressive Moving Average Model, ARMA)에 대해 설명한다. 제4장에서는 ARMA 모형화의 결과를 이용하여 MLP의 구조를 결정하는 문제와 MLP의 학습절차에 대해 설명한 후 제5장에서 111개의 M1-Competition 자료에 대한 MLP의 예측성과를 관측치의 수 시계열의 특성 등을 고려하여 분석한다.

## II. 데이터 및 분석방법

단일변수 시계열 분석을 위해 많은 예측기법들이 6, 70년대 부터 개발되어 발전해 왔다. Markridakis는 80년초 실세계의 다양한 출처로부터 1001개에 이르는 시계열들을 수집한 후, 여러 예측기법들의 정확성을 비교하기 위한 경쟁대회를 주관하였다. M1-Competition이라고 불리우는 이 경쟁대회에 많은 예측전문가들이 참여했는데 이들은 시계열의 출처나 성격 등에 관한 사전지식을 전혀 제공받지 못하고, 오로지 주어진 시계열 자료만을 대상으로 자신들이 적용할 예측기법에 관한 전문지식을 총동원하여 예측치를 작성하였다. Markridakis는 자신이 유보하고 있던 실제 관측치들을 제출된 예측치들과 비교하여 총 24가지 예측기법들의 정확도를 분석 발표하였다. Markridakis에 의해 유보된 각 시계열내의 관측치들의 숫자는 연간 시계열의 경우 6개, 분기별 시계열은 8개, 월간 시계열은 18개이었으며, 예측의 정확도는 평균 절대오차비율(Mean Absolute Percentage Error, MAPE)에 의해 측정되었다. M1-Competition의 결과는 전문가들 사이에 많은 논란을 불러 일으켜, 최근 Markridakis는 29개의 시계열을 새로이 준비해 제2차 대회인 M2-Competition을 개최하였다 (Markridakis 등, 1982, 1993).

M1-Competition에 사용되었던 시계열 자료들은 그 후 예측방법론들을 실증적으로 검증하는데 매우 유용한 자료로 인식되어 왔는데 특히 원래의 데이터 베이스로부터 연간자료 20개, 분기별 자료 23개 및 월간자료 68개로 축소된 111개의 시계열 자료들이 많은 연구에서 이용되어 왔다 (Foster, 1992; Sharda와 Patil, 1992). 본 연구에서도 이 111개의 시계열에 대해 M1-Competition에서 예측의 정확성을 측정하기 위해 사용하였던 분석방법을 그대로 채택하여 MLP의 예측성과를 분석하고 이를 자기회귀 이동평균 모형과

비교검토 하였다.

주어진 시계열  $\{Z_t, t \in T\}$ 에 대한 예측치를 구하기 위해서 반복적으로 단단계 예측치 (iterative one-step-ahead forecast)를 얻는 절차를 사용하였다. 즉 과거의 관측치를 이용하여 예측함수  $f$ 를 추정한 후,  $f$ 를 이용하여 예측치  $\hat{Z}_{t+r}$ 을 하나씩 차례로 얻어가는 것으로 이를 수식으로 표현하면 다음과 같다.

$$\begin{aligned} \hat{Z}_{t+1} &= f(Z_t, Z_{t-1}, \dots) \\ \hat{Z}_{t+2} &= f(\hat{Z}_{t+1}, Z_t, Z_{t-1}, \dots) \\ &\vdots \\ \hat{Z}_{t+r} &= f(\hat{Z}_{t+r-1}, \hat{Z}_{t+r-2}, \dots, \hat{Z}_{t+1}, Z_t, Z_{t-1}, \dots) \end{aligned}$$

여기서  $\{\hat{Z}_{t+r}, r \in L\}$ 은 예측기간  $L$  동안의 예측치를 의미한다. 본 연구에서는 예측함수  $f$ 를 MLP와 ARMA모형에 의해 추정되었으며 추정된 함수  $f$ 가 과거의 관측치를 얼마나 잘 설명하는가를 측정하기 위하여 다음과 같은  $R^2$ 를 사용하였다.

$$R^2 = 1 - \frac{\sum_{t \in T} (\hat{Z}_t - Z_t)^2}{\sum_{t \in T} (Z_t - \bar{Z})^2}$$

여기서  $Z$ 와  $\hat{Z}_t$ 는  $\{Z_t, t \in T\}$ 의 평균 및 추정함수  $f$ 의 출력값을 나타낸다. 한편 예측기간  $L$ 동안 얻어진 예측치의 정확도를 측정하기 위해서는 다음과 같이 정의 되는 평균절대오차비율 (MAPE)를 사용하였다.

$$MAPE = \frac{1}{L} \sum_{r=1}^L \frac{|\hat{Z}_{t+r} - Z_{t+r}|}{Z_{t+r}} \times 100$$

### III. 자기회귀 이동평균 모형

자기회귀 이동평균 모형 (Autoregressive Moving Average Model, ARMA)은 이론적 정교함과 단기예측의 우수성을 인정받고 있어, 실제 예측작업에 많이 사용되고 있을 뿐만 아니라, 새로운 시계열 예측방법

론을 개발함에 있어서 개발된 모형의 성능을 평가하기 위한 벤치마크 모형으로도 많이 채택되고 있다. 자기회귀 과정의 차수가  $p$ 이고 이동평균 과정의 차수가  $q$ 인 ARMA( $p, q$ ) 모형의 수식은 다음과 같이 표현된다.

$$\Phi(B)Z_t = C + \theta(B)a_t,$$

where  $\Phi(B) = U(B)\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  
 $U(B) = 1 - \phi_1 B - \dots - \phi_d B^d$ ,  
 $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ,  
 $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ ,  
 $BZ_t = Z_{t-1}$   
 $C$  is a constant  
 $a_t$  is a gaussian white noise process.

위 식들에서 만약  $U(B) = (1-B)^d$ 로 표현된다면 ARMA( $p, q$ ) 모형은 Box-Jenkins 접근방법에서의 ARIMA( $p-d, d, q$ ) 모형이 된다. 일반적인 ARMA모형의 수립절차는 (1) 자기회귀 과정 및 이동평균 과정의 차수( $p, q$ )를 정하는 모형 식별단계, (2) 선택된 차수( $p, q$ )에 근거해 모형의 파라메타인  $\phi$ 와  $\theta$ 들을 추정하는 파라메타 추정단계 및 (3) 추정된 모형이 주어진 시계열을 얼마나 정확히 묘사하는지를 검증하는 모형 검증단계의 세 단계를 거친다. ARMA 모형화의 세 단계 중에서도 모형 식별단계가 대단히 중요한데, 이는 만약 모형수립에 있어 주어진 시계열 자료에의 적합성 (fitness) 여부에만 중점을 두어 너무 많은 파라메타들을 사용할 경우, 추정된 ARMA모형의 예측력이 떨어질 수 있기 때문이다. 따라서 같은 적합성을 갖는 하나 이상의 ARMA( $p, q$ ) 모형들이 존재할 경우에는 가장 낮은 차수들을 갖는 모형, 즉 파라메타들의 개수가 가장 적은 모형을 선택한다는 ‘절제의 원리 (Principle of Parsimony)’가 ARMA모형 수립에 있어 중요한 모형수립지침으로 적용되고 있다.

ARMA모형의 식별단계에서는 Autocorrelation Function(ACF)와 Partial Autocorrelation Function(PACF) (Box 와 Jenkins, 1976), Akaike Information

Criterion (Akaike, 1974) 등이 사용되는데, 이 통계량들을 해석하기 위해서는 상당한 전문지식이 요구되고 최종차수(p,q)를 선정하기 까지는 앞서 언급한 ARMA모형 수립의 세 단계를 반복적으로 수행하는 것이 불가피하다. 이와 같은 ARMA모형 식별단계의 어려움을 해결하기 위해 모형수립 과정을 자동화하려는 많은 노력들이 있어 왔는데 이에 대한 상세한 내용은 Jhee 와 Lee (1993)를 참조하기 바란다. 최근 ARMA모형을 자동적으로 수립해 주는 상용 소프트웨어들 중의 하나로 Box-Jenkins 방법에 근거한 AUTOBOX 3.0이 많이 사용되고 있는데 (AFA Inc, 1991), 이 소프트웨어의 ARMA모형 수립기능은 전문가의 능력에 상응하는 것으로 평가받고 있다(Sharda 와 Patil, 1992). 본 연구에서도 1차적으로 AUTOBOX 3.0을 사용하여 M1-Competition 자료 중 111개의 시계열을 모형화하고 예측치를 구했는데, 이때 얻어진 ARMA 모형의 식별에 관한 정보 즉 자기회귀 및 이동평균 과정의 차수(p,q)를 MLP의 구조결정에 사용하였으며, AUTOBOX 3.0을 통해 얻은 예측치들은 MLP를 통해 얻은 예측치들을 시계열이 갖는 특성에 따라 분류하여 비교 평가하는데 이용하였다.

## IV. 다계층 퍼셉트론

### 4-1. 신경망을 이용한 시계열 분석

신경망은 다양한 응용분야에 적용되어 성공적인 결과를 보여 왔는데, 최근 문자인식, 음성인식, 제어 분야 등의 공학적 응용외에도 경제예측, 재무분석, 신용조회, 기업도산 예측 및 주가예측 등의 광범위한 경영 경제분야에서도 응용되고 있다. 신경망을 예측 업무에 응용하는 방법에는 두가지를 생각할 수 있다. 첫번째는 신경망을 간접적으로 이용하는 방법으로

이는 다시 (1) 몇개의 예측기법을 사용하여 얻은 예측치들을 결합하여 향상된 예측치를 얻는 과정에서 신경망을 여러 예측치들을 결합하는 가중치를 얻는 수단으로 이용하는 방법(Foster 등, 1992)과, (2) 신경망의 패턴인식 능력을 활용하여 주어진 시계열내에 존재하는 시계열 패턴들을 파악하고 이를 입력으로 사용하여 기존의 예측기법들 중에서 가장 적절한 예측모형을 선정하는 방법 (Chu 와 Widjaja, 1994; Lee 와 Jhee, 1994)의 두가지를 생각 할 수 있다.

두번째는 (그림 1)과 같이 시계열 자료를 바로 입력으로 받아 신경망으로부터 직접 예측치를 얻는 방법이다. (그림 1-a)에서 볼 수 있드시 n개의 입력공간에 의해 표현된 시계열 관측치들  $\{Z_{t-1}, Z_{t-2}, \dots, Z_{t-n}\}$  은 MLP에 의해 t 시점에서의 예측치  $\hat{Z}_t$  로 대응된다. MLP의 학습을 위해서는 (그림 1-b)의 비용함수 즉 실제 관측치 $Z_t$ 와 추정치  $\hat{Z}_t$ 의 차이를 최소화시키는 것이 필요한데 이를 위해서는 MLP 학습에 가장 많이 사용되는 백프로퍼게이션 알고리즘을 수정없이 사용하는 것이 가능하다. 이때 학습자료 집합의 구성은 (그림 1-c)에서와 같이 한 단위 시간씩 움직여 가며 입출력값들을 대응시켜 학습자료를 구성하는 슬라이딩 윈도우 기법이 사용된다. 한편, 다단계 예측치 (Multi-step-ahead forecasts)를 한번에 얻기 위해서는 (그림 1-a)에서 출력층 처리요소들의 개수를 원하는 만큼 늘리면 되는데 기존 연구결과에 의하면 출력층의 처리요소를 하나만 사용하여 축차적으로 예측치를 얻는 것이 더 좋은 예측성과를 보인다고 알려져 있다 (Sharda 와 Patil, 1992).

본 연구에서도 (그림 1-a)에서와 같은 MLP의 구조를 그대로 사용하고 있다. 즉 출력층의 처리요소는 하나만 사용되었고 은닉층의 층의 수도 하나만 사용되었으며 시그모이드 (Sigmoid) 함수가 처리요소내의 전이함수로 사용되었다. Hornik 등 (1989)의 연구에 의하면 MLP는 비록 하나의 은닉층만을 가진다 해도 어떠한 함수에 대해서도 자신을 근사시킬 수 있기

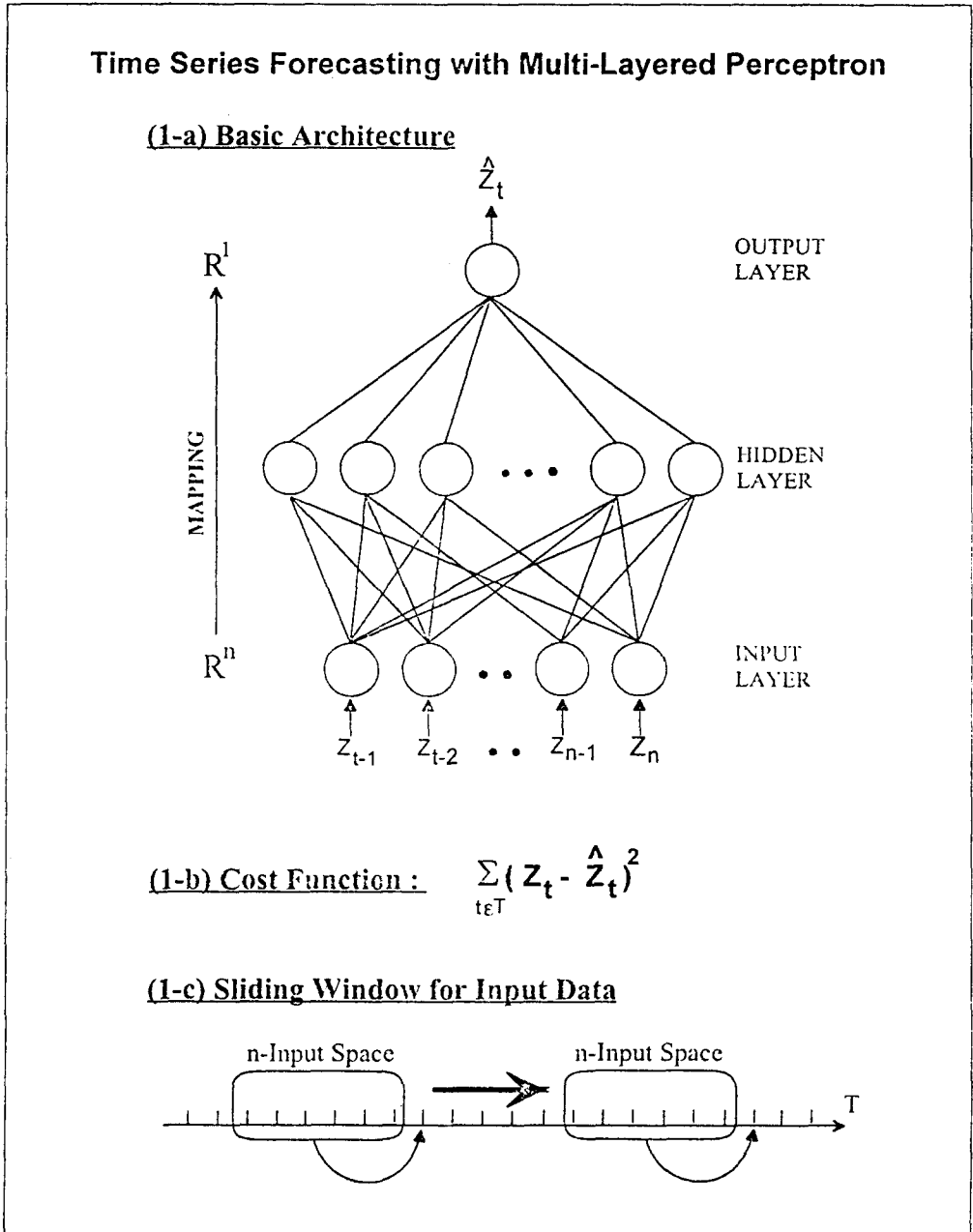


그림 1. 다계층 퍼셉트론을 이용한 시계열 분석

때문에 만능 근사기(Universal Approximator)로 불리운다. 하지만 앞서 ARMA모형에서의 '절제의 원리'와 마찬가지로 주어진 시계열에 잘 부합되었다고 반

드시 좋은 예측치를 얻는다는 보장은 없다. MLP가 과거의 관측치들을 최대한 설명할 수 있도록 하기 위해서는 큰 규모의 MLP를 사용하여 시계열내에 존

재하는 정보들을 가능한 많이 추출해 내도록 하는 것이 바람직해 보이지만, 이 경우 MLP는 유용한 정보 이외에도 시계열내에 있는 많은 잡음까지도 기억하게 되어 예측성능이 떨어지는, 즉 과잉적합(overfitting) 문제를 발생시킬 수 있다. 이 과잉적합문제의 해결은 잡음이 심하고 시계열내의 관측치가 적은 경우의 단일변수 시계열분석에 있어서는 매우 중요한 문제로 지적되고 있다 (Weigend, 1990). 따라서 적절한 규모의 MLP를 사용하는 것이 학습시 MLP가 주어진 시계열에 대해 과잉적합될 위험을 줄이고 좋은 예측치를 얻는 방법임에도 아직까지 이에 대한 구체적인 연구결과들이 없었다.

신경망의 많은 공학적 응용분야에 있어서 입력벡터의 크기에 대해서는 크게 신경을 쓰지 않는데 이는 응용문제에 의해 입력벡터의 크기가 사전에 결정되기 때문이다. 하지만 사회경제 시스템에의 응용에서는 선택된 모형에 상관없이 입력으로 사용될 변수들의 선택 및 입력벡터의 크기 결정이 더 중요한 문제로 인식되고 있는데, 이는 최근 신경망의 성공적인 응용을 위해서는 학습자료 집합을 구성하기 전에 입력자료에 대한 충분한 검토가 필수적이라는 주장과 일맥상통하는 것이다 (Caudill, 1989). 특히 입력벡터의 크기 결정은 단일변수 시계열 분석에서는 더욱 중요한 문제로 대두되는데 이는 시계열의 특성을 반영하지 못하는 입력공간으로는 좋은 예측치를 얻을 수 없기 때문이다. 본 연구에서는 ARMA모형 식별단계에서 파악된 ARMA모형의 차수( $p, q$ )를 이용하여 입력벡터의 크기를 정하는데, 이는 ARMA모형의 식별이 상당한 통계적 분석에 의하여 주어진 시계열의 특성을 잘 나타내는 ( $p, q$ )를 선택하기 때문에 신경망의 구조결정에 앞서 필요한 사전 자료분석과정을 대체해주는 것으로 생각할 수 있다.

## 4-2. 학습절차

111개의 시계열 자료마다 각기 다른 구조를 갖는 MLP가 설계되었으며 각 MLP들은 해당 시계열에 대해 (그림 1)에서와 같은 방법으로 백프로퍼게이션 알고리즘을 사용하여 학습시켰다. 시계열 자료들은 MLP에 입력되기 전에 0.1과 0.9사이의 값을 갖도록 선형변환되었으며 비안정적 시계열들을 처리하기 위한 추세나 계절성의 제거 작업은 하지 않았다. 학습시 사용된 학습멈춤규칙은 주어진 시계열로부터 MLP가 가능한 많은 정보를 이끌어 내도록 하기위해 학습오차 제곱의 평균(mean square error, 그림 1-b 참조)이 0.0001에 이르거나, 학습회수가 최대 허용회수인 5000번에 이르면 학습을 끝내도록 하였다. 이는 매우 간단한 학습멈춤규칙으로 MLP가 주어진 시계열에 대해 과잉적합될 위험성을 내포하고 있지만, 본 연구에서는 사전 자료분석의 역할을 담당해 주는 ARMA모형 식별정보를 이용하여 MLP의 크기를 제한함으로써 과잉적합의 위험은 크지 않다고 생각한다. 또 예측기간 동안 미래관측치들의 값을 사전에 알 수 없는 실제 예측작업에 있어서는 가장 좋은 예측결과를 보이는 학습회수를 알 수 없다는 사실을 고려하면 채택된 학습멈춤규칙은 실제 적용가능성이 매우 높은 규칙이라 할 수 있다.

학습 파라메타 및 연결 가중치(connection weight)의 조절간격을 결정하기 위해 111개의 시계열들을 추세와 계절성을 갖는가에 따라 4그룹으로 분류한 후 각 그룹으로부터 3개의 시계열들을 뽑아 모두 12개의 시계열을 사용해 예비실험을 하였다. 학습률과 관성항의 계수는 (0.1, 0), (0.1, 0.1), (0.1, 0.9) 및 (0.3, 0.5)의 네가지 조합을 사용했는데 (0.1, 0.1)을 사용했을 때 다른경우보다 예측결과가 약간 좋았다. 연결가중치의 조절간격에 대해서는 (1) 입력자료가 바뀔때마다 조절하는 경우 (2) 5회 입력후 조절하는 경우 및 (3) 시계열내의 모든 관측치가 한번 입력된 후 조절하는

경우의 세가지를 시험했는데, 입력자료가 변경될 때마다 연결가중치를 조절하는 것이 가장 좋은 예측치를 얻을 수 있었다. 일단 학습이 끝나면  $R^2$ 를 구하여 학습된 MLP가 얼마나 과거의 관측치를 잘 설명하는가를 측정하였고, 유보된 관측치들에 대한 예측치를 얻은 후 이를 원래 가졌던 크기의 값들로 재 변환한 후 MAPE를 구했다.

### 4.3. 다계층 퍼셉트론의 구조결정

입력층의 처리요소들의 개수를 결정하기 위하여 ARMA(p,q) 모형은 크기  $n = \max(p, q + 1)$ 을 갖는 상태벡터에 의한 상태공간 방정식(state-space form)으로 표현가능하고 (Harvey, 1984), 또 상태공간 방정식은 신경망에 의해 묘사 가능하다는 사실을 이용하였다. 따라서 3장에서 설명하였듯이 본 연구에서는 ARMA모형의 추정을 위해 Box-Jenkins방법에 근거한 AUTOBOX 3.0을 사용하므로 모형의 식별은 ARIMA(p,d,q)의 형태로 이루어진다. 여기서 differencing 차수를 나타내는 d는 자기회귀과정의 차수에 더해질 수 있으므로 주어진 시계열에 사용될 MLP의 입력층 처리요소의 수는  $n = \max(p + d, q + 1)$ 에 의해 결정하였다. 이와같은 접근방법의 타당성을 검증하기 위하여 우선 111개의 M1-Competition자료중 앞 절에서 설명

한 12개의 시계열에 대해 예비실험을 하였고, 이 과정에서 은닉층 처리요소 수의 결정에 관한 문제도 검토하였다.

(그림 2, 3, 4)는 예비실험에 사용되었던 시계열 들 중의 하나인 S184의 분석 결과를 나타낸 것이다. (그림 2)의 S184는 AUTOBOX 3.0에 의해 ARIMA(1,4,4)로 식별되었는데, 이는 ARMA(5,4)와 같으므로 MLP의 입력벡터 크기는 5로 결정되어야 한다. (그림 3)은 입력벡터의 크기에 따른 MLP의 예측능력을 도시한 것으로 우측 Y축에 표시된  $R^2$ 의 값은 1에 가까울수록 기존의 관측치들을 잘 설명하는 것이고, 좌측 Y축에 표시된 MAPE 값은 적을수록 MLP의 훈련에 사용되지 않고 유보된 관측치에 대한 예측의 정확도가 높음을

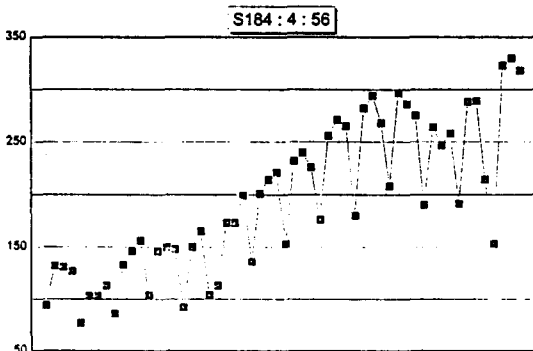


그림 2. 시계열의 예 : S184 (S184는 56개의 관측치를 가진 분기별 시계열이다)

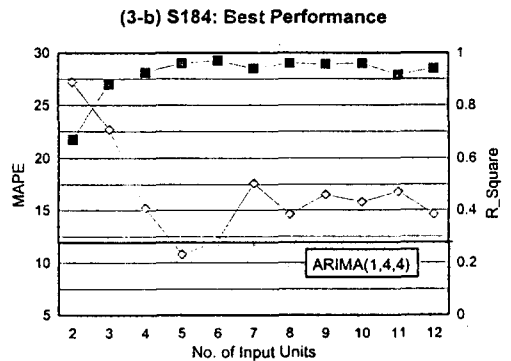
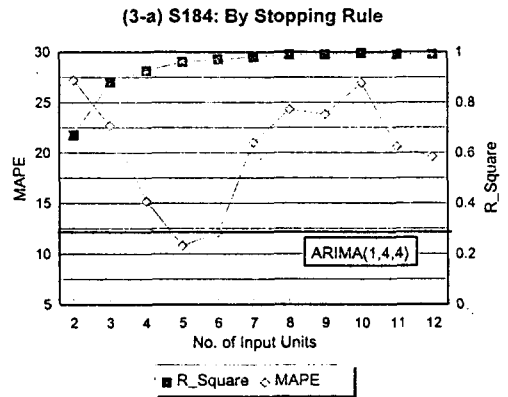


그림 3. 입력층 처리요소의 수가 예측성과에 미치는 영향 : 시계열 S184의 경우



나타낸다. ARIMA(1,4,4)라고 표시된 굵은 직선은 AUTOBOX 3.0에 의해 얻어진 MAPE 값이다. (그림 3-a)는 정해진 학습멈춤절차에 의해 MLP의 학습이 끝났을 때의 R2 및 MAPE 값을 입력벡터의 크기에 따라 변화시켜가며 나타낸 것이고, (그림 3-b)는 각기 주어진 크기의 입력벡터를 가지고 MLP를 훈련시켜는 과정에서 학습회수가 100단위가 될 때마다 유보된 관측치에 대한 MAPE를 구하고 그 중 가장 좋은 MAPE 값 (이 후 최저 MAPE (Best Performance)로 부름)과 그때의 R2 값을 함께 나타낸 것이다. 이 두 그림에서 알 수 있드시 입력벡터의 크기가 5일때 MLP는 가장 좋은 MAPE값을 가짐을 볼 수 있으며 이때 ARIMA(1,4,4) 모형보다 작은 MAPE값을 가진다.

하지만 입력벡터의 크기가 너무 작거나 커지면 MLP의 예측성능이 떨어짐을 알 수 있다. 즉 입력벡터의 크기가 커짐에 따라 R<sup>2</sup>의 값이 1에 가까워져 기존 관측치는 잘 설명하지만 MAPE 값은 작아지지 않고 오히려 커져서 예측성능이 나빠지는 것을 볼 수 있다. 또 입력벡터의 크기가 8, 9, 10, 11, 12의 경우 (그림 3-a, b) 두 그래프의 MAPE값을 비교하면 MLP의 학습회수를 어떻게 하는가에 따라서도 예측성능에 많은 차이가 있음을 볼 수 있으나 실제 예측작업에서는 (그림 3-b)의 최저 MAPE 값들을 찾을 수 없다는 점을 고려하여야 한다.

(그림 4)는 은닉층의 처리요소 수의 결정에 관한 것으로 입력벡터의 크기가 5와 12일 경우 은닉층 처

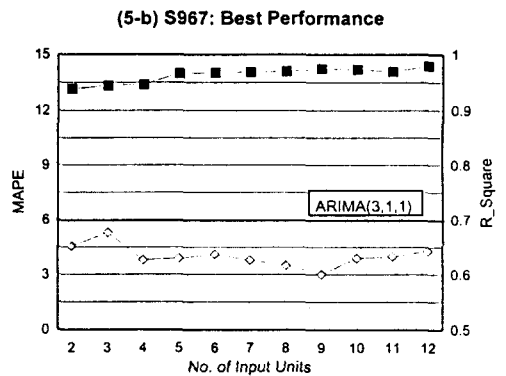
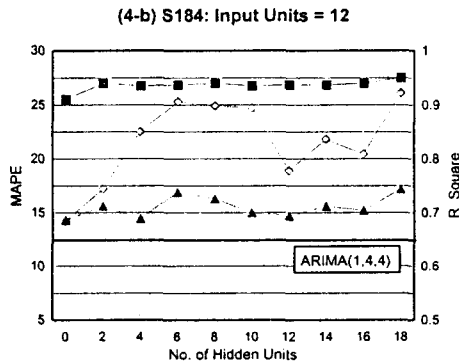
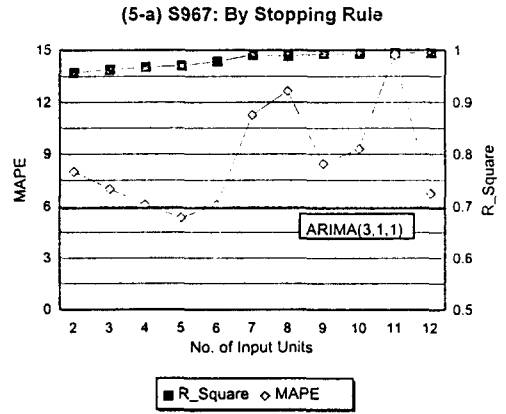
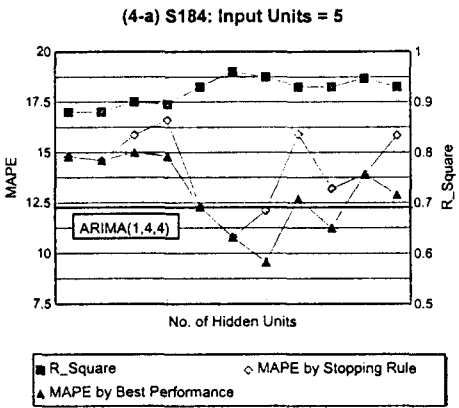


그림 4. 은닉층 처리요소의 수가 예측성능에 미치는 영향 : 시계열 S184의 경우

그림 5. 입력층 처리요소의 수가 예측성능에 미치는 영향 : 시계열 S967의 경우

리요소의 수를 변화시켜가면서  $R^2$ 와 두 가지 경우의 MAPE를 나타낸 것이다. (그림4-a)에서 보며 최저 MAPE는 6개의 은닉층 처리요소를 사용했을 때 얻어졌지만 채택된 학습멈춤절차에 의했을 경우에는 입력층의 처리요소 수와 같은 5개 일때 가장 적은 MAPE가 얻어졌다. 한편 입력벡터의 크기가 12일 경우에는 은닉층의 처리요소를 변화시켜도 ARIMA(1,4,4)보다 좋은 예측결과를 얻을 수 없었지만 처리요소의 수가 아주 적거나 12일때 비교적 좋은 결과를 얻을 수 있었다.

(그림 5)는 ARIMA(3,1,1)을 따르는 S967에 대한 분석결과인데 (그림 5-a)에서는 입력벡터의 크기가 4가 아니라 5일때 가장 작은 MAPE가 얻어졌으며 (그림 5-b)에서는 입력벡터의 크기가 9일때 최저 MAPE가 얻어져, 본 연구의 입력벡터 결정방법과는 일치된 결과가 얻어지지 않았지만 S967의 경우 입력벡터의 크기가 4일때와 5나 9일 경우의 예측성과들을 비교하면 큰 차이가 없었다. 예비실험에 사용된 나머지 시계열들에 대해서도 ARMA모형식별의 정보를 이용하여 MLP의 입력벡터 크기를 결정하고 같은 수의 은닉층 처리요소들을 사용함으로써 좋은 예측치들을 얻을 수 있었다.

## V. 실험결과

111개의 M1-Competition 자료 전체에 대한 MAPE의 평균 및 표준편차는 MLP가 12.387과 13.567이었고 AUTOBOX 3.0을 이용해 얻은 ARMA모형의 경우는 18.872와 36.017이었다. 두 예측방법간의 차이는 기존 관측치의 설명력 ( $R^2$ )에서 뿐만 아니라 예측성과 (MAPE)에서도 통계적으로 유의했다. 이 결과는 사회경제 시스템의 예측작업에서도 overfitting 문제를 적절히 해결하면 비선형 모형인 신경망이 유

용하게 이용될 수 있음을 입증하는 것이다. 111개의 M1-Competition 시계열 자료들 중 일부는 ARMA모형에 의한 모형화가 적절치 않다는 의견이 있어 왔는데 (Pack 와 Downing, 1983), 본 연구에서도 몇개의 시계열에 대해 AUTOBOX 3.0은 매우 큰 MAPE값을 보여주었다. 이와같은 주장을 반영하기 위해 Sharda 와 Patil (1992)의 연구에서와 같이 다시 72개의 시계열들을 가려내어 분석한 결과 MLP의 MAPE의 전체 평균은 10.585, 표준편차는 7.686이었으며, AUTOBOX 3.0의 경우는 15.699와 14.922이었다.

시계열의 특성이나 관측치의 수가 MLP의 예측성과에 영향을 주는가를 알아 보기 위한 분석을 계속 하였는데, (표 1)은 72개의 시계열들을 연간, 분기별 및 월간의 세 관측간격에 따라 구분하여, MLP와 AUTOBOX 3.0 둘 중에서 더 좋은 예측치를 보인 시계열들의 수를 나타낸 것이다. 전체적으로 53개의 시계열에 대해 MLP가 더 좋은 예측성과를 보였지만 상대적으로 평가해 볼때 MLP의 월간 시계열에 대한 예측성과는 연간이나 분기별 시계열의 경우보다 떨어지는 것을 볼 수 있다. M1-Competition 자료들 중 월간 시계열들이 분기별이나 연간 시계열들에 비해 더 많은 관측치들을 갖고 있음을 고려할때 MLP가 좋은 예측성능을 보이기 위해서는 관측치의 수가 많을수록 좋다는 사실과 위배되는 것이다.

MLP가 월간 시계열들에 대해 예측성과가 상대적으로 떨어지는 현상을 설명하기 위해 72개의 시계열들을 관측치들의 수에 따라 (1) 30개 미만 (2) 30개 이상 100개 미만 (3) 100개 이상의 세 그룹으로 나누

표 1. 예측성과 비교 : 관측간격에 따른 분류

(단위 : 시계열의 수)

관 측 간 격	연간	분	월간	합계
MLP	7	16	30	53
AUTOBOX 3.0	1	2	16	19
합 계	8	18	46	72

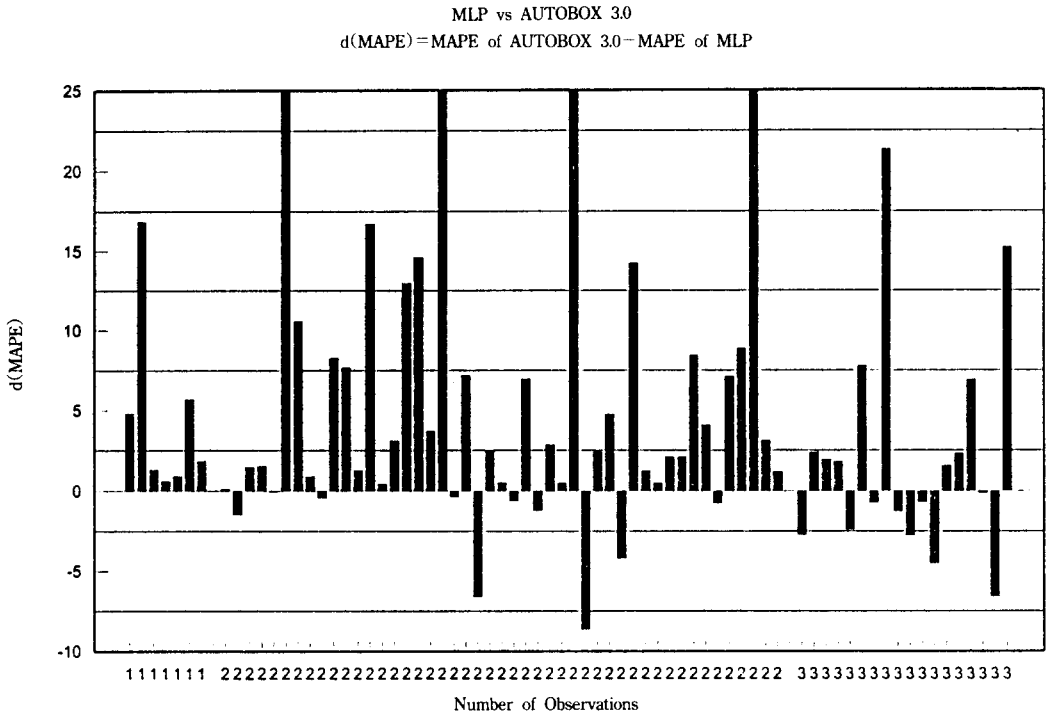


그림 6. 관측치 수(NOB)에 의해 예측성과의 비교 : 72개의 시계열  
 (1)  $NOB < 30$  (2)  $30 \leq NOB < 100$  (3)  $100 \leq NOB$

후 AUTOBOX 3.0과 예측성과를 비교하였다. (그림 6)에서 X축은 각 시계열들이 어느 그룹에 속하는가를 나타내고, Y축은 AUTOBOX 3.0의 MAPE를 MLP의 MAPE로 부터 차감한 값을 나타낸 것이다. 제 1, 2 그룹 즉 관측치가 100개 미만일 경우 MLP는 AUTOBOX 3.0에 비해 상당히 좋은 예측성과를 보여 본 연구에서 사용된 MLP구조 결정방법이 비교적 적은 양의 관측치들에 대해서도 잘 적용됨을 알 수 있다. 하지만 100개 이상의 관측치를 갖는 경우 두 방법간의 차이를 보이지 않았는데 이를 설명하기 위하여 M1-Competition자료들을 다시 시계열들이 갖는 특성에 의해 4가지 분류하였다. 즉 시계열이 추세나 계절성을 갖는가 여부에 따라 분류한 후 MLP와 AUTOBOX 3.0중 어느것이 더 좋은 예측결과를 보였는가를 정리한 것이 (표 2)이다. (표 2)로부터 시계열들이 추세나

표 2. 예측성과의 비교 : 시계열 특성에 의한 분류  
 (단위 : 시계열 수)

시계열특성	없음	추세	계절성	추세+계절성	합계
MLP	5	25	15	8	53
AUTOBOX 3.0	1	3	9	6	19
합 계	6	28	24	14	72

계절성을 갖지 않거나 추세만을 가지고 있을 경우에는 본 논문의 접근방법에 의해 설계된 MLP는 AUTOBOX 3.0보다 훨씬 좋은 결과를 보였지만 계절성을 갖거나 추세와 계절성을 동시에 갖는 경우에는 상대적인 예측성과가 떨어짐을 볼 수 있다. 따라서 관측치의 수가 많음에도 불구하고 MLP의 예측성과가 떨어지는 이유는 계절성에 원인이 있음을 알 수 있다. 실제로 M1-

Competition 자료들에서 100개 이상의 관측치를 갖는 시계열들은 대부분 월간 시계열들이고 이들은 계절성을 갖는 경우가 많았다.

이와 같이 계절성이 있는 경우 MLP의 예측성고가 떨어지는 이유에대한 가능한 설명은 ARIMA(p,d,q)로 식별된 시계열이 p와 d 즉 자기회귀 과정 및 differencing 차수에 계절성을 동시에 가질 경우 p+d를 입력벡터의 크기로 사용하므로 MLP는 상당히 큰 입력벡터를 가지게 되어 주어진 시계열에 대해 overfitting 되었던 것으로 볼 수 있다. ARMA모형화 과정에서는 파라메타 추정단계에서 비록 입력공간의 크기가 크더라도 설명력이 약한 변수들은 제외되는데 반해, MLP에서는 입력공간내의 모든 변수들이 무차별적으로

사용되고 파라메타의 수도 매우 많은 비선형 모형임을 감안하면 입력공간의 크기가 클 경우 시계열 내의 잡음들에 대해서도 과민하게 반응할 위험이 커지는 것이다. 따라서 계절성을 제거한 시계열을 사용하여 입력벡터의 크기를 줄일 경우 MLP의 예측성고가 개선될 수 있는가를 알아보기 위해 12개의 시계열들을 선택하여 분해기법(Decomposition Method)을 사용하여 계절성을 제거한 후 MLP의 예측성고를 정리한 것이 (표 3)이다. 3개의 시계열에 대해서는 계절성을 제거한 후 오히려 예측성고가 나빠졌으나 나머지 시계열들에 대해서는 상당히 좋아져 각 시계열별로는 최고 50% 가까운 예측성고의 개선을 얻을 수 있었다.

표 3. 계절성을 고려한 MLP의 예측성과 비교

시계열명	계절성 제거전			계절성 제거후		
	ARIMA치수	MLP구조	MAPE	ARIMA치수	MLP구조	MAPE
S409	0,12,13	14-14-1	18.567	1,0,2	3-3-1	13.396
S499	14,1,0	15-15-1	25.099	2,1,12	13-13-1	26.454
S508	2,13,12	15-15-1	7.732	12,1,2	13-13-1	8.425
S598	2,13,18	19-19-1	6.479	2,1,0	3-3-1	5.854
S643	1,12,12	13-13-1	11.997	1,0,12	13-13-1	6.141
S661	7,12,13	19-19-1	15.324	2,0,0	2-2-1	8.901
S787	14,13,0	27-27-1	2.815	2,0,12	13-13-1	3.794
S796	3,12,0	15-15-1	15.567	1,0,1	2-2-1	11.643
S877	13,13,0	26-26-1	3.499	1,0,1	2-2-1	2.690
S904	13,13,0	26-26-1	10.210	12,1,2	13-7-1	5.958
S958	2,13,12	15-15-1	6.783	2,1,0	3-3-1	3.919
S985	2,12,0	14-14-1	16.471	2,0,12	13-13-1	13.863
	평균 (표준편차)		11.712 (6.071)	평균 (표준편차)		9.253 (6.162)

## VI. 결 론

신경망이 사회경제 시스템의 새로운 예측수단으로 유효한가를 입증하기 위한 여러 연구들이 있었지만 단일변수 시계열 분석의 경우 연구결과들이 일치하지 않는 경우가 많았다. 그러나 각 연구에 사용된 신경망의 구조가 임의로 결정되어 시계열의 특성을 제대로 반영하지 못했던 사실과 분석에 사용된 시계열의 수가 극히 제한적인 것을 고려하면 아직도 기존의 연구결과들을 일반화하기는 어려운 실정이다.

본 연구에서는 다층 퍼셉트론의 구조를 결정하기 위하여 사전 자료분석의 중요성을 감안하여 ARMA 모형화의 모형식별단계에서 파악된 자기회귀 및 이동평균 과정의 차수들을 이용하여 입력벡터의 크기를 결정함으로써 다계층 퍼셉트론의 구조결정에 있어 시행착오를 줄일 수 있도록 했으며, 이를 111개의 MI-Competition 시계열 자료에 적용한 결과 좋은 예측성과를 얻을 수 있어서, 사회경제 시스템의 예측에서도 비선형 모형인 신경망을 유용하게 사용할 수 있음을 입증하였다.

본 연구의 다계층 퍼셉트론의 구조결정 방법은 시계열내의 관측치의 수가 적은 경우에도 과잉적합(overfitting)의 위험없이 사용가능했지만 계절성이 내재된 시계열의 경우에는 예측성과가 떨어지는 경향을 보였다. 이는 계절성으로 인해 상당히 큰 입력벡터를 사용하게 됨에 따라 다계층 퍼셉트론이 주어진 시계열에 과잉적합될 위험이 커지는 것으로 볼 수 있으며, 채택된 학습멈춤규칙으로 인하여 이 위험이 가중된 것으로 추정된다. 하지만 사용된 학습멈춤규칙은 현실적용가능성이 매우 크므로 이를 그대로 사용하고 대신 분해기법을 사용하여 계절성을 제거할 경우 다계층 퍼셉트론의 예측성과를 개선할 수 있음을 보였다. 따라서 본 연구 결과는 관측치의 수가 제한적이고 잡음이 심한 시계열에 대해서도 적용가능하며 앞으로

신경망을 이용하여 자동적으로 예측치를 생성하는 수단으로도 사용될 수 있을 것이다.

## 참고문헌

- 1) AFA Inc., *AUTOBOX 3.0 User's Manual*. Hatboro, PA (1991).
- 2) Akaike, H. "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, Vol. AC-19, (1974), 716~723.
- 3) Box, G.E.P., & Jenkins, G.M., *Time Series Analysis-Forecasting and Control*, Holden-Day, SF. (1976)
- 4) Caudill, M., *Neural Network Primer*. Miller Freeman Publications, San Francisco, CA (1989).
- 5) Chatfield, C., "Neural Networks : Forecasting Breakthrough or Passing Fad?" *Int'l J. of Forecasting*, Vol. 9, (1993), 1~3.
- 6) Chu, C.H., & Widjaja, D., "Neural Network System for Forecasting Method Selection," *Decision Support Systems*, Vol. 12, (1994), pp.13~24.
- 7) De Gooijer, J.G., & Kumar, K., "Some recent developments in non-linear time series modelling, testing, and forecasting," *Int'l J. of Forecasting*, Vol. 8, (1992), 135~156.
- 8) Fishwick, P.A., "Neural network models in simulation: a comparison with traditional modeling approaches," *Proceedings of the 1989 Winter Simulation Conference*, (1989), 702~710.
- 9) Foster, W.R., Collopy, F., & Ungar, L.H., "Neural network forecasting of short noisy time series." *Computers and Chemical Engineering*, Vol. 16, (1992), 293~297.
- 10) Harvey, A.C., "A Unified View of Statistical Forecasting Procedures". *Journal of Forecasting*, Vol. 3,

- (1984), 245~275.
- 11) Hornik, K., Stinchcombe, M. and White, H., "Multi-layer Feedforward Networks Are Universal Approximators," *Neural Networks*, Vol. 2, (1989), 359~366.
  - 12) Hoptruff, R.G., "The principles and practice of time series forecasting and business modelling using neural nets," *Neural Computing & Applications*, Vol. 1, (1993), 59~66.
  - 13) Jhee, W.C., & Lee, K.J., "Performance of neural networks in managerial forecasting," *Intelligent Systems in Accounting, Finance and Management*, Vol. 2, (1993), 55~71.
  - 14) Lapedes, A.S. & Farber, R.M., *Nonlinear signal processing using neural networks: prediction and system modeling*, Los Alamos Nat'l Lab. Technical Report, LA-UR-87-2662, (1987).
  - 15) Lee, K.J. & W.C. Jhee, "Two-Stage Neural Network Approach for ARMA Model Identification with ESACF," *Decision Support System*, Vol. 11, No. 5, (1994), 461~479.
  - 16) Makridakis et al, "The accuracy of extrapolation (time series) methods : results of a forecasting competition," *Journal of Forecasting*, Vol. 1, (1982), 111~153.
  - 17) Ripley, B.D., "Statistical aspects of neural networks", *Proceedings of SemStat*, Chapman & Hall, London, (1993).
  - 18) Rumelhart D.E., Hinton, G.E. & Williams, R.J., "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, Vol. 1 : Foundations (Eds. : Rumelhart D.E. and McClelland, J.L.), MIT, Cambridge MA, (1986).
  - 19) Sharda, R., & Patil, R.B., "A connectionist approach to time series prediction : an empirical test," *Journal of Intelligent Manufacturing*, Vol. 2, (1992).
  - 20) Tang, Z., Almeida, C., & Fishwick, P., "Time series forecasting using neural networks vs. Box-Jenkins methodology," *Simulation*, Vol. 57, (1991). 303~310.
  - 21) Tong, H., "*Non-linear time series : a dynamical system approach*," Oxford Univ. Press, (1990).
  - 22) Weigend, A.S., Rumelhart, D.E., & Huberman, B.A., "Back-propagation, weight-elimination and time series prediction. *Connectionist Models : Proc. of the 1990 Summer School*, (1990). 105~116.
  - 23) Werbos, P., "Generalization of backpropagation with application to recurrent gas market model," *Neural Networks*, Vol. 1, (1989). 339~356.
  - 24) White, H., "Economic prediction using neural networks: the case of IBM daily stock returns," *International Joint Conference on Neural Networks*, (1988), II 451~II458.