



### 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



**저작자표시.** 귀하는 원저작자를 표시하여야 합니다.



**비영리.** 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



**변경금지.** 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

**저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.**

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

物流學碩士學位論文

**컨테이너 물동량을 이용한 인공신경망과  
ARIMA 모형의 예측력 비교에 관한 연구**

**A Comparative Study on the Container Port  
Throughput Forecasting using Neural Network  
and ARIMA Models**

指導教授 辛 昌 勳

2008年 8月

韓國海洋大學校 大學院

東北亞物流시스템學科

李 芝 媛

# 목 차

제 1 장 서 론 .....	1
1.1 연구의 배경 및 목적 .....	1
1.2 연구의 방법 및 구성 .....	2
제 2 장 선행연구고찰 .....	3
제 3 장 분석기법 .....	5
3.1 인공신경망 .....	5
3.1.1 인공신경망의 개요 .....	5
3.1.2 인공신경망의 구조 .....	6
3.1.3 신경망 이론의 목적함수(Object Function) .....	9
3.1.4 신경망 이론의 은닉층과 은닉마디의 개수 .....	10
3.2 ARIMA .....	11
3.2.1 ARIMA모형의 개요 및 특성 .....	11
3.2.2 ARIMA모형의 구조 .....	12
제 4 장 실증 분석 .....	15
4.1 부두별 데이터 실증 분석 .....	15
4.2 인공신경망모형 분석 .....	19
4.2.1 원자료의 인공신경망 분석 .....	20

4.2.2 추세를 제거한 자료의 인공신경망 분석 .....	22
4.2.3 계절성을 제거한 자료의 인공신경망 분석 .....	24
4.2.4 추세, 계절성을 제거한 인공신경망 분석 .....	26
4.3 ARIMA모형 분석 .....	28
4.4 결과종합 .....	29
제 5 장 결론 및 연구의 한계 .....	31
참고문헌 .....	33

## 표 목 차

<표 4-1> 원자료의 인공신경망 최적화결과 .....	21
<표 4-2> 추세를 제거한 자료의 인공신경망 결과 .....	23
<표 4-3> 계절성을 제거한 자료의 인공신경망 결과 .....	25
<표 4-4> 추세와 계절성을 제거한 자료의 인공신경망 결과 .....	27
<표 4-5> ARIMA모형 결과 .....	28
<표 4-6> 모형종합 결과 .....	30

## 그림 목 차

<그림 3-1> MLP 신경망의 구조 .....	6
<그림 3-2> 인공신경망의 구조 .....	7
<그림 3-3> ARIMA모형의 분석 절차 .....	13
<그림 4-1> 부두별 수입 물동량 .....	16
<그림 4-2> 부두별 수출 물동량 .....	17
<그림 4-3> 부두별 환적 물동량 .....	18

*A Comparative Study on the Container  
Port Throughput Forecasting using  
Neural Network and ARIMA Models*

*Lee, Ji Won*

*Department of Logistics in Northeast Asia  
Graduate School of Korea Maritime University*

*Abstract*

The forecast of a container traffic has been very important for terminal plan and development. Generally, statistic methods, such as regression analysis, ARIMA, have been much used for traffic forecasting. Recent research activities in forecasting with artificial neural networks(ANNs) suggest that ANNs can be a promising alternative to the traditional linear methods. Time series data has trend and seasonality. In this paper, a ANNs methodology that make a consideration of trend and seasonal. The result with terminal traffic data indicate that effectiveness can differ according to the characteristics of terminals.

# 제 1 장 서 론

## 1.1 연구의 배경 및 목적

항만을 둘러싼 해운환경은 FTA의 확산과 중국의 WTO가입 이후 급속도로 변화하고 있다. 이러한 변화에 효과적으로 대처하기 위해 정부에서는 5년 단위의 무역항 기본계획을 수립하고 있다. 그러나 과거의 수요예측기법은 5년에서 10년 이상의 장기예측방법을 사용하고 있어 최근의 변화에 민첩하게 대처할 수 없는 실정이다.

항만 시설의 부족은 대형선사의 물량이 인접항만으로 빠져나가 손실이 발생하며 과도한 시설투자는 국가예산의 낭비를 초래하게 된다. 이를 위해 정부에서는 트리거물과 수요예측센터 등을 통해 이러한 문제를 해결하려고 노력하고 있으나 근본적인 문제해결을 위해서는 보다 정확성 높은 예측기법의 개발이 필요하다.

따라서 본 연구에서는 시계열(time series)자료를 이용한 대표적인 정량적 예측기법인 자기회귀이동평균(ARIMA; autoregressive integrated moving average)모형과 인공신경망 모형(ANN; artificial neural network)을 통해 보다 예측력 높은 모형을 개발하기 위한 방안을 모색하였다.

기존 연구에서 언급된 신경망모형의 단점을 해결하기 위해 추세를 제거한 모형(DT; de-trend Model)과 계절성을 제거한 모형(DS; de-seasonal Model), 추세와 계절성을 모두 제거한 모형(DTDS; de-trend de-seasonal Model)을 통해 이러한 신경망모형의 단점을 극복하는 방안을 살펴보았다. 또한 ARIMA 모형을 통하여 얻어진 예측 값을 비교하여 더욱 정확한 예측기법을 찾고자 하였다.



## 1.2 연구의 방법 및 구성

기존의 연구들에서는 시계열 자료의 특성인 추세와 계절성 부분에 대한 연구가 미흡하여 추세와 계절성을 각각 제거한 모형과 추세, 계절성을 모두 제거한 모형을 수립하여 좀 더 시계열 자료의 특성을 고려한 인공신경망모형을 사용하였다. 이를 통하여 얻어진 결과값과 ARIMA 모형을 통하여 얻어진 예측 값을 비교하였다.

부산항에 위치하고 있는 부두별로 2001년부터 2007년까지의 자료를 활용하였다. 부두는 자성대, 신선대, 우암, 감만, 감천한진, 일반부두 등으로 분류하였고, 데이터는 2001년부터 2006년까지의 월별자료를 분석하고 2007년 자료를 가지고 예측력 검증에 활용하였다.

본 연구의 구성은 다음과 같다.

2장에서는 각 기법을 활용한 모형들에 대한 선행연구고찰을 하였고, 3장에서는 각각 모형들의 분석 기법들을 설명했다. 4장에서는 각 분석 기법별 실증데이터 분석을 하였고, 마지막으로 5장에서는 본 연구의 결론을 넘으로 마무리 지었다.

## 제 2 장 선행연구고찰

전찬영, 송주미(2007)의 연구에서는 우리나라 국내 수출입 화물 10개 품목에 대한 시계열 예측을 인공신경망모형과 회귀분석을 적합성, 신뢰성, 강건성 측면에서 비교하였다. 이 과정에서 상승이 둔화되거나, 정체, 혹은 추세를 판단할 수 없는 시계열 자료에 대해서는 신경망 모형이 타당한 결과를 보이고 있고, 상승하고 있는 시계열 자료에 대해서는 회귀분석이 타당한 결과를 보이고 있다고 밝혔다.

이철영, 박성영(2002)의 신경망을 이용한 컨테이너 물동량 예측에서는 인공신경망과 일반 시계열 분석법인 이동평균법과 지수평활법, 추세분석법들을 비교하였다. 분석에 사용된 변수는 수출입 화물량, 환적화물량, 입출항 선박수 등으로 시계열 자료가 아닌 인과변수를 사용하였다. 이를 통해서 ANN모형이 단순한 지수평활법이나 이동평균법들보다 더욱 좋은 결과를 나타낸다고 밝혔다.

신창훈, 강정식 외(2007)의 연구에서는 ARIMA모형, ANN모형, 하이브리드모형을 통해 부산항 및 국내 주요 항만을 대상으로 컨테이너물동량 예측에 했다. 연구 결과 모형 적합도 부분에서는 인공신경망모형이 높게 나왔고 하이브리드 예측력부분은 ARIMA모형이 높게 나왔다. 이를 해결하기 위하여 ARIMA모형과 ANN모형의 결합방법을 다양하게 조절하는 하는 것이 더욱 좋은 예측력을 보일 것이라고 하였다.

Lam(2004)은 홍콩의 품목별 물동량에 대해 1983년에서 2000년까지의 18개의 자료를 가지고 회귀 분석과 인공신경망모형에 대해서 각각 분석하여, 적합도 및 신뢰성을 비교 분석하였다.

Lapedes & Farber(1987)는 시계열 자료를 인공신경망 모형이 기존의 계량적인 기법들보다 우수하다고 밝힘으로써 인공신경망 모형을 통한 다양한 분야에서의 예측연구가 진행되었다.

신성환(1995)은 선형모형, 인공신경망모형, 이동평균법의 세 가지 방법에 의한 환율예측력을 비교분석하였는데, 인공신경망 모형과 이동평균법 둘 다 수익률 예측 측면에서 효과적인 것으로 나타났으며, 특히 인공신경망모형이 주별 예측에 유용한 것으로 나타났다. 김정훈(2007)의 연구에서는 부산항, 광양항, 인천항을 대상으로 ARIMA 모형을 통한 선박통행량 예측이 이루어지는 등 물류관련 예측에서도 많이 사용되고 있음을 밝혔다.

Lam(2004)은 홍콩의 품목별 물동량을 회귀분석과 인공신경망모형으로 각각 예측하여, 적합도 및 신뢰성을 비교분석하였다. 1983~2000년 동안의 18개 시계열 자료를 이용하여 2002~2011년의 물동량을 예측하였다. 적합도의 검증을 위하여 2001년 실측치를 이용하여 회귀분석 결과와 인공신경망모형의 결과를 비교하였다.

전찬영(1999)은 중장기 항만물동량을 거시적 방법에 의거하여 항목별(수입, 수출, 연안)로 예측함으로써 경제상황 변화가 중장기 항만수요에 미칠 영향을 구체적으로 판단하였다. 각 항목별 물동량 특성을 고려하여 예측모형과 거시변수들 간의 조합을 통해 많은 모형을 설정 및 추정하였으며, 여러 통계적 특성치들이 유의한 것으로 모형을 선정하였다.

기존의 연구를 통해서 인공신경망은 상승이 하고 있는 자료에 대해서는 인공신경망의 예측력이 높지 않다고 밝혔다. 본 연구에서는 시계열의 특성을 고려하여 인공신경망분석을 하고 ARIMA모형 분석을 하여 그 결과를 비교해 볼 것이다.

## 제 3 장 분석기법

### 3.1 인공신경망

#### 3.1.1 인공신경망의 개요

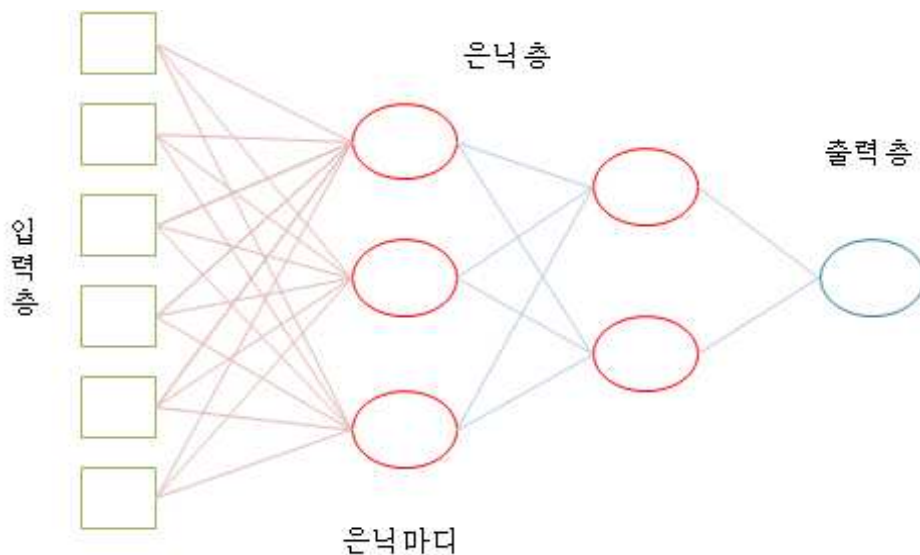
인공신경망(artificial neural networks)에 관한 연구는 뇌 신경생리학(neurophysiology)으로부터 영감을 얻어 시작되었다. 자료 분석 분야에서 신경망은 복잡한 구조를 가진 자료에서의 예측(prediction) 문제를 해결하기 위해서 사용되는 유연한 비선형모형(nonlinear models)의 하나로 분류될 수 있다. 그러나 신경생리학과 유사성 때문에 일반적으로 다른 통계적 예측모형에 비해 보다 흥미롭게 받아들여지고 있다.

신경망은 은닉마디(hidden units)라고 불리는 독특한 구성요소에 의해서 일반적인 통계모형과 구별되어진다. 은닉마디는 인간의 신경세포를 모형화한 것으로서, 각 은닉마디는 입력변수들의 결합(combination)을 수신하여 목표변수에 전달한다. 이때 결합에 사용되는 계수(coefficient)들을 연결강도(synaptic weights)라고 부르며, 활성화함수는 입력 값을 변환하고 이를 입력으로 사용하는 다른 마디로 출력하게 된다. 신경망 함수는 비선형 함수로 일반적으로 선형 함수의 예측 값보다 더 정교하다고 알려져 있다.

인공 신경망 모형은 항만물동량, 환율, 아파트가격, KOSPI 1200지수, BOD 등 다양한 분야에서 연구가 이루어지고 있다. 주로 회귀분석 및 시계열 분석 등의 통계적 모형과 인공신경망 모형의 우수성을 비교하고 있는데, 단기간 내에 심한 변동을 보이는 자료에 인공신경망 모형의 예측력이 두드러지는 결과를 보여주고 있다.

### 3.1.2 인공신경망의 구조

인공신경망에는 여러 가지 다양한 모형이 있으나, 그 중에서도 자료분석을 위해 가장 널리 사용되는 모형은 MLP(Multilayer Perceptron, 다층인식자) 신경망이다. MLP는 입력층(input layer), 은닉마디로 구성된 은닉층(hidden layer), 그리고 출력층(output layer)으로 구성된 전방향(feed-forward) 신경망이다. <그림3-1>은 입력층, 2개의 은닉층, 그리고 출력층으로 이루어진 MLP 신경망의 구조이다.

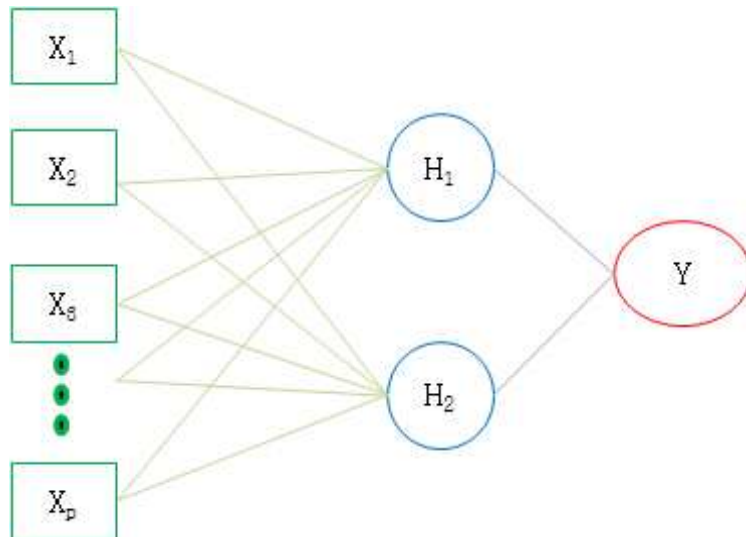


<그림 3-1> MLP 신경망의 구조

입력층: 각 입력변수에 대응되는 마디들로 구성되어 있다. 명목형 (nominal) 변수에 대해서는 각 수준에 대응하는 입력마디를 가지게 되는데, 이는 통계적 선형모형에서 가변수(dummy variable)를 사용하는 것과 같다.

은닉층: 여러 개의 은닉마디로 구성되어 있다. 각 은닉마디는 입력층으로부터 전달되는 변수 값들의 선형결합(linear combination)을 비선형함수(nonlinear function)로 처리하여 출력층 또는 다른 은닉층에 전달한다.

출력층: 목표변수(target)에 대응하는 마디들을 갖는다. 여러 개의 목표변수 또는 세 개 이상의 수준을 가지는 명목형 목표변수가 있을 경우에는 여러 개의 출력마디들이 존재한다.



<그림 3-2> 인공신경망의 구조

$$H_1 = f_1(b_1 + w_{11}X_1 + w_{21}X_2 + \dots + w_{p1}X_p) \quad (1)$$

$$H_2 = f_2(b_2 + w_{12}X_1 + w_{22}X_2 + \dots + w_{p2}X_p) \quad (2)$$

$$Y = g(b_0 + w_{10}H_1 + w_{20}H_2) \quad (3)$$

결합함수(combination function): 결합함수는 입력층 또는 은닉층의 마디들을 결합하는 형태를 의미한다. 예를 들어, 위 식에서 각 은닉마디  $H_1$ 과  $H_2$ 는 입력변수들을 선형결합 (즉,  $b_j + w_{1j}X_1 + w_{2j}X_2 + \dots + w_{pj}X_p$ )하여 이를 변환한다. 대부분의 신경망에서는 결합함수로 이와 같은 선형함수(linear function)를 사용하지만, 다른 형태의 결합함수를 사용하는 신경망들도 있다.

활성함수(activation function): 입력변수 또는 은닉마디의 결합을 변환하는 함수를 의미한다. 위의 식에서  $f_1$ ,  $f_2$ 와  $g$ 는 각각 활성함수와 출력활성함수(output activation function)라고 불리며, 보통 S-자형의 비선형곡선의 형태를 가지게 되어 스퀴싱함수(squashing function)라고 불리기도 한다. 활성함수와 출력활성함수는 동일한 함수를 사용하는 것이 일반적이다.

활성함수는 통계적 선형모형의 연결함수(link function)의 역함수와 유사한 의미를 가지며, 가장 보편적으로 사용되는 활성함수는 다음과 같은 로지스틱(logistic) 함수와 쌍곡탄젠트(hyperbolic tangent) 함수이다.

$$\text{로지스틱 함수 : } f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

$$\text{쌍곡탄젠트 함수 : } f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

한편, 목표변수가 제한된 범위를 가지지 않는 연속형 변수인 경우에는, 출력활성함수로 항등함수(identity function)를 사용하여 다음과 식(6)과 같이 출력마디가 생성되도록 하는 경우도 있다.

$$Y = b_0 + w_{01}H_1 + w_{02}H_2 \quad (6)$$

결합함수와 마찬가지로 목적에 따라 다양한 활성함수를 사용할 수 있다.

### 3.1.3 신경망 이론의 목적함수(Object Function)

적절한 수의 은닉 층과 은닉마디의 수가 결정되면, 그 다음에 해야 할 일은 데이터로부터 연결강도라고 불리는 계수(coefficient)들을 추정하는 것이다(이를 훈련(train)이라고 부른다). 즉, 신경망에 대해서 주어진 데이터를 가장 잘 반영하는 계수 값을 찾아내는 것인데, 이는 오차함수(error function)라고도 불리는 목적함수(object function)를 최소로 하는 문제가 된다. 일반적으로 주로 사용되는 목적함수는 일반적인 선형모형에서 흔히 오차 제곱합이라고 불리는 다음 식(7)과 같은 함수이다.

$$\sum_{i=1}^n (Y_i - P_i)^2 \quad (7)$$

여기서  $Y_i$ 는  $i$  번째 개체의 실제값이고,  $P_i$ 는  $i$  번째 개체의 예측값이다. 이 목적함수를 표현하면 다음 식(8)과 같다.

$$\sum_{i=1}^n (Y_i - P_i)^2 = \sum_{i=1}^m (Y_i - b_0 - w_{01}H_{1i} - w_{02}H_{2i} - w_{03}H_{3i})^2 \quad (8)$$

한편 통계적 모형에서는 목표변수의 분포를 가정하여 보다 다양한 형태의 목적함수를 고려할 수 있다.

신경망에서 목적함수를 최적화하는 계수 값을 찾는 것은 매우 어려운 작업이다. 신경망과 같은 비선형모형에는 이에 대한 분석적인 해가 존재하지 않기 때문에, 일반적으로 다음과 같은 수치해석적인 반복적 최적화방법을



사용한다.

- 1) 각 계수에 대해서 임의의 초기값을 부여한다.
- 2) “새로운 값 = 이전 값 + 수정값”과 같은 공식을 이용하여 계수 값을 개선한다.
- 3) 수렴할 때까지(즉, 이전 값과 새로운 값의 차이가 거의 없을 때까지) 반복적으로 수행한다.

추정의 비수렴성 문제는 신경망과 같은 비선형함수에 대한 최적화에서 매우 보편적으로 발생하는 문제이다. 만약 초기값이 국부 최소값(local minimum)에 가까우면 추정치는 전체 최소값(global minimum)이 아닌 국부최소값으로 수렴할 것이며, 국부 최소값에서의 모수 추정치는 실제 데이터를 정확하게 추정하지 못할 가능성이 많다. 이러한 문제를 해결하기 위한 한 방법은 여러 개의 초기 값으로부터 해를 구하고 이를 비교하여 사용하는 것이다. 그러나 어떤 알고리즘과 초기 값을 사용하던지 간에 전체 최소값을 보장하지는 않는다.

사실 활성화함수는 은닉마디나 출력마디의 범위를 일정한 범위 이내(흔히 0과 1 사이)로 제한하는 역할을 수행하기도 하는데, 신경망분석에 들어가기 이전에 입력변수와 출력변수에 대해 적절한 변환 또는 표준화(standardization)를 수행하는 것도 최적화 문제를 해결하기 위한 한 가지 방법이 될 수 있다.

### 3.1.4 신경망 이론의 은닉층과 은닉마디의 개수

신경망은 다양한 모형을 포함하는 매우 유연한 모형이다. 그러나 데이터로부터 계수를 추정해야 하기 때문에 실제로 MLP는 이론과 같이 유연하지 못하다. 또한 주어진 함수를 근사화하기 위해 매우 많은 수의 은닉마디가 필요할 수도 있다. 은닉층과 은닉마디가 많으면 많을수록 신경망은 더욱 복잡

해지며, 추정해야할 계수의 수가 급격히 증가하기 때문에 최적화가 훨씬 어렵게 된다. 사실, 적절한 은닉층과 은닉마디의 수를 결정하기 위해서 시행착오(trial - and - error)적인 방법 이외의 별 다른 대안이 없는 경우가 많다.

## 3.2 ARIMA

### 3.2.1 ARIMA모형의 개요 및 특성

일반적으로 ARIMA모형은 시계열 예측기법 중에서 가장 높은 예측 설명력을 보여 왔다는 학계와 일반 기업체의 많은 연구에서 나타나고 있으며 거시경제지표등의 예측이나 주식시세예측 등 그 적용사례가 매우 다양하게 나타나고 있다.

ARIMA모형은 예측의 정확성이 매우 높고 모형의 적합성을 검증할 수 있는 통계적 검진이 가능하며 예측의 신뢰구간설정이 가능한 장점이 있는 반면 상대적으로 많은 양의 데이터가 필요하며 새로운 데이터가 투입될 때 모형의 모수를 쉽게 업데이트할 수 있다는 방법에 만족스러운 모형을 개발하기 위해서는 많은 비용(시간과 자원)이 필요하다. ARIMA 접근법에 따르면 필요한 시계열 자료 수의 기준을 50개에 두고 있고, SAS/ETS 매뉴얼에서는 적어도 30개 이상의 시계열 자료가 필요하며 그 이하인 경우 추정 값에 문제가 있을 수도 있다고 한다. 본 연구에서는 2001년 1월부터 2006년 12월까지 부산항에서 처리된 컨테이너 물동량 72개의 관측 값을 사용하여 충분히 검증 가능한 데이터를 확보하였다.

ARIMA모형은 정상적, 비정상적 시계열을 모두 다룰 수 있어서 매우 융통성이 있으며, 모형의 적합성에 대한 통계적 검정이 가능하기 때문에 이론적으로 견고하고, 예측력 또한 상대적으로 우수하다고 평가된다. 그러나

복잡한 수학적 방법과 통계적 방법 그리고 주관적 판단을 필요로 하기 때문에 경험을 필요로 한다.

### 3.2.2 ARIMA 모형의 구조

ARIMA모형은 시계열자료가 추세나 계절성을 띠는 경우 이를 제거하여 정상적(stationary) 계열로 변환한 후, 자기회귀모형(autoregressive; AR), 이동평균모형(moving average; MA), 자기회귀이동평균모형 (autoregressive moving average; ARMA) 중에서 가장 적합한 모형을 선택할 수 있게 한다.

AR모형은 아래 식(9)와 같이 과거의 관측 값들을 기반으로 미래를 예측하는데 반해, MA모형은 식(10)에서 알 수 있듯이 과거의 오차 값들을 기반으로 미래를 예측하는 방법이다.

차수가 p인 자기회귀모형은 AR(p)의 형태로 아래 식(9)와 같이 표현된다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t \quad (9)$$

주:  $y_t$  : 종속변수,  $y_{t-1}, y_{t-2}, y_{t-p}$  : 독립변수,

$\Phi_1, \Phi_2, \Phi_p$  : 자기회귀계수,  $\epsilon_t$  : 잔차(백색잡음)

이동평균모형은 MA(q)의 형태로 아래 식(10)과 같이 표현된다.

$$y_t = \epsilon_t - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \dots - \Theta_q \epsilon_{t-q} \quad (10)$$

주:  $y_t$  : 종속변수,  $\epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-q}$  : 잔차의 이전값,

$\Theta_1, \Theta_2, \Theta_q$  : 이동평균계수

ARMA모형은 AR과 MA를 혼합시켜 놓은 모형으로 차수가 p와 q인 경우에는 ARMA(p, q)의 형태로 식(11)과 같이 표현된다. 즉 과거의 실제값의 오차 값을 동시에 고려하여 미래를 예측하는 것이다.

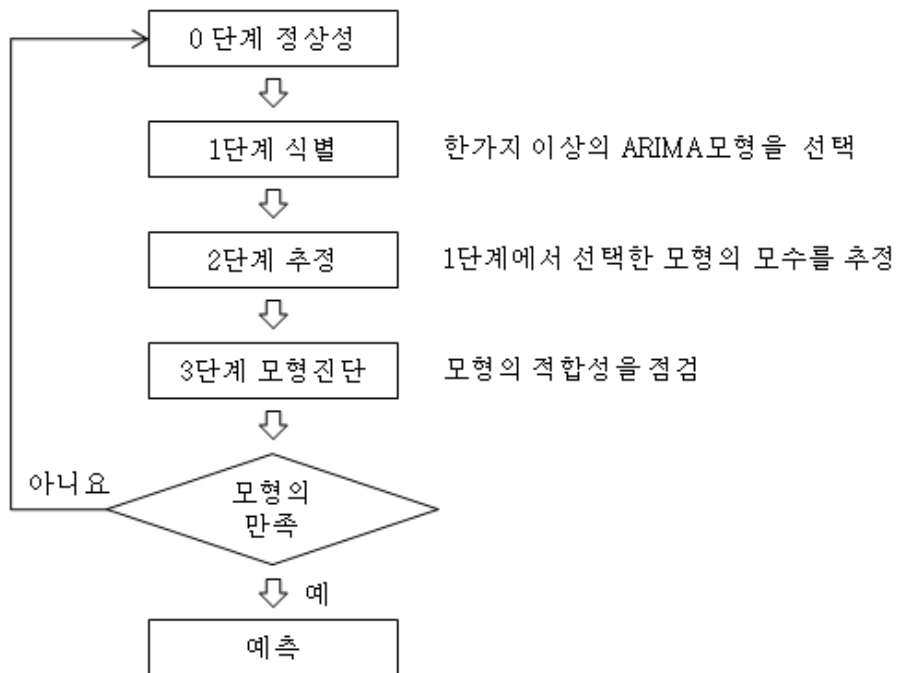
$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \dots - \Theta_q \epsilon_{t-q} \quad (11)$$

주:  $y_t$  : 종속변수,  $y_{t-1}, y_{t-2}, y_{t-p}$  : 독립변수,

$\Phi_1, \Phi_2, \Phi_p$  : 자기회귀계수,  $\epsilon_t$  : 잔차(백색잡음),

$\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$  : 잔차의 이전값,  $\Theta_1, \Theta_2, \Theta_q$  : 이동평균계수

ARIMA모형을 설정하기 위해 필요한 절차는 아래 <그림 3-3>과 같다.



<그림 3-3> ARIMA모형의 분석 절차

- 0단계 : 계열의 평균과 분산을 정상적(stationary)<sup>1)</sup>으로 변경시킴
- 1단계 : 일변량 자료계열의 상관관계를 자기상관함수(ACF, PACF)를 파악하고 ARIMA의 차수를 선택함
- 2단계 : 1단계에서 선택한 모형의 계수(들)를 정확히 추정하며, 모형적합성에 관한 몇가지 경고신호를 탐지하여 계수들이 주어진 수학적 부등식 조건을 만족하지 못하면 그 모형은 기각함
- 3단계 : 추정된 모형이 통계적으로 적절한지를 결정하며, 이 단계에서 나타난 결과들을 통해 모형이 얼마나 좋은지를 판단함(AIC<sup>2)</sup>, SBC 등을 통해)

이 3단계를 반복적으로 적용하여 가장 적절한 모형을 선택하여 예측시점들에 대한 예측을 실시하게 된다.

---

1) 정상성(stationarity)은 확률과정에서  $n$ 개의 시계열  $y_{t_1+k}, y_{t_2+k}, \dots, y_{t_n+k}$ 의 결합확률분포  $P(y_{t_1+h}, y_{t_2+h}, \dots, y_{t_n+h})$ 이  $t_1, t_2, \dots, t_n$  및  $h$ 와 관계없이 동일하다는 성질을 의미함

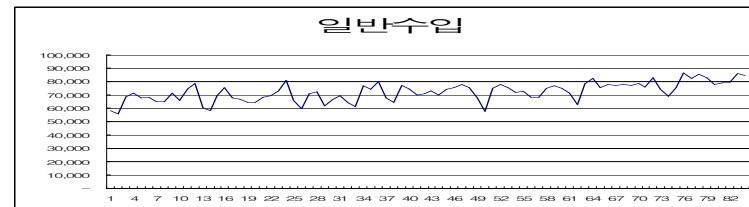
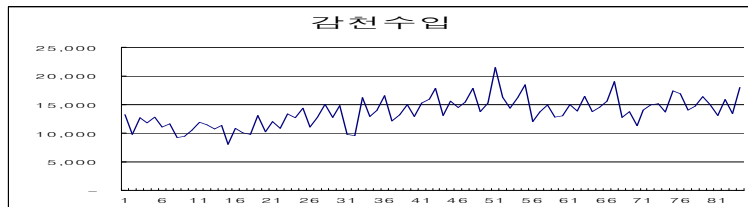
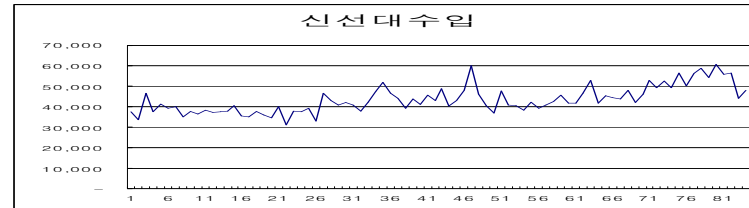
2) AIC(Akaike Information Criterion)는 모형의 최대 대수우도로부터 모수의 수를 줄인값을 근사적인 기대평균대수우도의 추정값으로 나타낸다. 최우법에 의해서 결정되는 모수를 가진 모형의 나쁜 정도를 비교하기 위한 지표이고 같은 데이터에 대해서 AIC가 최소가 되는 모형을 선택하는 것임

## 제 4 장 실증 분석

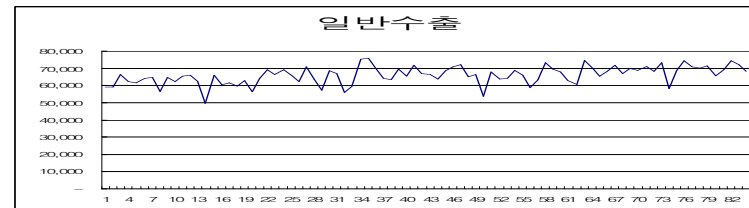
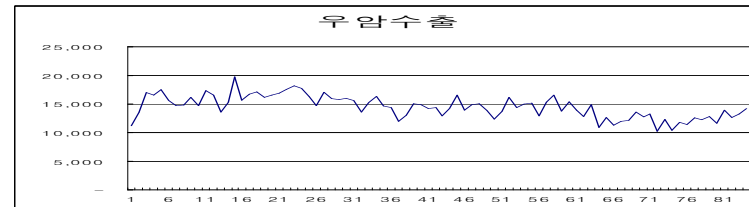
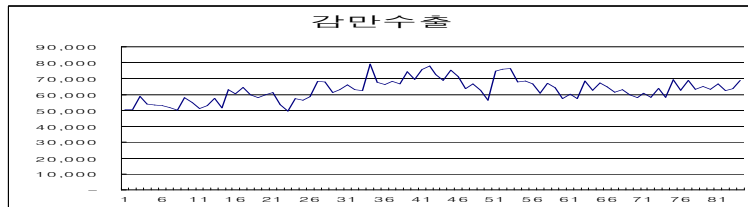
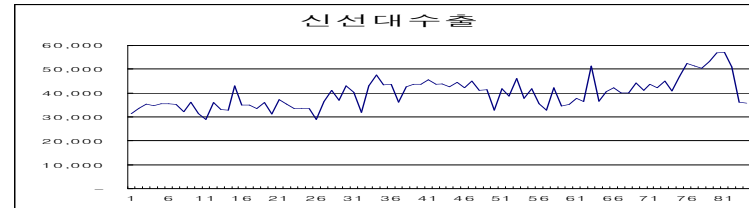
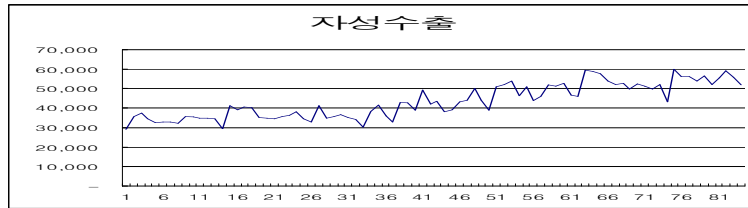
### 4.1 부두별 데이터 실증 분석

본 연구에 사용된 데이터는 부산항만공사에서 부산항만에 위치하고 있는 부두별 자료를 2001년부터 2007년까지 수입, 수출과 환적 물동량으로 수집하였다. 부산에 위치하고 있는 자성대, 신선대, 감만, 우암, 감천, 일반부두로 나누어 6개의 부두를 대상으로 분석하였다. 모형의 적합도를 알아보기 위해서 2001년부터 2006년까지의 월별 데이터 72개를 사용하였고, 2007년 12개의 자료를 예측하는데 사용하였다.

<그림 4-1>, <그림 4-2>, <그림 4-3>은 각각 부두별의 수입, 수출, 환적 물동량을 그래프로 나타내었다. 그래프를 살펴보면 수입 물동량에서 자성대부두는 꾸준한 증가를 하고 있고, 수출 물동량에서는 자성대, 신선대, 일반 부두가 증가하고 있으며 우암부두는 감소하고 있다. 환적 물동량에서는 신선대, 감만, 우암부두가 증가하고 있음을 알 수 있다.

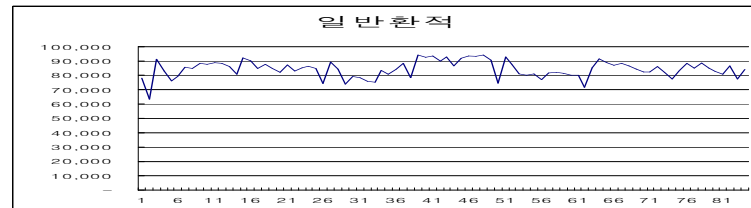
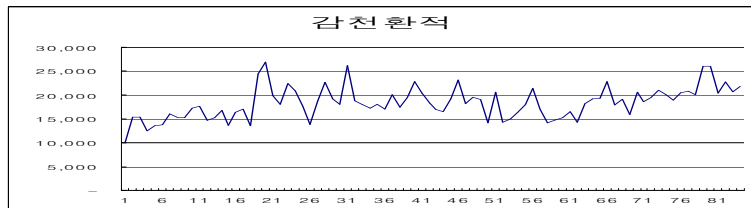
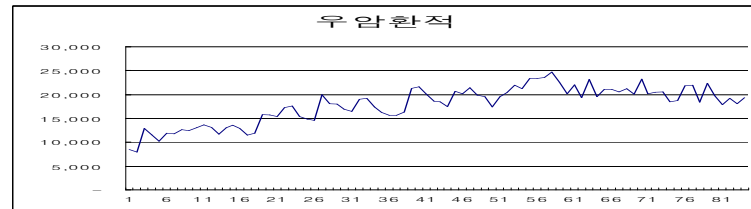
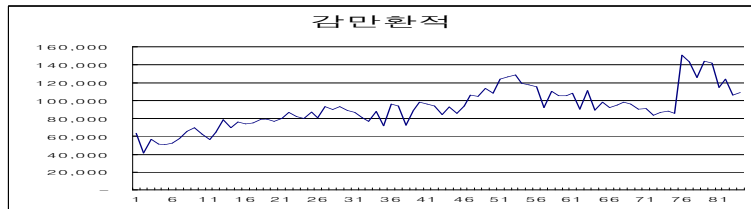
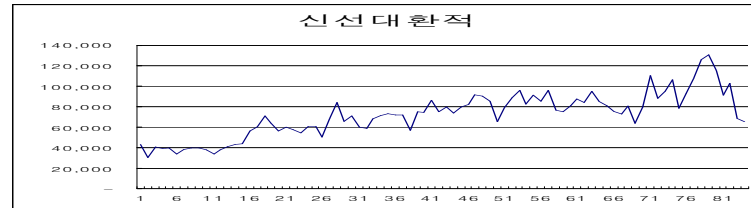
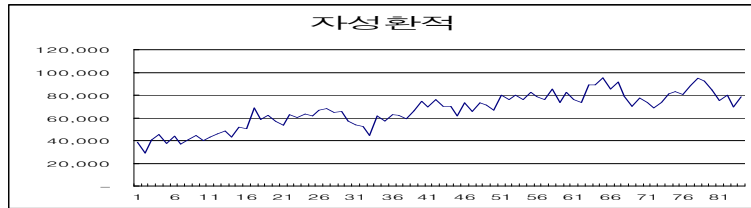


<그림 4-1> 부두별 수입 물동량



<그림 4-2> 부두별 수출 물동량





<그림 4-3> 부두별 환적 물동량

## 4.2 인공신경망모형 분석

시계열 자료의 분석에 나타나는 특성은 성질을 달리하는 몇 가지 변동의 합성으로 보고, 이들을 분해하여 각각을 측정하는 방법을 취한다. 시계열에 나타나는 구성요소는 김연형(2002)에 따라 추세변동, 순환변동, 계절변동과 불규칙부분으로 분류된다.

시계열 자료의 특성을 있는 그대로 고려한 인공신경망분석과 추세와 계절성을 각각 제거한 데이터에 대한 인공신경망분석과 추세와 계절성을 모두 제거한 데이터에 대해서 분석을 실행하였다.

정확한 신경망 모형의 구축을 위해서는 입력층의 유닛의 개수, 은닉층의 layer 개수와 유닛의 개수를 설정하여야 한다. 우선 입력층의 경우 Harvey(1984)는  $n = \max(p, q+1)$ 을 갖는 상태 공간 방정식으로 표현하였다. 이는 ARIMA의 AR모형의 차수와 MA모형의 차수+1의 개수 중에서 큰 수를 입력유닛의 수로 결정한다는 의미이다.

본 연구에서는 입력층의 유닛수는 2, 3, 4, 5, 6, 7개로 은닉층의 유닛수는 3, 4, 5, 6, 7, 8, 9로 변경하였다. 총 42개의 모형에 대해서 AIC(Akaike Information Criterion)값과 SBC(Schwarz's Bayesian Criterion)값을 비교하여, 가장 적절한 모형을 선정하여서 결과를 도출하였다.

#### 4.2.1 원자료의 인공신경망 분석

각 항만별 선택된 인공신경망 모형은 다음 <표4-1>와 같다. 데이터 실증분석에 사용된 데이터 분석그래프에서와 같이 자성대, 신선대, 우암, 감만, 감천의 환적물동량은 매년 증가하고 있는 추세이다. 증가하고 있는 물동량에 대한 AIC값과 SBC값의 최적화된 유닛의 수는 입력층 2개에 은닉층 3개인 모형이 각각 선택 되었다. 물동량이 감소하고 있는 감만의 수입과 우암의 수출 부분에 대해서는 입력층 5개와 은닉층 3개인 모형이 선택 되었다.

원자료의 인공신경망 분석은 증가하고 있는 물동량에 대해서는 입력층의 유닛은 적게 하고, 반대로 감소하는 물동량에 대해서는 입력층의 유닛을 많이 함으로써 최적화 모형을 찾을 수 있다.

<표 4-1> 원자료의 인공신경망 최적화 결과

부두	구분	ANN Model input units x hidden units	AIC값	SBC값
자성대	수입	lag-2x4 <sup>3)</sup>	-391.79	-350.46
	수출	lag-3x5	-341.84	-282.65
	환적	lag-2x3	-265.50	-235.91
신선대	수입	lag-6x4	-377.09	-301.96
	수출	lag-3x7	-352.23	-270.27
	환적	lag-2x3	-251.54	-221.94
감만	수입	lag-5x3	-335.02	-284.93
	수출	lag-2x3	-351.18	-321.58
	환적	lag-2x3	-276.31	-246.71
우암	수입	lag-2x3	-273.52	-243.92
	수출	lag-5x3	-339.54	-289.46
	환적	lag-2x3	-251.40	-221.81
감천	수입	lag-3x5	-313.02	-253.83
	수출	lag-2x3	-252.16	-222.56
	환적	lag-2x3	-256.12	-226.52
일반	수입	lag-2x3	-364.47	-334.87
	수출	lag-4x5	-429.40	-358.82
	환적	lag-2x3	-378.87	-349.27

주: ANN은 원데이터를 인공신경망으로 분석한 것을 의미함

3) lag-2의 의미는  $y_t$ 의 예측을 위해서  $\{y_{t-1}, y_{t-2}\}$ 을 입력층으로 사용했다는 의미임

#### 4.2.2 추세를 제거한 자료의 인공신경망 분석

추세성분(trend component)은 시간의 경과에 따라 증가하거나 감소하는 등 어떤 추세를 갖고 움직이는 장기변동을 의미한다. 이러한 추세를 제거하는 과정은 회귀분석을 하게 되면 추세를 파악할 수 있게 된다. 이 때 추세를 제거한 값인 잔차를 가지고 인공신경망 분석을 하였다.

추세를 제거한 자료의 인공신경망 분석을 나타낸 것은 <표 4-2>와 같다. 거의 대부분의 부두들이 AIC 통계량과 SBC통계량이 0에 가장 가깝게 나왔다.

증가 패턴을 보이고 있는 자성대, 신선대, 일반 부두의 수출 물동량은 원 데이터 인공신경망 분석과는 반대로 많은 입력층을 필요로 하고 있다. 그리고 일반부두의 수입, 수출, 환적 물동량이 상대적으로 입력변수와 은닉층의 유닛을 많이 필요로 하고 있다.

<표 4-2> 추세를 제거한 자료의 인공신경망 최적화 결과

부두	구분	DT Model input units x hidden units	AIC값	SBC값
자성대	수입	lag-3x3	-30.01	6.41
	수출	lag-6x4	-75.41	-0.28
	환적	lag-3x3	-21.52	14.91
신선대	수입	lag-4x8	-97.88	13.67
	수출	lag-7x4	-78.15	6.09
	환적	lag-6x3	-59.78	-2.86
감만	수입	lag-4x3	-39.00	4.25
	수출	lag-6x3	-46.28	10.63
	환적	lag-4x4	-55.14	1.78
우암	수입	lag-7x3	-32.87	30.87
	수출	lag-3x3	-18.01	18.42
	환적	lag-2x6	-47.19	9.73
감천	수입	lag-4x9	-152.89	-27.67
	수출	lag-4x8	-128.87	-17.31
	환적	lag-2x4	-4.18	34.52
일반	수입	lag-6x3	-42.74	14.18
	수출	lag-5x4	-7.68	42.41
	환적	lag-7x3	-44.53	19.22

주: DT는 추세를 제거한 후 인공신경망으로 분석한 것을 의미함.

### 4.2.3 계절성을 제거한 자료의 인공신경망 분석

계절성분(seasonal component)은 기온, 강수량, 월별효과 등과 같이 1년, 1개월, 1주 등 일정한 주기를 가지고 규칙적으로 반복되는 변동성분을 말한다. 이는 보통 1년 이내의 주기적인 변동을 말하며 삼각함수 등의 선형 결합에 의해서 설명될 수 있다. 계절성분을 구하기 위해서 본 논문에서는 중심이동평균 백분율을 사용해서 계절지수를 구했다. 1년 치인 12개의 월별 계절지수를 구한 후 실제 자료에 곱해준 다음 인공신경망 분석을 하였다.

계절성을 제거한 자료의 대부분은 AIC값과 SBC값의 최적화된 모형의 은닉층 2개와 입력층 3개를 사용하는 모델이 가장 적합하다고 나왔다.

<표 4-3> 계절성을 제거한 자료의 인공신경망 결과

부두	구분	DS Model input units x hidden units	AIC값	SBC값
자성대	수입	lag-2x3	-406.87	-375.27
	수출	lag-2x3	-399.43	-367.83
	환적	lag-2x3	-326.66	-295.06
신선대	수입	lag-2x3	-417.85	-386.25
	수출	lag-2x3	-386.82	-355.21
	환적	lag-2x3	-314.64	-283.04
감만	수입	lag-2x3	-363.91	-332.31
	수출	lag-2x3	-412.50	-380.90
	환적	lag-2x3	-319.79	-288.19
우암	수입	lag-2x3	-389.17	-357.57
	수출	lag-2x3	-450.48	-418.88
	환적	lag-2x3	-318.53	-286.93
감천	수입	lag-3x3	-342.97	-304.08
	수출	lag-4x3	-333.79	-287.60
	환적	lag-2x3	-308.21	-276.61
일반	수입	lag-3x3	-484.08	-445.18
	수출	lag-2x5	-488.95	-437.91
	환적	lag-3x3	-509.34	-470.44

주: DS는 계절성을 제거한 후 인공신경망으로 분석한 것을 의미함.



#### 4.2.4 추세, 계절성을 제거한 인공신경망 분석

추세, 계절성을 제거한 결과를 구하기 위해서는 원자료에서 추세를 제거한 잔차 값을 먼저 구한다. 잔차 값에 중심이동평균 백분율을 사용해서 구한 12개월 계절지수를 곱한 후 인공신경망 분석을 실행하였다. 그 결과는 다음 <표4-4>와 같다.

환적 물동량에서 증가패턴을 보이고 있는 신선대와 감만 부두는 입력층의 개수 4개에서 최적화 값이 나타났다. 물동량의 감소를 보이고 있는 우암 수출부분과, 일반 환적 부분에서는 입력층의 개수를 3개 필요로 하고 있었다. 이는 원자료의 인공신경망 분석과는 반대되는 결과로 증가 패턴을 보이고 있는 물동량의 입력층의 개수는 늘리고, 감소 패턴을 보이는 물동량의 입력층은 줄여야한다고 본다.

<표 4-4> 추세와 계절성을 제거한 자료의 인공신경망 결과

부두	구분	DTDS Model input units x hidden units	AIC값	SBC값
자성대	수입	lag-4x9	-163.2	-37.98
	수출	lag-5x5	-66.53	15.42
	환적	lag-4x3	-23.01	20.23
신선대	수입	lag-7x3	-57.53	6.21
	수출	lag-6x4	-71.24	3.88
	환적	lag-4x3	-45.59	-2.33
감만	수입	lag-3x3	-26.91	9.51
	수출	lag-3x4	-30.84	16.96
	환적	lag-4x4	-49.83	7.08
우암	수입	lag-3x3	-17.27	19.15
	수출	lag-3x3	-14.71	21.7
	환적	lag-2x3	-7.83	21.76
감천	수입	lag-4x9	-134	-8.79
	수출	lag-4x8	-129.6	-18.04
	환적	lag-3x3	9.81	46.24
일반	수입	lag-6x3	-41.03	15.87
	수출	lag-2x3	11.99	41.59
	환적	lag-3x4	-25.65	22.15

주: DTDS는 추세와 계절성을 모두 제거한 후 인공신경망으로 분석한 것을 의미함.

### 4.3 ARIMA모형 분석

ARIMA모형의 분석단계를 실시한 후 AIC(Akaike Information Criterion) 통계량 값과 SBC(Schwarz's Bayesian Criterion) 통계량 값을 구했다. 또한 인공신경망 값과 비교하기 위하여 예측치 와 실제 값의 차이(RMSE; root mean square error)값을 구했다. 아래의 <표 4-5>에서 ARIMA모형 분석 결과 값을 나타내었다.

<표 4-5> ARIMA모형 결과

부 두	구 분	ARIMA Model	AIC	SBC	RMSE
자 성 대	수 입	(0,1,1)(1,0,0) <sub>12</sub>	1,161.5	1,166.0	3,467.6
	수 출	(2,1,0)(0,1,1) <sub>12</sub>	982.1	988.4	3,913.6
	환 적	(2,1,0)(0,1,1) <sub>12</sub>	1,034.2	1,040.4	6,084.4
신 선 대	수 입	(2,1,2)(0,1,1) <sub>12</sub>	994.5	1,004.9	4,200.8
	수 출	(0,1,1) <sub>12</sub>	1,032.1	1,034.2	5,344.8
	환 적	(2,1,2)(0,1,1) <sub>12</sub>	1,094.0	1,104.4	9,761.9
우 압	수 입	(0,1,2)(0,1,1) <sub>12</sub>	850.3	856.6	1,281.0
	수 출	(0,2,2)(0,1,1) <sub>12</sub>	861.8	867.9	1,599.2
	환 적	(0,1,2)(0,1,1) <sub>12</sub>	903.3	909.5	2,006.1
감 만	수 입	(2,1,0)(0,1,1) <sub>12</sub>	1,013.9	1,020.2	5,124.9
	수 출	(0,1,1) <sub>12</sub>	1,077.5	1,079.6	7,803.4
	환 적	(0,1,2)(0,1,1) <sub>12</sub>	1,076.9	1,083.1	8,736.9
감 천	수 입	(0,1,2)(0,1,1) <sub>12</sub>	917.6	923.8	2,264.8
	수 출	(0,1,1) <sub>12</sub>	944.8	946.9	2,582.9
	환 적	(0,1,2)(0,1,1) <sub>12</sub>	967.9	974.1	3,469.3
일 반	수 입	(0,2,2)(0,1,1) <sub>12</sub>	967.3	973.4	3,970.9
	수 출	(0,1,1) <sub>12</sub>	1,011.3	1,013.3	4,493.7
	환 적	(0,1,1)(1,0,0) <sub>12</sub>	1,221.8	1,226.4	5,304.6

## 4.4 결과종합

앞서 분석한 인공신경망과 ARIMA모형을 비교하여서 <표 4-6>과 같은 결과를 얻었다. 모형의 적합능력을 오차(RMSE)를 통해 비교하였다.

18개의 자료에 대해서 ARIMA모형이 9개가 예측력이 좋았고, 인공신경망은 나머지 9개 부분에서 뛰어난 예측력을 보였다. ARIMA 모형이 뛰어난 예측력을 보여준 자료는 주로 상승 지속형이었고, 불규칙 상승형과 불규칙 정체형의 자료에 대해서는 인공신경망의 예측력이 뛰어난 것을 보이고 있다.

기존의 인공신경망을 이용한 연구에서 시계열 자료가 만약 지속적인 증가패턴이 나타나는 경우 타 모형에 비해 예측력이 떨어진다는 연구결과와는 일치하지 않는다. 자성대, 신선대, 감만, 우암 부두의 환적 물동량은 지속적으로 증가하는 추세인데 각각은 인공신경망에서 계절성과 추세를 고려하는 부분에서 더욱 더 예측력이 높게 나왔다. 그리고 감천부두 같은 경우에는 추세와 계절성을 모두 고려한 인공신경망의 예측력이 높게 나왔다. 이를 통해서 알 수 있는 것은 시계열자료는 각각의 특성을 고려했을 때 예측력이 높아진다는 것이다.

<표 4-6> 모형종합 결과

부두	구분	ANN	DT	DS	DTDS	ARIMA
자성대	수입	15,390	54,982	5,132	7,237	3,736
	수출	6,707	4,037	7,257	5,642	3,838
	환적	6,720	12,309	5,630	11,583	8,745
신선대	수입	19,384	5,486	11,733	7,098	8,227
	수출	10,725	7,873	9,072	8,265	10,345
	환적	25,212	21,835	22,627	22,176	24,987
감만	수입	7,611	10,283	3,907	10,266	3,244
	수출	4,513	6,672	8,328	7,093	3,870
	환적	25,679	27,337	22,847	27,955	35,271
우암	수입	1,040	1,693	1,158	1,715	865
	수출	1,351	2,418	907	2,629	2,954
	환적	2,397	5,114	2,190	5,552	3,849
감천	수입	2,213	6,242	2,318	6,093	1,686
	수출	1,575	1,476	1,483	1,557	1,867
	환적	3,835	2,074	3,969	1,946	3,012
일반	수입	9,056	6,457	9,362	4,647	3,328
	수출	5,346	6,553	4,819	5,065	4,615
	환적	5,602	8,862	5,962	8,720	4,739

주: ANN(인공신경망모형), DT(추세제거 ANN), DS(계절성제거 ANN), DTDS(추세+계절성제거 ANN)  
 표 안의 회색 셀은 최소오차를 나타냄.

## 제 5 장 결론 및 연구의 한계

예측기법은 분석에 사용되는 자료의 종류나 형태에 따라 그 예측력이 판이하게 달라질 수 있으며 어떤 모형이라도 항상 높은 예측력을 지닌 모형이란 평가를 기대하기 어렵다.

본 연구는 시계열 자료의 특성을 고려한 인공신경망과 ARIMA 모형의 예측력을 비교하였다. 자료의 특성을 고려한 분석기법의 예측력이 뛰어나서 기존에 해결하지 못했던 선형과 비선형적인 자료에 대해서 좋은 해법이 될 수 있을 것이라고 생각한다.

시계열 자료의 예측력을 높이기 위하여 추세를 제거하거나, 계절성을 제거하거나, 추세와 계절성 모두를 제거한 후에 인공신경망으로 분석하였다. 이러한 것들을 고려한 후의 예측력은 원자료의 인공신경망 분석보다 좋게 나왔고, ARIMA 모형과 비교했을 때도 좋게 나왔다. 기존 연구에서는 꾸준히 증가하고 있는 자료에 대해 인공신경망의 예측력이 좋지 않았지만 본 연구에서 분석한 특성을 고려한 모델에서는 뛰어난 예측력을 보였다.

또한 분석을 통하여 알 수 있었던 것은 원자료의 인공신경망 분석에서는 증가하고 있는 물동량에 대해서는 입력층의 개수를 줄이는 것이 분석이 좋게 나왔고, 추세와 계절성을 모두 제거한 인공신경망 분석에서는 반대로 늘리는 것이 분석이 좋게 나왔다. 감소추세의 물동량에 대해서도 반대의 연구결과가 나왔다. 이를 통해서 앞으로 인공신경망 분석을 하기 위하여 입력층과 은닉층의 유닛의 개수를 선정할 때 자료의 증가와 감소를 고려한다면 예측의 시행을 좀 더 줄일 수 있을 것이라고 본다.

인공신경망모형의 특성상 많은 자료를 필요로 하는데 현재는 제한된 관측개수를 사용하여서 한계점을 가진다. 추후의 연구에서는 이러한 한계를 극복할 수 있는 다양한 자료 수집에 힘써야 할 것이다. 본 연구에서는 한정된 입력층과 은닉층을 사용해서 자료의 예측기법을 선정하여 예측력을

알아보는데 초점을 맞추고 있다. 후에 예측을 하기 위해서는 더욱 더 다양한 입력층과 은닉층을 사용하여, 자료의 특성에 맞는 분석기법을 사용하게 되면 더욱 정확한 예측을 할 수 있을 것이라고 생각한다.

## 참고문헌

- 강현철 외 6명(2006), “고객관계관리(CRM)을 위한 데이터 마이닝 방법론,” 자유아카데미, pp.167-190.
- 김연형(2002), “시계열 예측,” 형설출판사, pp.15-306.
- 박성영, 이철영(2002), “신경망을 이용한 컨테이너 물동량 예측에 관한 연구,” 한국항해항만학회지 제 26권 제2호, pp.183-188.
- 부산항만공사, 인터넷 웹사이트(<http://www.busanpa.com>)
- 송일호, 정우수(2002), “SAS와 E-VIEWS를 이용한 계량경제실증분석,” 삼영사, pp.317-318.
- 신창훈, 강정식, 박수남, 이지훈(2007), “하이브리드 ARIMA-신경망 모델을 통한 컨테이너물동량 예측에 관한 연구,” 한국항해항만학회지 제 32권 제 1호, pp.81-88.
- 전찬영, 송주미(2007), “인공신경망모형의 항만물동량 예측 적용에 관한 연구,” 해운물류연구 제53호, pp.65-82.
- 조인호(2004), “SAS강좌와 통계컨설팅,” 영진닷컴, pp.742-754.
- 지원철(1995), “신경망을 이용한 시계열 분석 : M1- Competition Data에 대한 예측성과 분석,” 한국전문가시스템학회지 창간호, pp.135-148.
- 최봉호(2005), “수출입 컨테이너화물량의 예측,” 국제통상연구 제10권 제 2호, pp.121-140.
- 해운항만 물류 정보센터, 인터넷 웹사이트([http:// www.spidc.go.kr](http://www.spidc.go.kr))
- Box, G. E. P. and Jenkins, G. M.(1976), “Time series Analysis Forecasting and Control,” Holden-Day, San Francisco.
- G. Peter Zhang(2003), “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing 50, pp.159-175.
- Harvey, A. C.(1984), “A Unified View of Statical Forecasting Procedures,” Journal of Forecasting, Vol. 3, pp.245-275.



- Lapedes, A. & Farber, R.(1987), "Nonlinear Signal Processing Using Neural Networks Prediction and System Modeling," Los Almos National Laboratory Report, LA-UR-87-2662.
- Nelson, C. R. and C. I. Plosser(1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," Journal of Monetary Economics, 10 pp.139-162.
- S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibdon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler(1982), "The accuracy of extrapolation (time series) methods: results of a forecasting competition," Journal of Forecasting, Vol 9, pp.5-29.