하이브리드 ARIMA-신경망 모델을 통한 컨테이너물동량 예측에 관한 연구

신창훈* · 강정식** · 박수남** · †이지훈

* 한국해양대학교 물류시스템학과 교수, **, † 한국해양대학교 물류시스템공학과 대학원

A study on the forecast of port traffic using hybrid ARIMA-neural network model

Chang-Hoon Shin* · Jeong-Sick Kang** · Soo-Nam Park** · †Ji-Hoon Lee

* Department of Korea Maritime University, Busan 606-791, Korea
**, † Graduate school of Korea Maritime University, Busan 606-791, Korea

요 약 : 컨테이너항만의 물동량 예측은 항만의 개발 및 운영계획을 위해 매우 중요한 과정이다. 일반적으로 회귀분석, ARIMA모형 등의 통계적 방법론을 통해 많은 예측이 이뤄져왔다. 최근의 연구에서는 인공 신경망(ANN)기법을 통한 예측이 이뤄지고 있으며 기존의 선형적인 기법을 대신하고 있다. 본 연구에서는 선형모형과 비선형모형에 강점이 있는 ARIMA모형과 신경망모형을 결합해 보다 효과적인 예측 모형을 개발하고자 한다. 실제 항만의 과거 자료를 통해 모델의 적합성을 측정하였고 항만의 특성에 따라 모형의 적합성이 다양하게 나타났다.

핵심용어 : 컨테이너항만, 물동량예측, ARIMA모형, 인공신경망모형, 하이브리드모형

Abstract : The forecast of a container traffic has been very important for port 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. In this paper, a hybrid methodology that combines both ARIMA and ANN models is proposed to take advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling. The results with port traffic data indicate that effectiveness can differ according to the characteristics of ports.

Key words : Container port, Forecast, ARIMA model, ANN model, Hybrid model

1. 서 론

과거 부산항의 컨테이너부두 부족 문제는 1970년대 초 컨테이너제도가 도입된 이래 최근까지 계속되고 있는 문제로 2006년 신항의 개장으로 시설부족 문제는 차츰 해결되고 있는 추세다. 최근에는 오히려 물동량의 증가율이 둔화되고 있어 항만의 정책수립에 어려움을 주고 있다. 또한 최근 동북아지역의 항만 간 경쟁은 점차 치열해지고 있으며 이러한 경쟁 환경에서 막대한 투자가 요구되는 항만시설의 과부족을 명확히 판단하고 계획하는 일은 매우 중요한 과제이다. 해양수산부에서 추진 중인 트리거룰(trigger rule)은 이러한 개발시기의 조정을 위해 검토되고 있다. 이처럼 과거와는 달리 급변하는 환경에서 항만의 물동량의 예측은 보다 단기적이고 높은 정확성을 요구하고 있다.

ARIMA모형은 많은 시계열 분석 방법 중 사회경제적 시스

템의 단기예측에 있어 이론적 우수성과 예측의 정확성을 인정받아왔다. 최봉호(2005)의 연구에서는 오차수정모형과 ARIMA모형을 통해 부산항을 비롯한 국내 주요 항만의 물동량을 예측했으며 분석결과 ARIMA모형을 이용한 결과가 보다 높은 적합도와 정확성을 보이는 것으로 나타났다.

최근에는 데이터마이닝(data mining) 기법에 대한 연구가 진행됨에 따라 비선형예측기법에 대한 관심이 증가하고 있다. Lapedes & Farber(1987)의 연구에서 시계열분석에 인공신경망(Artificial Neural Network ; ANN)이 이용될 수 있음을 입증한 후 이를 활용한 예측이 많이 진행되고 있다. 전찬영과 송주미(2007)의 연구에서는 주요 수출입품목을 대상으로 회귀분석(regression analysis)과 인공신경망모형의 예측력을 비교하였다. Zhang(2003)의 연구에서는 기존의 선형예측기법과 비선형예측기법을 접목시킨 하이브리드모형을 소개하고 기법간의 모형의 적합도와 예측력을 비교하였다.

* 대표저자 : 신창훈(정회원), chshin@hhu.ac.kr 051) 410-4333

** 강정식(정회원), nodi21@dreamwiz.com 051) 465-4301

** 박수남(정회원), onlypsn@lycos.co.kr 031) 955-6760

† 교신저자 : 이지훈(정회원), k9410032@korea.com 051) 410-4930

본 논문에서는 여러 시계열 예측방법들 중 단기예측에 가장 많이 사용하는 모형중 하나인 ARIMA모형과 인공 신경망 모형을 통해 모형의 추정 및 물동량을 예측하고 ARIMA모형과 ANN모형을 결합한 하이브리드 모형을 통해 새로운 항만 물동량 예측방법을 검토하였다. 모형간의 성능 비교를 위해 부산항, 광양항, 인천항, 울산항, 마산항의 1997년 이후 월별 물동량 자료를 수집하여 예측모형별로 추정하였고 이를 통해 모형의 적합도와 예측력을 산출하였다.

2. 선행연구 고찰

수요예측방법은 크게 정성적인(qualitative) 방법과 정량적인(quantitative) 방법으로 구분할 수 있다. 정성적인 방법은 일반적으로 전문가의 의견을 사용해서 미래의 결과를 주관적으로 예측하는 방법으로 델파이(delphi) 방법과 시나리오방법이 대표적이다. 정량적인 방법은 크게 시계열분석(time series analysis)과 인과분석(causal analysis) 방법이 있다. 인과분석은 투입변수와 산출변수의 인과관계를 통해서 나타나며, 시계열모형에서 산출변수는 투입변수의 자체생성과정을 통해서 나타나게 된다.

Box & Jenkins(1976)에 의해 개발된 ARIMA모형은 시계열 예측기법 중에서 매우 높은 예측 설명력을 보여 왔고 학계와 일반 기업체의 많은 연구에서 사용되고 있다. 이를 통한 거시경제지표 등의 예측이나 주식시세예측 등 그 적용사례가 매우 다양하다. 최봉호(2005)는 국내 주요 항만의 물동량을 계절 ARIMA 모형을 통해 예측한바 있고 오차수정모형과의 예측력을 비교하였다. 분석결과 ARIMA모형의 예측력이 오차수정모형에 비해 높은 것으로 나타났다.

Lapedes & Farber(1987)는 신호처리분야의 학자로 시계열 자료를 인공신경망모형으로 예측한 결과 기존의 계량적 기법들보다 우수함을 보인다고 밝혔다. 이후 ANN모형을 통해 다양한 분야에서 예측연구가 진행되고 있다. 박성영과 이철영(2002)의 연구에서는 ANN모형을 통해 우리나라 항만의 컨테이너물동량을 예측하였고 단일변수가 아닌 수출입화물량, 환적화물량, 입출항선박수 등의 시계열자료를 입력 값으로 분석하였다. 분석결과 지수평활법 등 단순시계열예측기법에 비해 높은 예측력을 보이는 것으로 나타났다. Lam et al.(2004)은 1983년에서 2000년까지의 홍콩항의 18개 품목별 물동량 자료를 토대로 회귀분석과 인공신경망 모형으로 각각 예측해서 모형의 적합도 및 신뢰성을 비교 분석하였다. 전찬영과 송주미(2007)의 연구에서도 국내 주요 수출입품목 10개를 선정하여 ANN모형과 회귀분석모형의 적합성과 예측력을 비교 분석한 바 있다. 비교결과 품목별 시계열의 특성에 따라 ‘상승지속형’의 경우 회귀분석이 ‘불규칙 상승’, ‘상승둔화’, ‘정체’의 경우 ANN모형이 높은 적합도를 나타내는 것으로 분석되었다.

하이브리드 모형은 Makridakis et al.(1982)의 연구에서 1,001개의 시계열 데이터베이스를 통해 다양한 연구자들이 다

양한 예측기법을 통해 그 예측력을 비교한 것에서 시작되었다. 자료에 대한 다양한 예측기법들이 사용되었으며 이후 많은 학자들은 기존의 계량예측기법과 인공신경망예측을 비교 또는 접목시키려고 노력하고 있다. 여기서 하이브리드 모형은 단일한 예측모형을 사용하는 것이 아닌 두 가지 이상의 예측모형을 접목시켰다는 의미에서 사용되었다.

Zhang(2003)의 연구에서는 ARIMA모형의 선형적 자료에 대한 장점과 비선형적인 자료의 경우 모형의 구축이 어렵다는 단점을 ANN모형을 통해 보완하였다. 연구를 위해 Makridakis et al.(1982)의 대표적인 시계열자료인 Wolf's 흑점자료, 환율자료를 통해 비교분석하였다. 연구결과로 거의 모든 자료에서 하이브리드모형이 우수한 것으로 나타났다. 이는 대부분의 시계열 자료는 선형과 비선형적인 특성을 지니고 있으며 두 가지 특성을 반영한 모형을 구축할 수 있다면 기존의 단일모형의 예측력을 상회하는 예측력을 얻을 수 있다는 이론을 바탕으로 많은 접근방법이 연구되고 있다.

3. 연구 모형

3.1 ARIMA모형

ARIMA모형은 자기회귀모형(autoregressive; AR), 이동평균모형(moving average; MA), 자기회귀이동평균모형(auto-regressive moving average; ARMA) 중에서 가장 적합한 모형을 선택하는 방법이다. 우선 AR모형은 아래 식 (1)과 같이 과거의 관측값들을 기반으로 미래를 예측하는데 반해, MA모형은 식 (2)에서 알 수 있듯이 과거의 오차값들을 기반으로 미래를 예측하는 방법이다.

차수가 p인 자기회귀모형은 AR(p)의 형태로 아래 식(1)과 같이 표현된다.

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

현 시점 t에서 시계열 y_t 는 p개의 과거 값들의 가중합들과 이들로 설명되지 않는 잔차 ϵ_t (백색잡음)으로 구성된다.

또한 이동평균모형은 MA(q)의 형태로 아래 식(2)과 같이 표현된다.

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

MA모형은 AR모형과 달리 t시점의 시계열 y_t 는 현재와 과거 q개의 백색잡음에 대한 가중선형결합으로 표현된다.

ARMA모형은 AR과 MA를 혼합시켜 놓은 모형으로 차수가 p와 q인 경우에는 ARMA(p, q)의 형태로 식 (3)과 같이 표현된다. 즉 과거의 실제값의 오차값을 동시에 고려하여 미래를 예측한다.

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \dots - \Theta_q \epsilon_{t-q} \quad (3)$$

ARIMA모형을 이용한 예측을 하기 위한 절차는 아래 Fig. 1과 같다.

Fig. 1 Flow chart of ARIMA process

ARIMA모형의 분석과정은 우선 준비단계로 시계열 자료의 평균과 분산을 통해 정상성(stationary)¹⁾ 여부를 판단하고 불 안정(non-stationary) 시계열일 경우 안정적으로 변경한다. 이때에는 추세를 제거하기 위한 차분(difference) 또는 로그변환이나 제곱근변환을 통한 분산의 안정화가 이뤄진다. 다음으로 모형의 식별과정으로 일변량 자료계열의 상관관계를 자기상관함수(ACF), 부분자기상관함수(PACF)로 파악하고 ARIMA 모형의 차수를 선택한다.

두 번째 단계로 모형의 식별단계에서 선택한 모형의 계수들을 정확히 추정하며, 모형적합성에 관한 몇 가지 경고신호를 탐지하여 계수들이 주어진 수학적인 부등식 조건을 만족하지 못하면 그 모형은 기각한다. 모형의 진단과정에서는 추정된 모형이 통계적으로 적절한지를 결정하며, 이 단계에서 나타난 결과들을 통해 모형이 얼마나 좋은지를 판단한다. 이때 AIC(Akaike Information Criterion), SBC(Schwarz's Bayesian Criterion) 통계량 등을 이용한다.

끝으로 이 3단계를 반복적으로 적용하여 가장 적절한 모형을 선택하여 예측시점들에 대한 예측을 실시한다.

3.2 인공신경망(ANN) 모형

인공신경망모형은 생물학적 신경계를 단순화시켜 수학적으로 모사한 모델로서 신경세포와 연결들로 구성된다. 과거에 수집된 자료를 통해 학습과정을 거치고 패턴을 찾아내는 분석 방법이다.

인공신경망은 일반적으로 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)의 다층퍼셉트론(multi layer perceptron)으로 구성되어 있다.

현시점 t 기의 시계열은 $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ 의 과거 값들을 입력층으로 q 개의 은닉층을 거쳐 α_j, β_{ij} 의 가중치를 통해 구할 수 있다.

Fig. 2 A model for ANN process

이를 수식화 하면 다음 식 (4)와 같다.

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i}) + \epsilon_t, \quad (4)$$

여기서 $g(x)$ 는 입력층과 은닉층 사이에 비선형적 관계를 위한 활성화함수로 주로 logistic함수나 쌍곡탄젠트(hyperbolic tangent)함수를 사용하며 본 연구에서는 활성화함수로 쌍곡탄젠트함수를 식 (5)와 같이 사용하였다.

$$g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

3.3 하이브리드모형

본 연구에서 사용한 하이브리드모형은 앞서 서술한 ARIMA 모형과 ANN모형을 결합한 형태로 ARIMA모형의 경우 선형 모형의 예측에 매우 강력한 것으로 평가되며 ANN모형의 경우 비선형문제 해결에 탁월하여 이 두 모형을 결합하여 두 모형과 비교분석하였다.

모형의 결합방법은 Zhang(2003)의 연구에서와 같이 우선 ARIMA모형을 수립하고 여기서 발생하는 오차값을 신경망모형을 통해 분석하였다. 이를 통해 ARIMA모형을 통해 시계열의 추세와 계절성이 우선 모형에 반영되고 이후 오차항의 비선형 패턴이 있을 경우 ARIMA모형과 ANN모형보다 높은 설명력을 보일 것으로 전망한다.

하이브리드모형을 수식으로 표현하면 다음 식 (6)과 같다.

$$y_t = L_t + N_t \quad (6)$$

1) 정상성은 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 의미한다. 즉 시계열의 평균과 분산이 시간에 따라 증가 혹은 감소하지 않고 일정한 경우를 의미하는 것이다.

t기의 시계열 값은 선형함수 L_t 와 비선형함수 N_t 의 합으로 구성되어 있고, t기의 오차 e_t 는 ARIMA모형의 추정값 $\hat{L}_t$ 의 차이로 구할 수 있고,

$$e_t = y_t - \hat{L}_t, \quad (7)$$

오차는 ANN모형을 통해 과거 n기까지의 오차값을 입력층으로 하는 비선형모델로 식 (8)과 같이 나타낼 수 있다.

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \epsilon_t, \quad (8)$$

따라서 하이브리드모형을 통한 t기의 추정치 $\hat{y}_t$ 는 ARIMA 모형의 추정치와 ANN모형의 추정치의 합으로 다음 식 (9)과 같이 나타낼 수 있다.

$$\hat{y}_t = \hat{L}_t + \hat{N}_t, \quad (9)$$

4. 실증 분석

4.1 자료의 구성

예측모형의 비교를 위해 해양수산부에서 운영 중인 해운항만 물류 정보센터에서 국내 주요 항만의 수출, 수입 컨테이너물동량 자료를 수집하였다. 부산항을 비롯해 광양항, 인천항, 울산항, 마산항을 대상으로 1997년 이후부터 2006년까지 물동량을 기준으로 분석하였고 각각의 모형 적합을 위해 1997년부터 2004년까지 96개의 자료를 사용했다. 모형의 예측력을 알아보기 위해서 2005년~2006년 구간을 예측구간으로 사용하였다.

Fig. 3의 수입화물의 시계열도표를 살펴보면 부산항, 광양항, 인천항은 지속적인 상승추세가 보이고 울산항은 2003년 이후 불규칙한 패턴을 보이고 있고 마산항의 경우 매우 불규칙한 증감패턴을 보이는 것으로 나타났다. 또한 대부분의 항만에서 계절적인 패턴도 보이는 것으로 분석되었다.

Fig. 3 Time plot of import data

Fig. 4 Time plot of export data

수출화물의 경우에도 수입화물과 마찬가지로 울산항과 마산항을 제외하고 모두 지속적인 증가추세를 보이는 것으로 나타났다.

분석에 사용된 자료 중 광양항의 수입과 수출의 경우 1999년 이전의 물동량 자료는 이후의 물동량 자료와는 큰 차이를 보여 분석에서는 제외하였다. 연구대상 항만의 물동량의 기초 통계량은 Table 1과 같다.

Table 1 Characteristics of raw data(1997~2006)

port	variable	mean	std dev.	skew.	kurtosis
부산-수입	BSI	230,592	42,900	0.043	-1.103
부산-수출	BSE	235,123	37,392	0.147	-1.038
광양-수입	GYI	28,747	16,615	-0.326	-0.737
광양-수출	GYE	29,813	17,810	-0.340	-0.875
인천-수입	ICI	36,178	11,988	0.897	0.078
인천-수출	ICE	29,598	12,227	1.132	0.553
울산-수입	USI	9,186	3,064	-0.473	-0.830
울산-수출	USE	10,853	4,303	-0.372	-1.009
마산-수입	MSI	1,688	560	0.474	-0.369
마산-수출	MSE	2,088	753	0.421	-0.597

4.2 ARIMA모형 분석

Nelson & Plosser(1982)는 대부분의 주요한 시계열 거시경제변수는 불안정적이라는 연구결과를 발표하였다. 이 연구에 따르면 불안정한 상태의 시계열 변수들은 아무런 상관관계가 없다고 할지라도 회귀계수의 T-값이 표본수가 커짐에 따라 증가하여 분석결과를 오도하는 허구적회귀(spurious regression)의 문제가 발생하게 된다. 따라서 이러한 시계열자료의 분석에 앞서 단위근 검정을 통해 각 항만의 물동량 자료가 안정적인지를 검증해 보도록 한다.

단위근 검정의 단순 DF(Dickey-Fuller)검정법은 오차항이 백색잡음인 경우에 한하여 유효하기 때문에, 계열상관 뿐만 아니라 이분산을 조정하기 위하여 DF검정법의 t통계량을 조정한 Phillips-Perron 검정법을 이용할 수도 있다. Phillips-Perron 검정법은 불안정하다는 귀무가설을 기각하는 경향을 강하게 갖고 있는 문제가 있다. 따라서 본 연구에서는 ADF(Augmented Dickey-Fuller) 검정을 통해 시계열변수의 안정성을 검정하였다. 단위근 검정결과 마산항의 수입물량을 제외한 모든 변수에서 단위근이 있다는 귀무가설을 기각하지 못해 모든 변수는 비정상적인 시계열로 판단되었다.

Table 2 Result of Unit root test

항만	구분	Y_t		ΔY_t	
		ADF t-통계량	Probability	ADF t-통계량	Probability
부산항	수입	-0.835	0.804	-3.680	0.005
	수출	-0.767	0.823	-3.488	0.010
광양항	수입	-0.552	0.875	-12.115	0.000
	수출	-0.543	0.877	-11.334	0.000
인천항	수입	3.777	1.000	-10.525	0.000
	수출	4.708	1.000	-6.976	0.000
울산항	수입	-2.113	0.239	-12.551	0.000
	수출	-1.539	0.510	-15.640	0.000
마산항	수입	-4.223	0.000	-8.218	0.000
	수출	-2.384	0.148	-18.646	0.000

주: 유의수준 0.05

일반적으로 시계열변수 Y_t 가 단위근을 갖는다면 차분된 변수 ΔY 는 안정적으로 나타나는 차분 정상성(difference stationary)을 갖는 것으로 알려져 있다. 따라서 각 항만의 변수를 1차 차분하여 단위근검정을 재 실시 하였다. 재검정 결과 대부분의 자료가 단위근이 없는 것으로 나타나 시계열 자료가 안정화 된 것으로 나타났다.

자료의 정상성의 확보된 후 1997년에서 2004년간 각 항만 물동량 자료를 통해 ARIMA모형의 순서대로 모형의 식별, 추정, 진단을 실시하였다. 각 항만별 선택 모형은 다음 Table 3과 같으며 모든 ARIMA모형에서 계절성이 발견되어 모형에 반영하였다. AIC값과 SBC값을 통해 최적의 모형을 선택하였고 마산항의 경우 모형의 식별결과 MA(1)모형을 따르는 것으로 나타났다.

Table 3 Result of ARIMA model identification

항만	구분	ARIMA Model (p,d,q) (P,D,Q) ₁₂	AIC 통계량	SBC 통계량
부산항	수입	(0,1,2)(0,1,1) ₁₂	1558.7	1566.0
	수출	(0,1,2)(0,1,1) ₁₂	1541.6	1548.8
광양항	수입	(0,1,1)(1,0,0) ₁₂	1148.5	1153.0
	수출	(0,1,1)(1,0,0) ₁₂	1143.6	1148.1
인천항	수입	(0,1,2)(0,1,1) ₁₂	1336.5	1343.7
	수출	(0,1,2)(0,1,1) ₁₂	1275.9	1283.2
울산항	수입	(2,1,0)(0,1,1) ₁₂	1195.7	1202.9
	수출	(0,1,2)(0,1,1) ₁₂	1181.7	1188.9
마산항	수입	(0,1,1)(1,0,0) ₁₂	1137.1	1096.7
	수출	(0,1,1)(1,0,0) ₁₂	1173.7	1178.8

4.3 ANN모형 분석

앞서 분석한 ARIMA모형의 결과와 비교하기 위해 동일한 자료를 통해 신경망모형을 구축하였다. 신경망 모형의 선정을 위해서는 입력층의 유닛의 개수와 은닉층의 layer수와 유닛의 개수를 선택하여야 한다. 우선 입력층의 경우 Harvey(1984)는 $n = \max(p, q + 1)$ 을 갖는 상태공간 방정식으로 표현하였다. 이는 ARIMA의 AR모형의 차수와 MA모형의 차수+1의 개수 중에서 큰 수를 입력유닛의 수로 결정한다는 의미이다. 또한 지원철(1995)의 연구에서는 $n = \max(p + d, q + 1)$ 를 신경망의 입력층으로 사용하였다.

본 연구에서는 입력층의 유닛수를 2, 3, 4개로 은닉층의 유닛수를 1개에서 6개로 변경해서 총 18개의 모형을 비교하여 표본오차와 AIC, SBC 값을 비교해 가장 적절한 모형을 선정하였다.

Table 4 Result of ANN model identification

항만	구분	ANN Model input unit×hidden unit	AIC 통계량	SBC 통계량
부산항	수입	lag-3 × 2 ³	1,892.94	1,921.15
	수출	lag-3 × 2	1,886.31	1,914.52
광양항	수입	lag-2 × 1	1,608.86	1,621.68
	수출	lag-2 × 1	1,607.37	1,620.19
인천항	수입	lag-3 × 5	1,590.93	1,657.60
	수출	lag-4 × 4	1,531.23	1,595.33
울산항	수입	lag-2 × 2	1,406.44	1,429.52
	수출	lag-2 × 1	1,405.92	1,418.74
마산항	수입	lag-4 × 5	1,115.23	1,194.73
	수출	lag-4 × 6	1,137.67	1,232.55

2) 여기서 lag-3의 의미는 y_t 의 예측을 위해 $\{y_{t-1}, y_{t-2}, y_{t-3}\}$ 을 입력층으로 사용한다는 의미이다.

각 항만별 선택된 신경망모형은 다음 Table 4와 같다. 부산항의 수출과 수입화물의 경우 입력층 유닛 3개와 은닉층의 유닛이 2개일 때 AIC, SBC 값이 가장 최적으로 나타났고 광양항의 경우 입력층의 유닛 2개와 1개의 은닉유닛으로 매우 간단한 모형이 선택되었다. 마산항의 경우 자료의 변동이 심하게 나타나 비교적 많은 입력유닛과 은닉유닛을 필요로 하는 것으로 추정되었다.

4.4 하이브리드모형 분석

하이브리드모형에서는 우선 수식 (7)에서 살펴본 ARIMA 모형의 오차항 자료를 신경망모형을 통해 분석한다. 모형의 구축 방법은 앞서의 ANN모형의 구축 방법과 같다.

Fig. 5 Time plot of ARIMA error-term data

오차항의 자료를 분석한 결과 앞서 ANN모형에 비해 더 많은 입력층과 은닉층을 선택한 모형일 경우 평균오차가 감소하는 것으로 나타났다. 따라서 추가적으로 입력층의 수를 5개에서 12개까지 늘렸고 은닉층의 수도 최대 12개로 증가시켜 총 152개의 신경망모형을 비교하였다.

분석결과 선택된 오차항의 신경망모형은 다음 Table 5와 같다.

Table 5 Result of hybrid error-term model identification

항만	구분	None-linear term model input unit×hidden unit	AIC 통계량	SBC 통계량
부산항	수입	lag-5 × 7	1,309.57	1,423.41
	수출	lag-5 × 8	1,286.29	1,416.06
광양항	수입	lag-6 × 6	1,122.44	1,234.00
	수출	lag-5 × 6	1,088.03	1,185.92
인천항	수입	lag-6 × 8	1,079.27	1,227.25
	수출	lag-5 × 7	1,062.07	1,175.90
울산항	수입	lag-5 × 8	1,002.48	1,132.25
	수출	lag-6 × 7	912.46	1,042.24
마산항	수입	lag-5 × 9	735.09	880.80
	수출	lag-6 × 7	752.28	882.05

4.5 분석결과와 해석

앞서 분석한 ARIMA모형, ANN모형, 하이브리드모형의 결과를 비교해 본다. 우선 모형에 사용된 1997년에서 2004년까지의 자료를 통해 구축된 모형이 실제 자료와 어느정도 차이가 있는지 절대표준편차(Mean Absolute Deviation ; MAD)³⁾를 통해 알아보면 다음 Table 6과 같다.

5개 항만의 수출입 자료 모두 하이브리드모형이 ARIMA모형과 ANN모형보다 매우 낮은 오차를 보여 적합도 면에서는 매우 높게 나타났다. 이는 신경망모형의 특성상 실제 모형에 가장 가까운 모델링 능력에 의한 것으로 해석된다.

하이브리드모형을 제외하고 ARIMA모형과 ANN모형을 비교하면 부산항의 경우를 제외하고 다른 8개의 자료에서 ANN에 의한 모형이 ARIMA에 의한 것 보다 높은 적합도를 보이는 것으로 나타났다.

Table 6 Result of models MAD(fitting range: 1999-2004)

항만	구분	ARIMA	ANN	Hybrid
부산항	수입	9,057.0	10,673.8	1,513.6
	수출	7,939.8	9,810.4	589.8
광양항	수입	2,355.8	1,834.1	383.7
	수출	2,176.3	1,664.9	482.9

$$3) \text{MAD} = \frac{\sum_{t=1}^n |e_t|}{n} = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}$$

인천항	수입	2,340.2	1,787.6	151.2
	수출	1,752.3	1,212.4	274.4
울산항	수입	1,041.4	823.9	76.9
	수출	972.0	854.8	76.3
마산항	수입	320.6	175.6	23.6
	수출	359.6	176.6	30.7

앞서 Table 6에서 추정된 모형의 파라메타로 2005년-2006년의 24개월 자료를 예측하고 이를 실제 데이터와 비교해 각각의 모형의 예측력을 분석해 보면 다음 Table 7과 같다.

Table 7 Result of models MAD(test range: 2005-2006)

항만	구분	ARIMA	ANN	Hybrid
부산항	수입	7,180.1	13,996.1	19,530.3
	수출	8,620.5	18,736.1	18,185.7
광양항	수입	3,690.5	3,989.3	4,533.7
	수출	2,558.3	4,084.8	3,576.2
인천항	수입	4,012.6	14,239.5	5,974.3
	수출	3,047.9	7,561.3	5,724.1
울산항	수입	1,344.7	1,455.7	2,350.7
	수출	1,284.4	1,473.1	3,068.2
마산항	수입	288.6	285.6	742.9
	수출	351.7	419.3	768.8

모형의 예측력에서는 거의 모든 항만에서 ARIMA모형이 보다 낮은 편차를 보여 가장 뛰어난 예측 모형으로 선택되었다. 마산항의 경우 수입화물의 경우 ANN모형에 의해 가장 높은 예측력을 보이고 있고 수출화물의 경우에도 ARIMA모형과 큰 차이를 보이지 않았다.

전찬영과 송주미(2007)의 연구와 많은 신경망모형에서도 시계열자료가 만약 지속적인 증가패턴이 나타나는 경우 타 모형에 비해 낮은 예측력을 보인다는 연구결과가 동일한 현상이 나타났다. 마산항의 경우 물동량 시계열 패턴이 불규칙한 증감을 반복하고 있어 ANN모형의 적합도가 높게 나타나고 반면 ARIMA를 통해 MA(1)으로 매우 단순한 모형이 식별되고 있다. 이는 자료가 불규칙적으로 움직일 경우 선형모형인 ARIMA모형에서는 모형의 추정이 사실상 어렵다는 것을 보여주는 것이다.

끝으로 하이브리드모형의 높은 모형 적합도에 비해 가장 낮은 예측력을 보이는 이유를 살펴보면 다음과 같이 해석할 수 있다.

우선 하이브리드모형에 사용된 ARIMA모형의 오차가 비선형적인 패턴이 나타나지 않는 경우이다. 이는 ARIMA모형에서 모형의 패턴이 모두 결정되었기 때문에 오차의 패턴이 백색잡음(white noise)으로 나타났을 가능성이 있기 때문이다. 지원철(1995)의 연구에서도 MLP의 크기가 커지면 추정해야

할 파라메타들인 연결가중치(connection weights) 들이 너무 많아져 주어진 시계열내의 무의미한 잡음까지도 학습하여 MLP의 예측력을 손상시킬 수 있는 문제를 지적했다. 또한 ANN모형의 특성상 과적합(overfitting) 되었을 수 있다. ANN모형은 실제 값과 가장 일치하는 모형을 구축하기 위해 개발되었기 때문에 시계열분석과 같이 자료의 패턴이 과거의 패턴과 미래의 패턴이 완벽히 일치하지 않는 경우에는 예측력이 떨어질 수 있는 경우도 있다. 이는 하나의 은닉층만 가지고 있어도 어떠한 함수에 대해서도 근사시킬 수 있는 신경망모형의 장점이자 단점으로 많은 연구에서도 지적되어온 문제이다.

만약 Zhang(2003)의 연구에서 사용되었던 태양의 흑점자료나 환율자료와 같이 박스컨내에서 일정한 패턴을 지니고 변동하는 자료였다면 하이브리드모형이 높은 적합도와 마찬가지로 높은 예측력을 보였을 것으로 판단된다.

이를 해결하기 위해서 ARIMA모형과 ANN모형의 결합방법을 다양하게 조절한다면 시계열 자료의 선형특성과 비선형특성을 모두 감안한 보다 예측력 높은 모형의 탄생도 예상해 볼 수 있겠다.

5. 결 론

우리나라 항만을 둘러싼 해운환경은 FTA의 확산과 중국의 다국적 물류기지화 등으로 인해 더욱 급속도로 변화하고 있다. 이에 대처하기 위한 효과적인 항만계획의 기본이라 할 수 있는 물동량의 수요예측은 매우 중요한 문제이다. 그러나 과거의 수요예측기법은 5년에서 10년 이상의 장기예측을 다루고 있어 이러한 변화에 민첩하게 대처할 수 없었다.

항만 시설의 부족은 대형선사의 물량이 인접항만으로 빠져나가 항만의 급속한 쇠퇴를 가져올 수 있고 과도한 항만시설은 국가예산의 낭비를 초래하는 문제가 발생한다. 이를 위해 정부에서도 트리거물과 수요예측센터 등을 통해 이러한 문제를 해결하려고 노력하고 있고 근본적인 문제해결을 위한 정확성 높은 수요예측기법이 필요한 실정이다.

본 논문은 최근 가장 많은 연구에서 이용되고 있는 단기에측모형인 ARIMA모형과 ANN모형과 두 모형을 결합한 하이브리드모형으로 국내 주요항만의 물동량을 예측하였다. 분석결과 예측에 사용된 자료에 대한 각각의 예측모형의 적합도는 하이브리드모형이 가장 좋게 나타났으며 다음으로 ANN모형, ARIMA모형의 순으로 나타났다. 모형의 예측력은 이와는 반대로 ARIMA모형이 가장 높은 예측력을 보였으며 다음으로 ANN모형이 뛰어난 것으로 나타났다. 이러한 연구결과는 ANN모형을 통한 예측 연구에서 많이 나타나는 경우로 자료의 증감추세가 있을 경우 모형의 예측력이 낮게 나타나는 모형의 문제점에 기인한다고 하겠다. 가장 낮은 예측력을 보인 하이브리드모형의 경우 ARIMA모형의 오차항을 지나치게 과적합시킴으로 인해 예측력을 손상시키는 것으로 해석할 수 있겠다.

항만 물동량 예측에는 시계열 모형이 아닌 인과모형에 의

한 예측이 많이 사용되었다. 이러한 이유는 항만물동량은 거시경제변수와 매우 밀접한 관계에 있기 때문이다. 기존의 많은 항만관련 예측은 GDP와 같은 거시경제변수나 인구통계 등의 변수를 통해 예측되었다. 그러나 인과모형은 예측에 사용되기 위해 각각의 변수를 재예측해야 하는 문제점이 있다. 즉 항만물동량을 예측하기 위해서는 GDP, 인구예측이 선행되어야 한다. 물론 이러한 예측치들은 공신력있는 기관에서 매년 발표되고 있으나 인과모형의 근본적인 문제해결은 아닌 것이다. 시계열예측 또한 과거자료를 바탕으로 하고 있어 미래의 급변하는 상황을 반영할 수 없는 모형상의 문제를 지니고 있고 이를 극복하기 위한 다양한 모형개발 또한 이뤄지고 있는 실정이다.

어떤 예측기법도 자료의 종류나 형태에 따라 그 예측력이 각양각색이며 항상 높은 예측력을 지닌 모형이란 존재하지 않는다는 것이 정설이다. 본 연구에서 제시한 하이브리드모형은 비록 예측력에서는 타 예측기법에 비해 낮은 예측력을 보였지만 증감추세가 없는 일정한 패턴자료의 인식에는 높은 모델적합도로 인해 도움을 줄 수 있을 것이다. 또한 하이브리드모형의 설명에서도 언급했듯이 시계열 자료에 필연적으로 내포되는 선형적인 특성과 비선형적인 특성을 적절히 감안한 모형의 개발은 기존 ARIMA모형과 ANN모형에 새로운 대안이 될 수 있을 것으로 전망한다.

참 고 문 헌

- [1] 박성영, 이철영(2002), “신경망을 이용한 컨테이너 물동량 예측에 관한 연구,” 한국항해항만학회지 제 26권 제2호, pp.183-188.
- [2] 송일호, 정우수(2002), “SAS와 E-VIEWS를 이용한 계량경제실증분석,” 삼영사, pp.317-318.
- [3] 진찬영, 송주미(2007), “인공신경망모형의 항만물동량 예측 적용에 관한 연구,” 해운물류연구 제53호, pp.65-82.
- [4] 지원철(1995), “신경망을 이용한 시계열 분석 : M1-Competition Data에 대한 예측성과 분석,” 한국전문가시스템학회지 창간호, pp.135-148.
- [5] 최봉호(2005), “수출입 컨테이너화물량의 예측,” 국제통상연구 제 10권 제2호, pp.121-140.
- [6] 해운항만 물류 정보센터, www.spidc.go.kr.
- [7] Box, G. E. P. and Jenkins, G. M.(1976), “Time series Analysis Forecasting and Control,” Holden-Day, San Francisco.
- [8] G. Peter Zhang(2003), “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing 50, pp.159-175.
- [9] Harvey, A. C.(1984), “A Unified View of Statical Forecasting Procedures,” Journal of Forecasting, Vol. 3, pp.245-275.
- [10] Lapedes, A. & Farber, R.(1987), “Nonlinear Signal Processing Using Neural Networks Prediction and System Modeling,” Los Alamos National Laboratory Report, LA-UR-87-2662.
- [11] Nelson, C. R. and Plosser, C. I.(1982), “Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications,” Journal of Monetary Economics, 10 pp. 139-162.
- [12] Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibdon, M., Lewandowski, R. Newton, J., Parzen, E. and Winkler, R. (1982), “The accuracy of extrapolation (time series) methods: results of a forecasting competition,” Journal of Forecasting, Vol 9, pp.5-29.

원고접수일 : 2007년 12월 1일

원고채택일 : 2008년 2월 26일