

부산항 컨테이너 물동량을 이용한 시계열 및 딥러닝 예측연구

이승필* · † 김환성

*한국해양대학교 대학원 물류시스템공학 박사과정생, † 한국해양대학교 물류시스템공학과 교수

Time series and deep learning prediction study Using container Throughput at Busan Port

Seung-Pil Lee* · † Hwan-Seong Kim

*Major of Logistics, Graduate school of Korea Maritime and Ocean University, Busan, 49112, KOREA

† Professor, Dept. of Logistics, Korea Maritime and Ocean University, Busan, 49112, KOREA

요 약 : 최근에는 딥러닝과 빅데이터를 기반으로 한 수요예측 기술이 전자상거래, 물류, 유통 분야의 스마트화를 가속화하고 있다. 특히, 글로벌 운송 네트워크와 현대적인 지능형 물류의 중심인 항만은 4차 산업혁명으로 인한 세계 경제 및 항만 환경의 변화에 발 빠르게 대응하고 있습니다. 항만물동량 예측은 신항만 건설, 항만확장, 터미널 운영 등 다양한 분야에서 중요한 영향을 담당하고 있다. 따라서 본 연구의 목적은 항만 물동량 예측에 자주 쓰이는 시계열 분석과 타 산업에서 좋은 결과를 도출해내고 있는 딥러닝 분석을 비교하여 부산항의 미래 컨테이너 예측에 적합한 예측모델을 제시하는 것이다. 부산항 컨테이너 물동량을 이용하여 학습시키고 그 이후 물동량 예측을 진행하였다. 또한, 상관관계 분석을 통해 물동량 변화와 관련된 외부변수를 선정하여 다변량 딥러닝 예측모델에 적용하였다. 그 결과 부산항 컨테이너 물동량만 이용한 단일변수 예측모델에서 LSTM의 오차가 가장 낮았고, 외부변수를 이용한 다변수 예측모델에서도 LSTM의 성능이 가장 우수하였다.

핵심용어 : 항만물동량 예측, 시계열분석(ARIMA, SARIMA), 딥러닝분석(RNN, LSTM, GRU, ESN)

Abstract : In recent years, technologies forecasting demand based on deep learning and big data have accelerated the smartification of the field of e-commerce, logistics and distribution areas. In particular, ports, which are the center of global transportation networks and modern intelligent logistics, are rapidly responding to changes in the global economy and port environment caused by the 4th industrial revolution. Port traffic forecasting will have an important impact in various fields such as new port construction, port expansion, and terminal operation. Therefore, the purpose of this study is to compare the time series analysis and deep learning analysis, which are often used for port traffic prediction, and to derive a prediction model suitable for the future container prediction of Busan Port. In addition, external variables related to trade volume changes were selected as correlations and applied to the multivariate deep learning prediction model. As a result, it was found that the LSTM error was low in the single-variable prediction model using only Busan Port container freight volume, and the LSTM error was also low in the multivariate prediction model using external variables.

Key words : Port throughput Forecasting, Time Series analysis(ARIMA, SARIMA), Deep Learning analysis(RNN, LSTM, GRU, ESN)

1. 서 론

2016년 세계 경제 포럼(WEF:World Economic Forum)에서 주창된 용어인 ‘제4차 산업 혁명’은 기존 1-3차 산업과 정보통신기술(ICT)의 융합으로 이루어지는 차세대 산업 혁명이다. 이 혁명의 핵심은 인공지능(AI), 빅 데이터(Big Data), 로봇공학, 사물인터넷(IoT), 블록체인, 무인 운송 수단, 3D 프린팅과 같은 7대 분야에서 새로운 기술의 혁신이다. ……(중략)

컨테이너 물동량 예측 연구에서는 전통적인 시계열 예측모델에서 나타나는 국내의 해운항만물류산업의 급격한 변화로 발생하는 오차에 대한 해결점을 딥러닝 예측모델에서 찾고 있다. 또한 딥러닝 기술이 발전함에 있어 다양한 딥러닝 모델을 이용해서 컨테이너 물동량 예측 비교 연구도 활발히 이루어지

고 있다. ……(중략)

그러나 컨테이너 물동량 변동요인을 내·외부적 요인으로 한정하고 있고, 한정된 물동량 변동요인들을 다변량 딥러닝 예측모델의 외부변수로 설정하는 연구들이 대다수이다. 이는 부산항의 강점인 지리적 위치의 영향을 배제한 예측이라고 생각되어 본 연구에서는 내·외부적 요인과 함께 지리적 요인을 추가하여 예측한다. ……(중략)

ESN 모델을 추가하여 단일 변수를 사용한 기존 시계열 및 딥러닝 예측모델과 다양한 외부변수를 사용한 딥러닝 예측모델의 예측 정확도를 비교하여 가장 예측력이 높은 모델을 도출하고자 한다. ……(중략)

본 연구의 서론은 연구의 배경과 연구목적 및 절차로 구성되며 물동량 예측 관한 선행연구 검토 후, 기존 연구와 차별성을

† 교신저자 : 종신회원, kimhs@kmou.ac.kr

* 정회원, pmpsos@kmou.ac.kr

도출한다.(중략)

본 논문에서 지정한 내부적, 외부적, 지리적 요인과 부산항 컨테이너 물동량 간의 상관관계 분석하여 상관성이 높은 요인을 다변량 딥러닝 예측모델의 외부변수로 활용한다.(중략)

2. 분석방법론

물동량 예측을 주제로 한 본 연구에서 크게 두 가지로 분석이 이루어진다. 첫째는 시계열 데이터 분석이고 두 번째는 다양한 예측모델을 활용하여 시계열 데이터를 예측하고 예측모델별 성능 비교분석이다.(중략)

2.1 Hurst 지수

Hurst(1951)는 나일강의 저수정착을 조사 및 분석하기 위해서 1907년부터 40년에 걸쳐 저수량이 시간에 따라 평균 저수량에서 어떻게 변화하는지를 측정하였다.(중략)

2.2 상관관계 분석

통계학과 확률론에서 두 변수가 어떤 선형적 관계인지를 분석하는 방법인 상관관계 분석은 변수가 서로 상관된 관계이거나 독립적 관계일 수 있으며, 이때 변수 간의 관계 강도를 '상관관계'라고 한다.(중략)

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

2.3 시계열 예측모델

Box and Jenkins(1976)에 의해서 이론적으로 체계화된 ARIMA 모델은 AR(Auto-regressive)모델과 MA(Moving Average) 모델을 합친 모델로써 시계열 분석에서 가장 많이 활용되고 있고 과거의 많은 연구를 통해 그 우수성이 입증되었다.(중략)

3. 실증분석

3.1 물동량 데이터

본 연구에서 말하고자 하는 시계열 및 딥러닝 예측연구를 하기 위해 해양수산부 해운항만물류정보시스템인 항만물류통계시스템(PORT-MIS)을 사용하였으며, 2001년 1월부터 2020년 12월까지의 부산항 월별 컨테이너 물동량을 사용하였다.(중략)

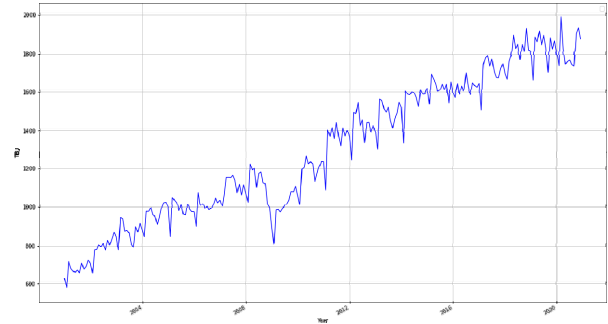


Fig. 1 부산항 컨테이너 물동량 그래프

3.2 단변량 시계열 예측모델

부산항 컨테이너 물동량 예측에 적합한 ARIMA 모델은 AIC 값이 2147.658로 가장 작게 측정된 ARIMA(0,1,1)(0,0,0)이며, SARIMA 모델의 경우 AIC 값이 1895.292로 가장 작게 측정된 SARIMA(2,0,0)(2,1,0)이 예측에 적합한 모델로 도출되었다.

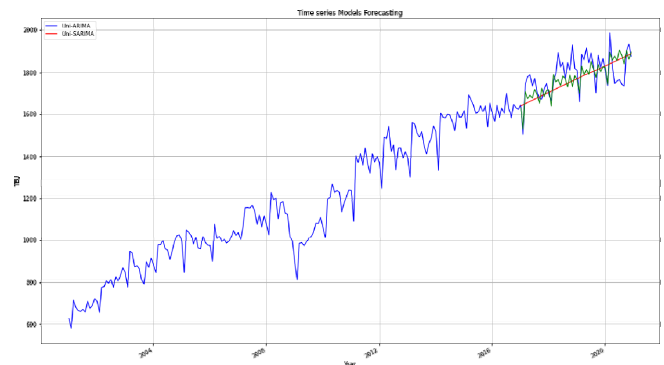


Fig. 2 부산항 컨테이너 물동량 그래프

3.3 실증 결과

[Table 1에서 보여지듯이 단변량 예측모델의 경우 부산항 컨테이너 물동량 예측 성능은 LSTM 예측모델의 MAPE 값이 0.0325, RMSE 값이 73.7017으로 다른 단변량 예측모델에 비해 우수한 모델임을 보여준다. 외부변수를 포함한 다변량 예측모델의 경우 LSTM 예측모델의 MAPE 값이 0.0486, RMSE 값이 109.1219으로 도출되어 다른 다변량 예측모델에 비해 우수한 모델임을 보여준다. (중략)

Table 1 최종 예측모델별 성능 비교

		예측모델명	MAPE	RMSE
단변량	시계열	ARIMA	0.0508	111.5017
		SARIMA	0.0458	102.9992
	딥러닝	RNN	0.0396	85.0243
		LSTM	0.0325	73.7017
		GRU	0.0377	80.6494
		ESN	0.0921	190.4384
다변량	딥러닝	LSTM	0.0486	109.1219
		GRU	0.0529	118.9076
		ESN	0.0648	144.0826

4. 결론

시계열 예측모델과 딥러닝 예측모델 비교에서 단변량 예측모델 중 LSTM 모델이 오차가 작게 도출되었고 다변량 예측모델에서도 LSTM이 가장 오차가 작게 나와 단변량과 다변량 예측에 있어 가장 우수함을 확인하였다. (중략)

또한 데이터 사이언스 관점에서 시계열 예측은 과거 데이터와 현재 데이터의 관계를 분석하여 학습시키는 RNN 종류의 모델이 축적 컴퓨팅의 한 종류인 ESN보다 높은 정확도를 나타냈다. (중략)

참 고 문 헌

- [1] Weikai Yao, "Prediction of container throughput of Dalian Port Based on factor analysis and ARIMA model", IOP Conf. Series: Earth and Environmental Science 831(2021), pp.1-5.
- [2] JianChen Zhang, "Forecast of Shanghai Port throughput Based on ARIMA", IOP Conf. Series: Earth and Environmental Science 831(2021), pp.1-7.
- [3] 김창범, "개입 승법계절 ARIMA와 인공신경망모델을 이용한 해상운송 물동량의 예측", 「한국항만경제학회지」 제31권 1호, 한국항만경제학회(2015), 69-84쪽.