

항만 공컨테이너 재고량 예측을 위한 ARIMA, 머신러닝 적용 연구

백지오*, 강민철**, 설민욱***, 임서정*

*세종대학교 컴퓨터공학과

**서울시립대학교 컴퓨터공학과

***세종대학교 수학과

giopaik@naver.com, minchul46@naver.com, blo9040@naver.com, sjytis14@naver.com

ARIMA, Machine Learning Approach to Forecasting Empty Container Volumes

Gio Paik*, Min-Chul Kang**, Min-Wook Soul***, Seo-Jeong Lim*

*Dept. of Computer Science, Sejong University

**Dept. of Computer Science, University of Seoul

***Dept. of Mathematics, Sejong University

요 약

공컨테이너(Empty Container)는 적컨테이너(Full Container)와 달리, 화물이 적재되지 않은 비어있는 컨테이너로 공컨테이너 재고는 수출에 비해 수입이 많은 항만에서, 수요는 수입에 비해 수출이 많은 항만에서 발생한다. 그러나 수입과 수출은 기간, 지역에 따라 유동적이기 때문에 수요와 재고량 예측에 어려움이 있는데, 본 연구에서는 자기회귀누적이동평균(ARIMA)과 머신러닝 기법을 활용하여 이를 예측하는 방법을 제시한다. 본 연구에 활용된 데이터와 프로그램 소스코드는 [Kaggle](https://www.kaggle.com)에 공개되어 있다.

1. 서론

효율적인 항만 운영을 위해, 필요한 만큼의 공컨테이너를 확보해두는 것은 필수적이다. 항만에 공컨테이너 재고가 부족할 경우 다른 항만에서 공컨테이너를 공수하는데 시간적/금전적 비용이 소요되며, 공컨테이너 재고가 과도하게 편중되어 있을 경우에는 공컨테이너 보관을 위한 비용과, 타 항만으로 공컨테이너를 운반하기 위한 비용이 소요된다.

본 연구는 이러한 비용 소모를 최소화하고, 불필요한 공컨테이너 이동을 최소화하기 위한 공컨테이너 수요 예측 서비스 개발 과정에서, 수요 예측을 위한 미래의 공컨테이너 재고량 예측을 위해 진행되었다.

공컨테이너 재고량 예측에 적합한 방법론을 찾기 위해, 대표적인 시계열 예측 모델인 자기회귀누적이동평균(ARIMA)과 기계학습 기법 중 선형 회귀(Linear Regression) 기법을 시도하여 오차의 절댓값을 비교하였다.

$$(\text{오차}) = |(\text{실제값}) - (\text{예측값})|$$

재고량 예측을 위해 해양수산부 행운항만물류정보 시스템 PORT-MIS¹에서 제공되는 항만별 월간 공컨테이너/적컨테이너 재고량 데이터를 활용하였다.

2. 항만별 컨테이너 재고량 데이터

본 연구에서 구축한 항만별 컨테이너 재고량 데이터는 2018년 1월부터 2020년 5월까지 부산, 인천, 평택/당진, 경인 등 국내 12개의 항만별로 각각 10ft, 20ft, 40ft, 기타 규격별, 국적별 적/공 컨테이너 재고량을 포함하고 있다.

다만, 2020년 2월~3월을 기점으로 폭발적인 코로나-19 바이러스의 유행으로 인해 국제 화물 물동량이 감소한 것을 감안하여, 본 연구에서는 2018년 1월부터 2020년 1월까지의 데이터만을 사용하였다.

데이터는 PORT-MIS에 공개된 국내 23개 항만의 데이터 중, 결측치가 전체 기간 중 50% 이상 존재하는 항만을 제외한 12개 항구의 데이터를 수집하였다. 수집된 데이터에는 각 항구의 10ft, 20ft, 40ft, 기타

¹ <https://new.portmis.go.kr/>

공동저자: 강민철, 설민욱, 임서정

<표 1> ARIMA 예측 결과와 실제의 비교

항만	부산	인천	평택, 당진	경인	대전	군산	목포	광양	포항	마산	울산
실제값	118998	27724	5138	231	1165	844	80	13038	562	44	3599
예측값	119418	26960	5216	232	1168	757	73	12951	455	48	3573
에러(에러율)	420(0.35)	763(2.75)	78(1.52)	1(0.8)	3(0.26)	86(10.28)	6(7.58)	86(0.66)	106(18.95)	4(10.57)	25(0.69)

<표 2> 선형 회귀 예측 결과와 실제의 비교

항만	부산	인천	평택, 당진	경인	대전	군산	목포	광양	포항	마산	울산
실제값	118998	27724	5138	231	1165	844	80	13038	562	44	3599
예측값	105759	25201	4863	126	745	697	67	13133	946	113	3281
에러(에러율)	13238(11.12)	2522(9.09)	274(5.34)	104(45.19)	419(36.00)	146(17.37)	12(15.20)	95(0.73)	384(68.50)	69(158.24)	317(8.82)

규격의 적/공 컨테이너 월별 보유량이 포함되어 있다.

데이터에는 컨테이너가 국적선에서 비롯되었는지 여부가 포함되었으나, 본 실험에서는 컨테이너의 출처는 고려하지 않았고, 가장 물동량이 많은 40ft 컨테이너를 기준으로 실험하였다.

3. ARIMA 모델

ARIMA(Autoregressive Integrated Moving Average) 모델은 y_t 가 어떤 변수 x 가 아니라 y 의 과거값, 시차값과 확률적 오차항에 의해 설명된다고 보는 시계열 예측 방법론이다. [1]

Yasmine et al.은 2011년 4월 ~ 2015년 5월의 유럽 앤트워프 항만의 물류를 ARIMA를 활용하여 예측한 결과, MAPE(평균 절대 백분율 오차) 5.84%의 성능을 얻었다. [2]

한편 Dejan et al.은 유럽 북 아드리아해의 항구들의 물류량을 ARIMA로 예측하였고, 항구에 따라 최소 2.8%에서 최대 7.91%의 MAPE를 얻었다. [3]

ARIMA 모델은 예측하고자 하는 데이터의 과거 데이터만을 기반으로 값을 예측하는데, 이때 자기회귀항의 지수와 이동평균항의 지수, 차분의 차수를 필요로 한다.

이 값들은 예측하고자 하는 데이터에 분포나 특성에 따라 변화하기에 예측하고자 하는 값에 따라 다르게 설정해야 한다. 본 연구에서는 각 변수에 0~3의 값을 대입해보고, 절대 오차가 가장 적은 변수를 취하는 그리드 탐색(Grid Search) 방법을 활용하여 ARIMA의 변수를 설정하였다

ARIMA 모델을 활용하여 국적선, 국외선에서 비롯된 40ft 규격 공컨테이너의 2018년부터 2019년 간의 보유량을 기반으로 2020년 1월의 보유량을 예측한 결과, 실제값 대비 평균 4.95%의 오차를 보이는 예측 결과를 보였다. <표 1>

4. 선형 회귀 모델

선형 회귀(Linear Regression)은 예측하고자 하는 값 y 가 어떤 변수 x_i 들의 선형 결합으로 이루어졌음을 가정하고, 각 x_i 와 y 간의 회귀 계수 w , 오차항 ε 를 경사 하강법을 통해 찾아내어 최종적으로 아래 식

(1)과 같은 예측 모델을 만들어내는 방법이다.

$$y = \varepsilon + \sum_{i=0}^n w_i x_i \quad (1)$$

본 연구에서는 예측하고자 하는 달의 이전 달의 데이터를 기반으로, 예측하고자 하는 달의 공컨테이너 재고량을 예측하는 모델을 개발했다.

선형 회귀는 학습에 사용되는 학습률(learning rate), 학습 횟수(Epochs)를 연구자가 설정해야 하는데, 본 연구에서는 학습률 1000, 학습 횟수를 300으로 통일하고 실험을 진행했다. 각 항만에 따라 학습률과 학습 횟수를 그리드 탐색 등의 방법으로 최적화할 경우, 성능이 현재보다 향상될 것으로 기대된다.

선형 회귀 모델을 활용하여 국적선, 국외선에서 비롯된 40ft 규격 공컨테이너 재고량을 전월 규격별 컨테이너 보유량으로부터 예측한 결과, 모델 학습 데이터에는 포함되지 않은 2020년 1월 데이터를 예측할 시, 평균 34.15%의 오차를 보이는 예측 결과를 보였다. <표 2>

5. 결론

본 연구에서는 국내 항만별 공컨테이너 예측에 ARIMA와 선형 회귀 방법을 적용하고, 모델 학습에 사용되지 않은 2020년 1월 데이터에 대한 각 방법의 정확도를 측정하여 비교하였다.

두 방법을 비교하였을 때, ARIMA의 평균 오차율이 4.95%, 선형 회귀 모델의 평균 오차율이 34.15%로 ARIMA가 우세했다.

다만, 선형 회귀 결과에 일부 항만의 오차율이 비정상적으로 높은 것을 감안하였을 때, 각 항만별 선형 회귀 모델을 최적화할 경우, 모델 성능이 개선될 여지가 있다.

본 연구에서는 항만별 공컨테이너 재고량 예측에 시계열 분석 기술인 ARIMA와 그리드 탐색을 적용하여 지난 2년간의 데이터를 활용하여 평균 4.95%, 최대 18.95%의 오차율로 다음달의 공컨테이너 재고량을 예측하는 것이 충분히 가능성을 실험적으로 보였다.

그러나 본 연구에는 시계열 데이터의 주기적 특성을 반영하는 SARIMA(Seasonal ARIMA) 모델이나, 딥러닝 모델, 특히 시계열 데이터 처리에 유용한 LSTM(Long Short-Term Memory) 모델 등에 대한 연구

가 포함되지 않았기에, 실제 산업 현장에의 적용을 위해서는 해당 분야에 대한 추가 연구를 통해 충분한 신뢰도를 확보할 필요가 있다.

다만, 본 연구에서 항만 공컨테이너 재고량 예측에 시계열 예측 기법의 적용이 가능함을 보였으므로 후속 연구가 충분히 이루어진다면, 국내 항만/항공 물류 관리 등의 분야에 시계열 예측을 도입하여 효율적인 물류 관리가 가능해질 것이라 기대된다.

본 논문은 해양수산부 실무형 해상물류일자리지원 사업의 지원을 통해 수행한 ICT 멘토링 프로젝트의 결과물입니다.

참고문헌

- [1] 민경창, 전영인, 하현구 "계절성 ARIMA 모형을 이용한 항공화물 수요예측: 인천국제공항발 유럽 항공노선을 중심으로" 대한교통학회지 제 31 권 제 3 호, 2013 년
- [2] Yasmine Rashed, Thierry Vanelander "Short-term forecast of container throughput: An ARIMA-intervention model for the port of Antwerp" Maritime Economics & Logistics, 19, 4, 2016
- [3] Dejan Dragan, Tomaž Kramberger, Marko Intihar "A comparison of Methods for Forecasting the Container Throughput in North Adriatic Ports" IAME, Norfolk VA USA, 2014

부 록

- A. 국내 항만별 컨테이너 재고량 데이터셋:
<https://www.kaggle.com/skyl7/monthly-container-holding-of-ports-in-south-korea>
- B. ARIMA 실험 소스코드:
<https://www.kaggle.com/skyl7/forecasting-empty-container-using-arma>
- C. 선형 회귀 실험 소스코드:
<https://www.kaggle.com/skyl7/forecasting-empty-container-with-linear-regression>