Application of ARIMA to Forecast Spatio-Temporal Ship Density

Ho-Kun JEON*,**, Gi Seop LEE*, Hong Yeon CHO*,**,***, Yong Gil PARK*, Chol Yong LEE*
* Korea Institute of Ocean Science and Technology(KIOST)
** University of Science and Technology(UST)
*** Korea Maritime and Ocean University(KMOU)
hkjeon@kiost.ac.kr

Key words: Ship Density, Forecasting, ARIMA

ABSTRACT

With the increased marine traffic in the recent decade, it has become essential to monitor the amount of traffic in the territorial sea because of maritime accidents and overfishing. The state-of-art technologies such as marine telecommunication and satellite observation have been drastically interested in researchers and have enhanced maritime surveillance power. Despite that, the forecasting of ship density is required to arrange the human resources of the coast guard efficiently. Therefore, this paper introduces Auto-Regressive Integrated Moving Average (ARIMA), an analysis model for time-series data, to forecast ship density. Prior to the application of the model, V-Pass data is preprocessed, and then the number of fishing ships within 12 hours (day/night each) and within specific spatial grids is counted. ARIMA model has been generated after checking the stationarity of the data for every spatial grid. The proposed method is revealed that forecast ship density within reasonable error bounds.

1. Introduction

Auto-Regressive Integrated Moving Average (ARIMA) is a time-series model that forecasts values through modeling of past observation (AR), past error (MA), and the difference between two consequent and adjacent observations (I). There have been researches to predict ship traffic density from the record of harbor entrance/leaving and Automatic Identification System (AIS) data (Yoo et al., 2014; Ramin, A., 2020). However, since the ARIMA has only been applied to time-series density on appointed target points in previous research, this paper proposes to expand its application to a broader area than previous research by constructing individual models on every point.

2. Data & Methodology

2.1 Preprocessing

Ship density can be categorized into two categories with the questions. 1) how did long ships stay in a specified space and duration? 2) how did many ships stay in a specified space and duration? For the first question, marine traffic data(hereafter MTD) records should be interpolated and resampled to make the interval of MTD records equal. Meanwhile, MTD data should be ensured whether the interval of MTD is enough to reflect its existence in specified space and duration.

This paper defines ship density with the second question above and the MTD is interpolated with 5 minutes intervals accordingly. Then, Maritime Mobile Service Identity(MMSI) number is checked for the specified space 0.05°×0.05°grid and duration of 12 hours(day: 0600-1800, night: 1800-0600). After counting the number of ships in the specified spatial and temporal condition, time-series MTD is extracted from every grid.

2.2 ARIMA

ARIMA model consists of AR, MA, and difference-based models. AR is based on the concept that past value tells the future value and is expressed as Eq. 1. Meanwhile, MA is based on the concept that past error tells the future value and is expressed as Eq. 2.

$$y_t = b_0 + b_1 y_{(t-1)} + b_2 y_{(t-2)} + \cdots + e_{(t)} \quad \text{Eq. (1)}$$

b : auto-regressive coefficient
y: value observation. $(t-p)$ to $(t-1)$ and forecast(t)
e : error dependent on time

$$y_t = \mu + e_{(t)} + \theta_1 e_{(t-1)} + \theta_2 e_{(t-2)} + \ldots + \theta_q e_{(t-q)} \quad \text{Eq. (2)}$$

μ : arithmetic mean of y

θ: coefficient

e : error depend on time

The difference is obtained from Eq. 3, and the difference is used to make data in stationarity characteristics.

$$z_t = y_{(t)} - y_{(t-1)} \quad \text{Eq. (3)}$$

To construct the ARIMA model, the order of AR, MA, and different should be obtained in advance. Here, the overall procedure is made through Python library {pmdarima}. The pmdarima repeat to model and predict changing its parameter recursively, then automatically derive the best parameter composition.

3. Results and Discussions

The density of ships is calculated for Busan adjacent sea (Long 128.63-129.68E, Lat 34.66-35.25N) in 2018. This paper presents a sample case on the Southeastern sea of Yeongdo island(red dash square).

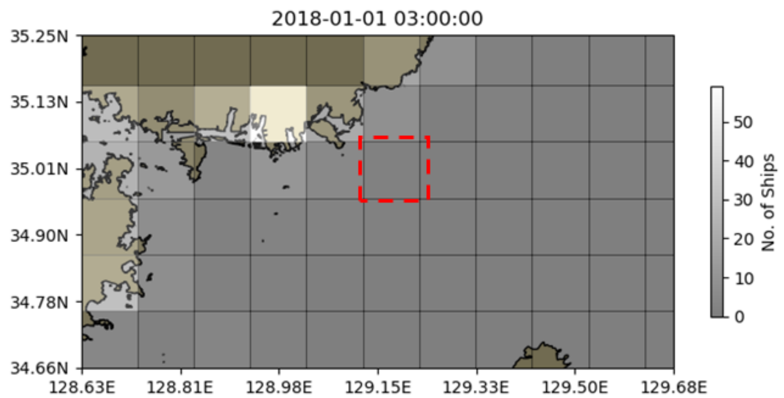

Figure 1. Study area. Red dash square indicates the sample grid of this paper.

The training data is the number of ships at night from Jan. 01 to Feb. 28, 2018, and the test data is from Mar. 01 to Apr. 29. The orders of AR, MA, and the order of difference were derived as P(3), D(1), and Q(0), respectively. Seven-day seasonality is also reflected in this model. From Mar. 01, the observation was updated daily, then a prediction was made for the day after the last observation.

For the traffic density at night, the prediction performance is RMSE 8.6 ships, MAE 6.5 ships, and the correlation coefficient(R) 0.41 were obtained (Fig. 2). Meanwhile, RMSE 15.2 ships, MAE 11.3 ships, and R 0.47 were obtained for daylight prediction (Fig. 3). Although the error is relatively high as 8.6 and 15.2, the R means the prediction is not random. A confidence interval (CI) within 95% with confidence level is displayed on the graphs of day and night. Both graphs show that the predictions were made in the CI.

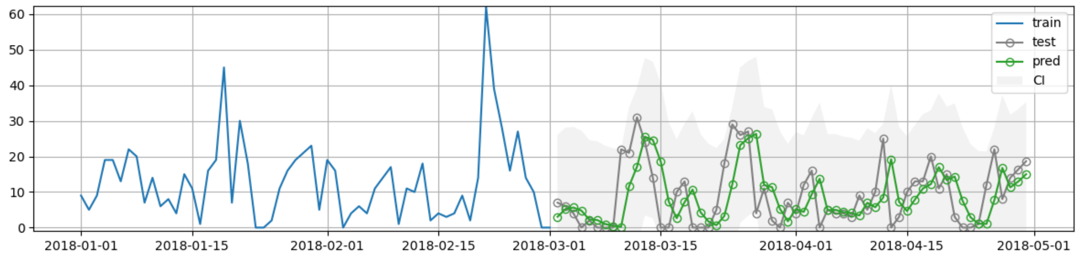

Figure 2. Prediction of traffic density at night.

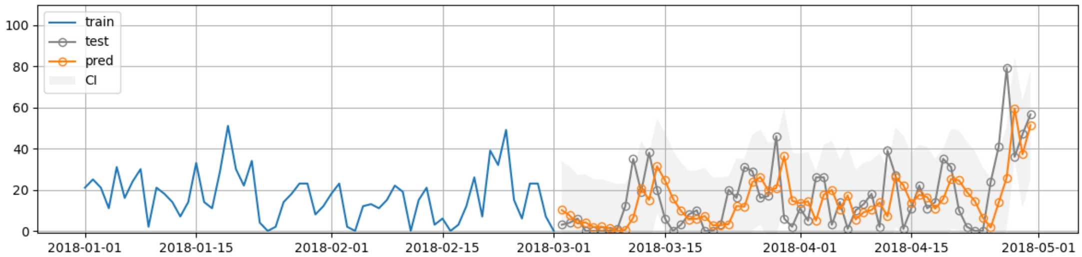

Figure 2. Prediction of traffic density at daylight.

4. Conclusion

ARIMA model was composed after determining the order of AR, MA, and the difference and there were predictions on a daily basis of the ship density by using the past observation of ship density. The prediction is not random but has a low correlation with observation. Thus, the result is not perfect yet, but it shows the feasibility of ship density prediction through the ARIMA model. In the future study, the spatial and temporal resolution will be extended to increase the accuracy to enhance the performance.

Acknowledges

This research is a part of the project entitled "Development of satellite-based system on monitoring and predicting ship distribution in the contiguous zone", funded by the Korea Coast Guard

References

(1) Yoo, S.-L., Kim, J.-S., Jeong, J.-S., Jeong, J.-Y. (2014): A Prediction of Marine Traffic Volume using Artificial Neural Network and Time Series Analysis, Journal of the Korean Society of Marine Environment & Safety, **20**(1), 33-44. (in Korean)

(2) Ramin, A., Mustaffa, M., Ahmad, S. (2020): Prediction of Marine Traffic Density Using Different Time Series -Model From AIS data of Port Klang and Straits of Malacca, Transactions on Maritime Science, **9**(2), 217-223.

(3) Cho, H.-Y. (2021): SARIMA Model Setup and Evaluation Using Monthly CO2 Data, KIDS Report, Korea Institute of Ocean Science and Technology, **3**(3), 38-45.

Author's Biography

Ho-Kun Jeon currently works at the Marine Bigdata Center, Korea Institute of Ocean Science and Technology (KIOST). He is on a Ph.D. at the University of Science and Technology (UST), majoring in Ocean and Coastal Engineering. He has a unique career as a navigation officer in merchant ships. The background enables him to study further maritime matters related to vessel operation. His current research field is marine safety and environment spatial and temporal analysis with remote sensing data.