

Research Article

The Daily Container Volumes Prediction of Storage Yard in Port with Long Short-Term Memory Recurrent Neural Network

Yinping Gao ¹, Daofang Chang,¹ Ting Fang,² and Yiqun Fan³

¹*Institute of Logistics Science and Engineering, Shanghai Maritime University, 201306 Shanghai, China*

²*School of Economics and Management, Shanghai Maritime University, 201306 Shanghai, China*

³*Shanghai Municipal Engineering Design Institute (Group) Co., Ltd., 200092 Shanghai, China*

Correspondence should be addressed to Yinping Gao; gaoyinping@stu.shmtu.edu.cn

Received 16 August 2019; Accepted 5 November 2019; Published 25 December 2019

Academic Editor: Stefano de Luca

Copyright © 2019 Yinping Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The effective forecast of container volumes can provide decision support for port scheduling and operating. In this work, by deep learning the historical dataset, the long short-term memory (LSTM) recurrent neural network (RNN) is used to predict daily volumes of containers which will enter the storage yard. The raw dataset of daily container volumes in a certain port is chosen as the training set and preprocessed with box plot. Then the LSTM model is established with Python and Tensorflow framework. The comparison between LSTM and other prediction methods like ARIMA model and BP neural network is also provided in this study, and the prediction gap of LSTM is lower than other methods. It is promising that the proposed LSTM is helpful to predict the daily volumes of containers.

1. Introduction

In the times of big data, a good forecasting result is helpful to provide decision-makers with strong decision-making basis. The daily container volumes of storage yard refer to the amount of containers which enter the storage yard every day before the container ships enter the port. The prediction of daily container volumes is of great significance to the terminal yard operation plan and the ship loading plan. On the one hand, in the process of making the yard plan, the storage area needs to be preplanned for the container that will enter the yard. The planning of the area depends on the correct predictions of the amount of the containers entering the yard. On the other hand, during the shipment process, the storage location of containers is too centralized or decentralized, which will affect the loading efficiency of the container ship. This requires an accurate prediction of the approach containers so that the yard space can be properly planned. Traditional prediction methods, such as time series prediction methods including exponential smoothing [1–4], grey prediction [5–7], and regression analysis [8, 9], are difficult to make accurate predictions for nonlinear systems with multiple influencing factors such as container voyage volume. Artificial neural

network (ANN) has a good ability of nonlinear approximation and adaptive self-learning. ANN can be divided into three layers which are input layer, hidden layer, and output layer respectively. The Back propagation algorithm is described as follows: by inputting training data into the neural network model, predicted data is the output, and compared with the real data. Then the gap between the predicted data and the real data calculated with loss function is propagated back to the model in order to adjust the parameters which can achieve the goal of improving the accuracy of the model. ANN has been widely used in the planning and prediction of container terminals. There have been studies shown that artificial neural networks can be used to simulate port planning problems associated with container terminals based on historical data, and the predictions can be considered as acceptable [10]. The ANN has also been established between the operational parameters and the static heeling angle, and can provide an accurate estimate of the static heeling angle in order to assess the anchor handling vessel stability [11]. In order to avoid the bottleneck and the smooth integration of container terminals in the supply chain, the ANN has been applied to predict the container dwell time, so as to help terminal operators to make daily decisions regarding stacking

policies, optimal equipment, and human resources allocation [12]. However, under the conditions of finite samples and computational units, it is impossible to simulate more complex mathematical operations, and the data characteristics of a large number of samples are selected according to the prior knowledge of a specific field, ignoring the effective use of the characteristics of the sample data itself.

Deep learning attempts to learn in multiple levels on the basis of different abstraction levels, instead of seeking the functional relationship between the input features and output results directly like RNN [13]. The recurrent neural network (RNN) is mainly used for the analysis and prediction of time series data. The RNN memorize the previous information and apply it to the calculation of the current output which is superior to simple neural network. That is to say, the hidden layer not only has a connection with the current input layer, but also has a connection with the hidden layer at the previous moment. However, the historical information retained by RNN decayed over time, which is called the historical gradient dissipation in the back-propagation process [14]. The vanishing/exploding gradients problem of RNN shows that it cannot be used for data modeling for longer time series. The Long Short-Term Memory (LSTM) is an improved Recurrent Neural Network (RNN) to overcome the vanishing/exploding gradients problem [15]. LSTM can learn long-term and short-term dependence information of time series. Because neural network includes time memory unit, it is suitable for processing and predicting interval and delay events in time series.

In this paper, a container volumes prediction model using deep learning method is proposed and applied to predict the daily volumes of containers which enter the storage yard at the container terminal. The work of this study includes the following: (1) The daily volumes of containers from 2013 to 2017 are chosen as the research dataset, and the box plot is applied to identify the outliers of all data. Then the outliers are replaced with the means before and after the outliers, instead of deleting the outliers directly. (2) The LSTM model is established with the Tensorflow framework and the Python. The preprocessed data is put into the established LSTM network and used to train the LSTM model. Then the model is applied to predict the daily container volumes of future days. (3) The comparison between LSTM, ARIMA, and BP neural network is also given to demonstrate the superiority of LSTM when dealing with the prediction of daily container volumes.

Through observing and training the historical data about container volumes which were transported by container ships to the container terminal, the prediction of daily container volumes is given in this study for the purpose of providing the data support when designing the yard storage plan. In the future research, the same prediction method is projected to make predictions about the volumes of containers which are transported to the storage yard by container trucks. In addition, the scheduling of yard cranes and quay cranes will be provided to complete containers entering and leaving the terminal in an efficient way.

The remainder of this paper is organized as follows. Relevant literatures are reviewed in Section 2. The detailed problem description of daily container volumes prediction is presented in Section 3, and the methodology of LSTM is

provided in Section 4. The experimentation and results are given in Section 5. Conclusions and future research are summarized in the last section, Section 6.

2. Literature Review

There has been some research about the prediction of container throughput. The original series of container throughput was divided into the low-frequency components and the high-frequency components, and the two components were predicted by Autoregressive Integrated Moving Average Model (ARIMA) and Support Vector Regression (SVR) respectively [16]. Chen and Chen [17] compared the prediction results of genetic programming (GP), decomposition approach (X-11), and seasonal auto regression integrated moving average (SARIMA) based on historical data from Taiwan ports and concluded that the GP model is superior to the other two methods. Zhang et al. [18] proposed a combined model composed of grey-forecast model and Logistic-growth-curve model to predict port cargo throughput and improved the accuracy of the forecast model. On electricity price forecasting, Weron [19] made a review and concluded several methods which were divided into multiagent models, fundamental models, reduced-form models, statistical models, and computational intelligence models. Cincotti et al. [20] applied three methods to predict the electricity spot-prices and verified that support vector machine had the best predictive capacity among the ARMA-GARCH, Neural Network, and Support Vector Machine. Twrdy and Batista [21] found a simple but efficient model, which refers to a Markov-chain annual growth rate model, a time-series trend model, a time-series trend model with periodical terms, and a grey system model, to forecast the container throughput based on available data in the Northern Adriatic ports. However, the daily container volumes has the characteristics that are affected by many factors, such as the capacity of container ships, the inland transit time of containers, and seasonal variation of freight transport. It is difficult to express with a certain function relation. The data of daily container volumes is a nonlinear time series, and the time series relationship cannot be explored easily by regression and nonlinear fitting. The recurrent neural network (RNN) in the deep learning domain can generate a memory state of past data when learning sequential data with inherent dependencies. In addition, the long short-term memory which is an improvement RNN can overcome the vanishing/exploding gradients problem caused by RNN.

The LSTM has been widely used in prediction. Cortez et al. [22] proposed the prediction model for emergency event on the basis of LSTM architecture, and made a comparative analysis on the effectiveness of LSTM and traditional time series. The LSTM network was applied to predict out-of-sample directional movements for the constituent stocks of the S&P 500 from 1992 until 2015 [23]. It has reached the conclusion that LSTM network is superior to memory-free classification methods, referring to a random forest, a deep neural net, and a logistic regression classifier. There were two issues in cloud datacenter and they were solved by a workload prediction model, which was developed with LSTM by Kumar

et al. [24]. Moreover, the mean square has been reduced and the prediction has achieved high accuracy. Oehmcke et al. [25] also used LSTM and a unique time dimensionality reduction method to reduce computation time and prediction errors. Nigri et al. [26] applied LSTM architecture into the Lee-Carter model to improve the predictive accuracy of mortality. The comparison with ARIMA was also given and the superiority of LSTM was demonstrated. On the forecast of solar energy, it has been proved that LSTM outperforms a large number of alternative methods with substantial margin and an average forecast skill of 52.2% over the persistence model [27]. Chen et al. [28] made the predictions of returns in the Chinese stock market with the LSTM model, and demonstrated the LSTM is superior to the feedforward neural network model as a time-series model. Based on LSTM model, Tian et al. [29] learned the characteristics of time-dependent and made prediction on the traffic flow. The results with LSTM outperformed than other approaches on traffic flow prediction.

In conclusion, LSTM is a special type of RNN, and it is an RNN that adds long-term and short-term memory functions. It can maintain the durability of the RNN and enable the model to depend on it for a long time. The LSTM network was born to overcome the problem of gradient disappearance. Then LSTM can more effectively capture the nonlinearity and randomness of the data series, and overcome the problem of back propagation of the error through memory blocks. Moreover, LSTM can satisfy the dependence of the data source on the time series, and achieve higher prediction accuracy.

3. Problem Description

With China's Belt and Road Initiative (BRI), port plays a significant role in the construction of the Belt and Road. It can be found that the container throughput has the growing trend in recent years. Taking a port as an example, the annual container throughput in the port was 9.18 million TEU in 2015, 9.6 million TEU in 2017, and 10.30 million TEU in 2017. Continuously increasing container throughput brings pressure to container terminals, and it is essential to manage container terminals for better operations. The management of storage yard is an important part of the entire terminal operations. There are many equipment and resources in the yard, including the yard trucks, yard cranes, and container blocks. In order to improve the operational efficiency, the yard resources, and space must be planned and used rationally. In particular, before the ship enters the anchorage, the container terminal begins to implement the port gathering plan. The storage location and the storage area of containers entering the yard every day need to be planned in advance. Moreover, the area where the containers are located will directly affect the efficiency of the ship loading operations. Therefore, it is necessary to make predictions of daily container volumes for the purpose of planning the storage area, so as to be convenient for ship loading operations after entering the port.

In fact, the volumes of containers to be loaded are unknown before the ship enters the anchorage. One week before the ship enters the port, the terminal starts to conduct the port gathering plan for the ships that are about to enter

the terminal. The container would be stopped for entry 6 hours before the closing time. In general, the staff on the yard predicts the volumes of containers entering the yard on the first day based on historical data and experience, and formulates a yard allocation plan. The amount of containers on the second day is predicted based on the actual amount of containers that entered on the first day. The remaining space on the first day of the yard is used to stack the arrival containers on the second day. If the space requirement of the yard is tight, then the reserved space provided to the approaching containers is very small, which will cause the decentralization of the yard plan. The containers will be piled up everywhere, which is not conducive to the loading and unloading operations of the ship. If a large amount of space is reserved for the containers, it will occupy the storage space of the containers which will be loaded on the other ships, leading to an increase in the turnover rate of the container and affecting the efficiency of the ship loading operations. Therefore, predicting the daily volumes of containers that will enter the storage yard is helpful to make operation plan of the storage yard, and allocate yard space and equipment resources reasonably. Accordingly, the operational efficiency of container terminals can be improved further.

As mentioned above, LSTM is an improvement of RNN. It can coordinate information distribution in historical memory units, and has stronger time series learning ability. Using LSTM for prediction can improve prediction accuracy and reduce errors. Therefore, the daily container volumes are predicted with LSTM-RNN, for the purpose of providing decision basis for operators when making plans at container terminals.

4. Long Short-Term Memory

Long short-term memory (LSTM) was first proposed by Hochreiter and Schmidhuber in 1997 [30]. LSTM is a special type of recurrent neural network (RNN) that adds long-term and short-term memory functions, while RNN is capable of processing sequence data. LSTM can maintain the persistence of RNN and enable the model to depend on it for a long time. In fact, the long-term memory information function is the own behavior of LSTM, different from what it learned through data training. As mentioned above, the standard RNN has a gradient disappearance problem, and LSTM is created to overcome the gradient disappearance problem. The long-term memory function is added to the neural network so that the information no longer decays. However, it will have some trouble in dealing with a huge magnitude of sequences and will be time-consuming when doing that.

The LSTM network consist of one input layer, one or more hidden layers, and one output layer. There are many memory cells in the hidden layers. The structure of LSTM memory cell is shown in Figure 1. The key to LSTM is cell status, with the horizontal red line running through the top of the structure. The cell state is similar to the conveyor belt and running from the previous block (C_{t-1}) to the current block (C_t). There is only a small amount of linear interaction and it is easy to keep the information on top. LSTM has the ability to remove or add information to the cell state through a well-designed structure

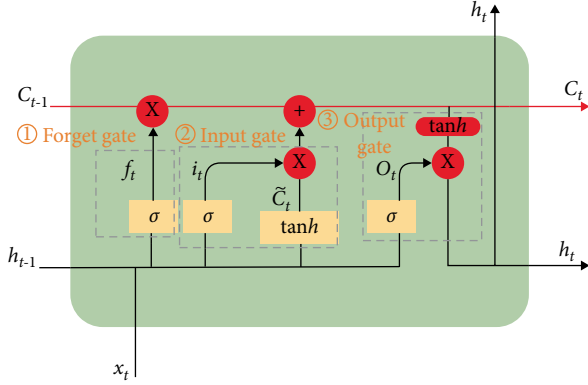


FIGURE 1: The structure of LSTM memory cell.

TABLE 1: The symbol meanings of LSTM algorithm steps.

Symbol	Meaning
f_t	The vectors for the activation values of the forget gate.
i_t	The vectors for the activation values of the input gate.
C_t	The vectors for the cell states.
\tilde{C}_t	The vectors for the candidate values.
o_t	The vectors for the activation values of the output gate.
h_t	The vectors for the output of the LSTM layer.
x_t	The input vector at time t .
$W_{f,x}, W_{f,h}, W_{i,x}, W_{i,h}, W_{\tilde{C},x}, W_{\tilde{C},h}, W_{o,x}, W_{o,h}$	The weight matrices between two gates.
$b_f, b_p, b_{\tilde{C}}, b_o$	The bias vectors of forget gate, input gate, candidate values, and output gate.

called a “gate.” Gates are a way to let information pass selectively. Each memory cell in the hidden layer has three gates, forget gate, input gate, and output gate respectively. The algorithms steps of LSTM are shown as Equations (1)–(6) and meanings of symbols are given in Table 1.

The first step in the LSTM is to decide what information will be discarded from the cell state. This decision was done through the forget gate. The gate reads h_{t-1} and x_t and outputs a value between 0 and 1 to each number in the cell state C_{t-1} . In addition, 1 means “completely reserved,” 0 means “completely discarded.” The function is shown in Equation (1).

$$f_t = \sigma(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f). \quad (1)$$

The second step is to determine what new information is stored in the cell state. This step includes two parts, as described in Equations (2) and (3). First, the sigmoid layer called “input layer” determines what value will be updated. Second, a \tanh layer creates a new candidate vector, which is added to the state.

$$i_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i), \quad (2)$$

$$\tilde{C}_t = \tanh(W_{\tilde{C},x}x_t + W_{\tilde{C},h}h_{t-1} + b_{\tilde{C}}). \quad (3)$$

The third step is to update the cell status. That is to update C_{t-1} to C_t . The old state is multiplied with f_t and the information will be discarded. Then the new candidate is added. The process is given in Equation (4).

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t. \quad (4)$$

The final step is to output. This output will be based on the cell status, but it is also a filtered version. First, a sigmoid layer runs’ to determine which part of the cell’s state will be the output. Next, the state of the cell is processed through the \tanh layer in order to get a value between -1 and 1 , and multiplied with the output of the sigmoid gate. The output h_t is denoted as the following Equations (5) and (6).

$$o_t = \sigma(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o), \quad (5)$$

$$h_t = o_t \times \tanh(C_t). \quad (6)$$

The input gate is the selective recording of the new information into the cell state, while the forgotten gate is the selective forgetting of information in the state of the cell. The output gate is used to control the output value of the cell. Through the cooperation of these three doors, the information contained in the cells can be continuously updated and forgotten, so it is very suitable for processing time series data. More precisely, the daily volumes of the containers which will be transported to the storage yard in the future days have a great relationship with the daily volumes of the containers in the previous period. The predicted features of the forecast have timing information, and then the LSTM can be used to predict the daily volumes of the containers.

Through the LSTM model, the historical data of daily container volumes is used as the input layer and divided into training sets and test sets. Then the data set is standardized, processed, and split into the lower layer processing to train and test the model. The hidden layer uses multilayer LSTM cells to build a circulating neural network, and the output layer outputs the historical data of daily container volumes which has been predicted. Therefore the LSTM method that has high prediction accuracy can be applied to predict the daily volumes of containers which will enter the storage yard for the future days, providing the decision basis when decision-maker made the storage yard plan in the port.

5. Experimentation

5.1. Data Preprocessing. All the experiments are performed on a computer with Intel Core i5 under the Windows 10 operating system using JetBrains PyCharm Community Edition 2017.3.1 x64. The dataset used for research in this paper is from the daily volumes of containers entering the storage yard of a certain port from 2013 to 2017. The raw dataset of the daily container volumes for five years is drawn in Figure 2. It can be

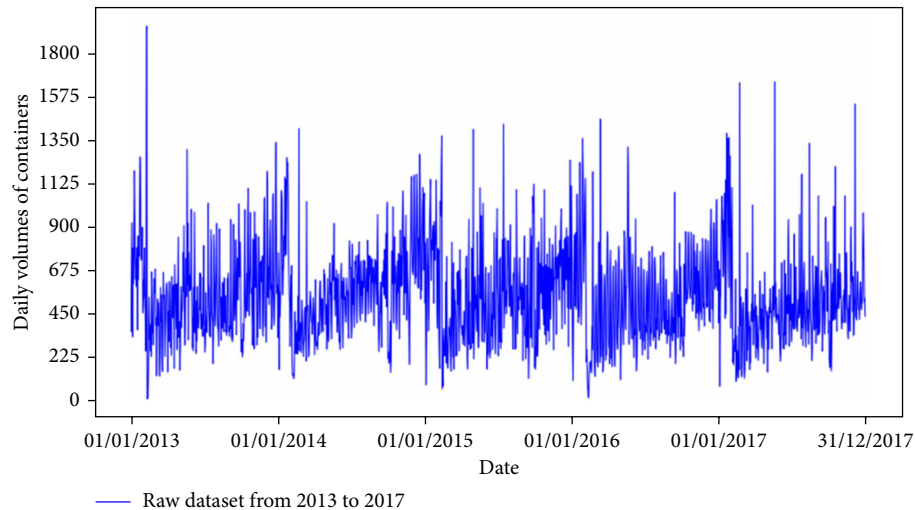


FIGURE 2: The raw dataset of daily container volumes from 2013 to 2017.

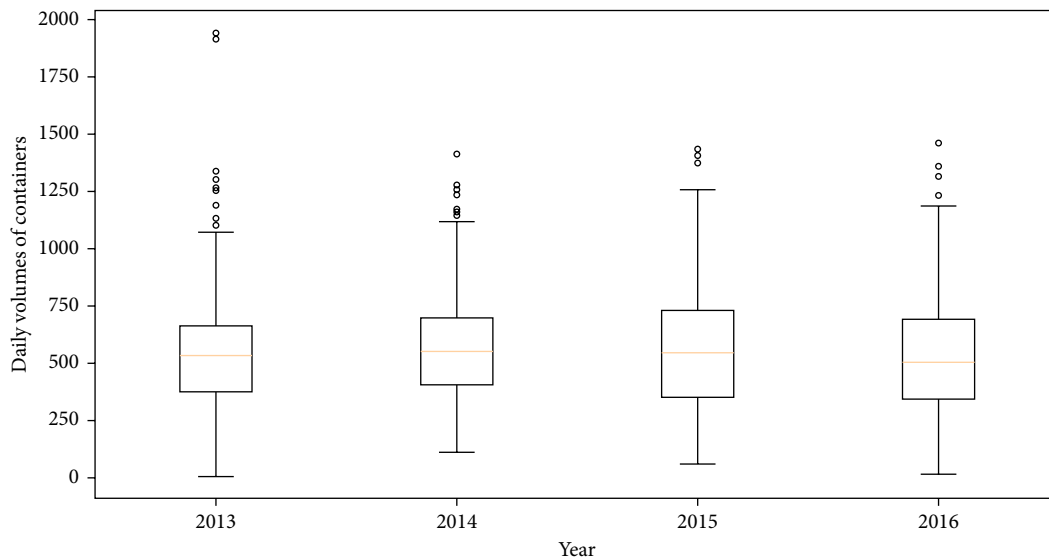


FIGURE 3: The box plot of daily container volumes from 2013 to 2016.

seen that the dataset has the periodical change characteristics, but there are also some data points deviating from the most data. In the process of data collection, there are human factors such as improper operation, aging of equipment, and so on, resulting in the generation of abnormal data and affecting the accuracy of the prediction model possibly. Therefore, it is necessary to identify and deal with the abnormal data in the raw dataset before training the prediction model.

The historical dataset from 2013 to 2016 is chosen to be the training set, and the training set is to be preprocessed. The first step is to detect the abnormal data, which will affect the experimental results. The abnormal data points are individual values in the sample dataset, whose values deviate significantly from the rest of the other data, also known as outliers. The common detection method is box plot, which can show the original appearance of data distribution visually. The criteria for determining outliers in box plot recognition are based on

quartile and interquartile range. Outliers are defined to be less than $(Q_L - 1.5 \times IQR)$ or more than $(Q_U + 1.5 \times IQR)$, where the Q_L , Q_U , and IQR are lower quartiles, upper quartiles, and interquartiles range respectively. The lower quartiles means that 1/4 of all data is smaller than Q_L , while the upper quartiles means that 1/4 of all data is larger than Q_U . The interquartile range IQR represents the difference between Q_L and Q_U , which contains half of all data. The daily container volumes of each year are ranked from the largest to the smallest, and the upper edge, the upper quartile Q_U , the median, the lower quartile Q_L , and the lower edge are calculated respectively. The values which are less than $(Q_L - 1.5 \times IQR)$ or more than $(Q_U + 1.5 \times IQR)$ can be considered as the outliers. The box plot of the daily container volumes from 2013 to 2016 is given in Figure 3.

The second step is to deal with the outliers. If deleting the outliers directly, it may result in the situation of insufficient

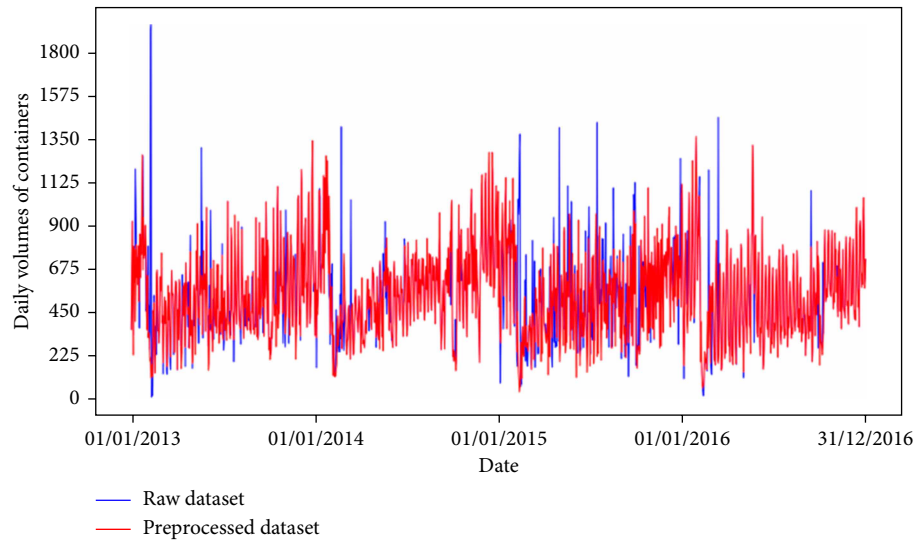


FIGURE 4: The preprocessed dataset of daily container volumes from 2013 to 2016.

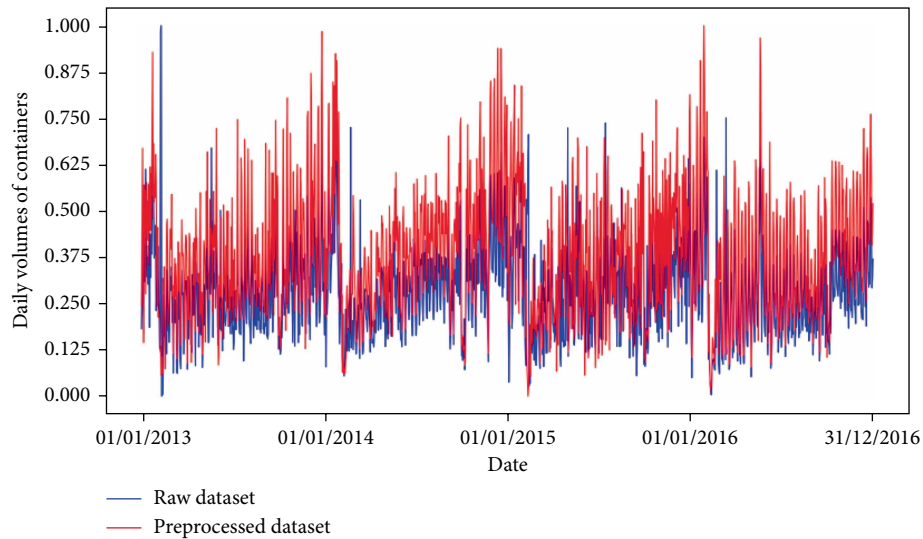


FIGURE 5: The normalized dataset of daily container volumes from 2013 to 2016.

samples and loss of useful information. Another method is to treat the outliers as missing values and then fill them with means. In addition, these outliers are viewed as deleted and filled with the mean of the data before and after one outlier in this paper. After identification and processing of outliers, the preprocessed dataset is shown in Figure 4.

The third step is to normalize the dataset. Normalization is mainly for the convenience of data processing, mapping the data within the range of 0–1. Before data analysis, it is usually necessary to normalize the data and use the standardized data for data analysis. The dataset used in this paper can be scaled to a specified range from 0 to 1, which can be achieved with preprocessing. MinMaxScaler class in Python, as shown in Figure 5.

5.2. Prediction by LSTM. After preprocessing the raw data of the daily container volumes from 2013 to 2016, the prediction

experiment is implemented by using Python and Tensorflow framework. The main purpose of this paper is to predict the daily container volumes in the next few days. As show in Figure 5, the horizontal axis is the time index from January 1st of 2013 to December 31st of 2016, and the vertical axis is the volumes of containers that entered the storage yard each day in a certain port. The data of daily container volumes changes periodically on a weekly basis. According to the changing characteristics of the data set, this paper adopts a rolling forecasting method to improve the accuracy of experiments. For example, for predicting the containers volumes on December 31st, the historical data from November 1st to 7th is used as the input and the historical value of November 8th is used as the label output for the first training. Then the historical data from November 2nd to 8th are taken as the input, and the historical data of the 9th is used as the label output for the second training. The rolling prediction is performed in this way until

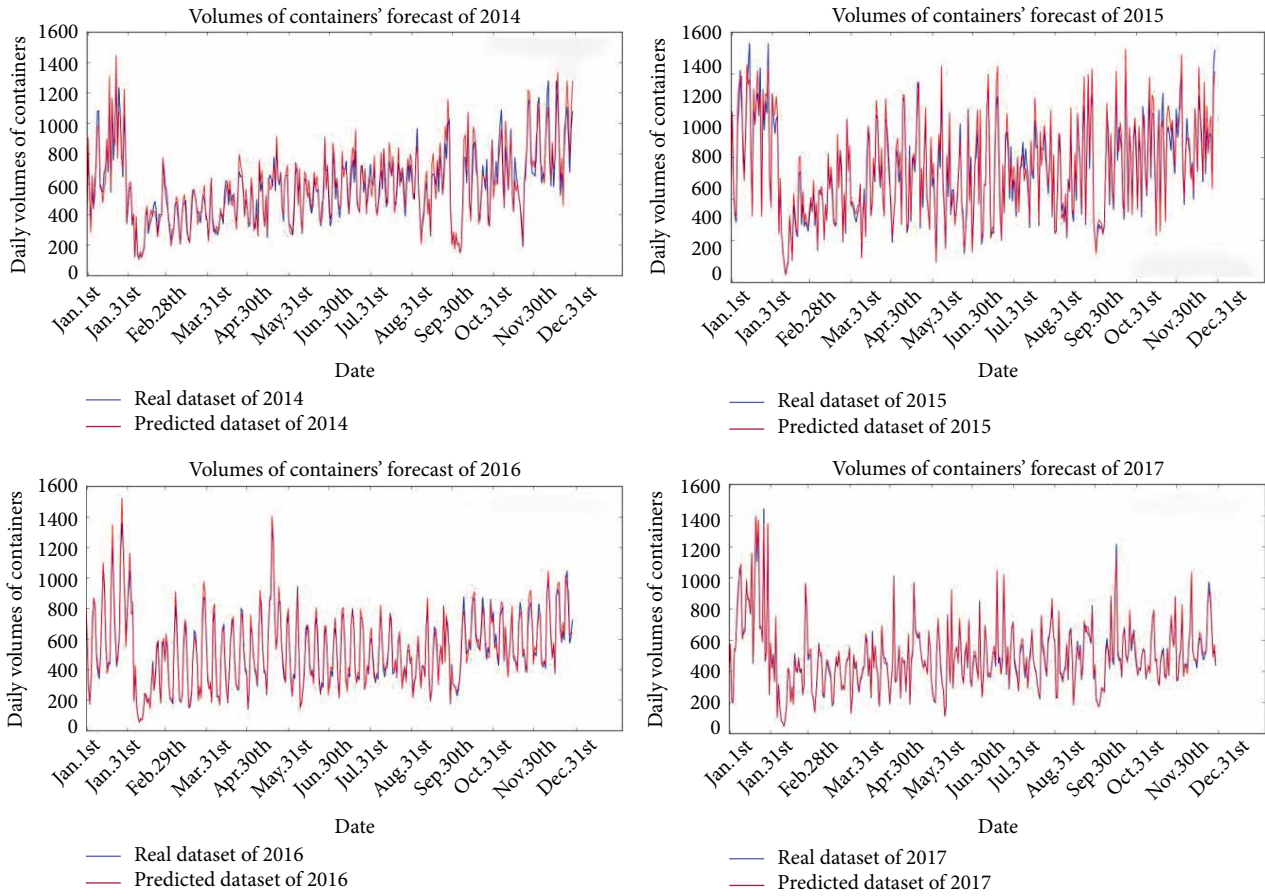


FIGURE 6: The comparison of the real data with the predicted data by LSTM of four years.

TABLE 2: The prediction gap between the real values and predicted values of four years.

The prediction gap	Year			
	2014	2015	2016	2017
RMSE	69.68	53.62	38.70	23.83
Prediction error percentage	9.37%	7.98%	5.69%	3.77%

it obtains the predicted value of December 31st. That is to say, the parameter of time step is set to be 7 in order to show the intrinsic feature of the data set. The input layer is one in the LSTM network, and the output of the LSTM network is the daily container volumes of the next period, so the number of output layer is also one. The batch size is 50, which refers there are 50 data samples to be trained in each batch.

The LSTM model is established with two hidden layers, which have 30 hidden neurons. In addition, the number of model iterations is set as 5000. The learning rate is set as 1×10^{-4} . The keep_prob function is used to handle the overfitting problem. The value of keep_prob at its active level adds a “switch” to each neuron in the layer. The opening probability of the “switch” is the value defined by keep_prob. Once the switch is closed, the output of this neuron will be blocked. Therefore it can balance the importance of each

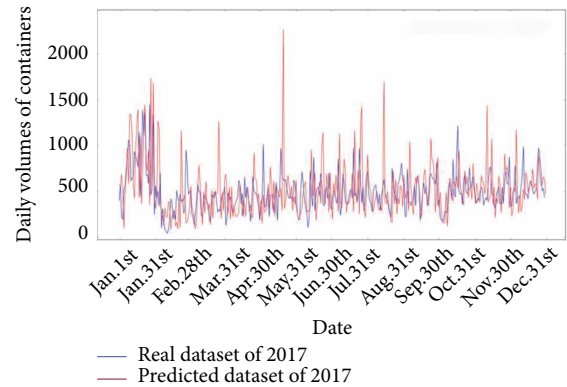


FIGURE 7: The comparison of the real data with the people-predicted data of 2017.

neuron’s role effectively and reduce the risk of overfitting. The activation function is sigmoid, loss function is mean_squared_error, and optimizer parameter to AdamOptimizer.

The total dataset should to be divided into two parts, the training set and test set respectively. The choices of the training and test set were based on multiple experiments. The data set was divided into training and test set according to different proportions, after several experiments, the average value was

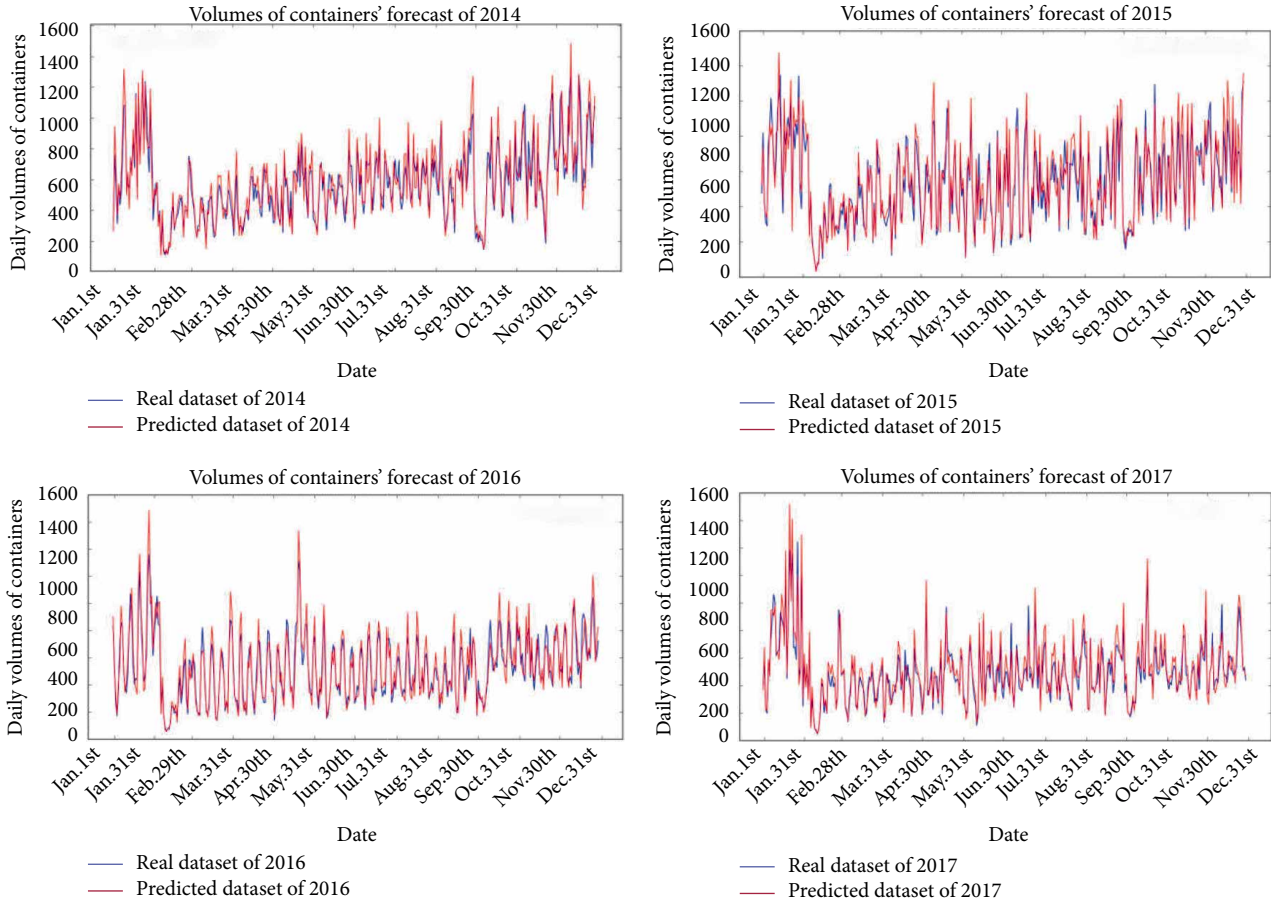


FIGURE 8: The prediction results with ARIMA model.

taken as the evaluation result to reduce the error. After repeating the experimental evaluation, the average value was chosen as the evaluation result of the retention method to reduce the error. Then this study chose 80% of all data as the training set and the rest of the dataset is used as the test set in order to evaluate the ability of the model to predict. The historical data of daily container volumes in 2013 is used to predict the daily container volumes of 2014, the real daily container volumes from 2013 to 2014 are used to predict the daily container volumes of 2015, the real daily container volumes from 2013 to 2015 are used to predict the daily container volumes of 2016, and the real daily container volumes from 2013 to 2016 are used to predict the daily container volumes of 2017. To compare the accuracy of the prediction data, the real data of 2017 needs to be retained and be made comparison with the predicted data by LSTM. Once the data is available, the data can be fed to the LSTM model to build a predictive model.

After training the LSTM model, it is used to make predictions on data, and the comparison of the real data with the predicted data is given in Figure 6. The blue line represents the values of the real data, while the red line is indicated to be the values of the predicted data.

The prediction gap between the real values and predicted values can be indicated by RMSE and prediction error percentage, as shown in Equations (7) and (8). The $x_{p,i}$ represents

each predicted daily container volume and the $x_{r,i}$ represents each real daily container volume. The total number of predicted value and real value is n .

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_{p,i} - x_{r,i})^2}{n}}, \quad (7)$$

$$\text{Prediction error percentage} = \frac{(\sum_{i=1}^n |x_{p,i} - x_{r,i}| / x_{r,i})}{n}. \quad (8)$$

The prediction gap between the real values and predicted values is provided in Table 2.

It can be found from Table 2 that the prediction gap decreases with the increase of training data set. At the beginning, only the historical data of daily container volumes in 2013 was trained to predict the daily container volumes in 2014, and the prediction gap is larger than any other data set. As for the predicted value in 2017, there is relatively less gap with the real data mainly because it is forecasted by the LSTM model which has been trained by large amount of data from 2013 to 2016. That is to say, the LSTM model can be applied to predict the daily container volumes in port and reduce the prediction gap.

In addition, the people-predicted dataset was provided in this study which has been collected from the employees who work in

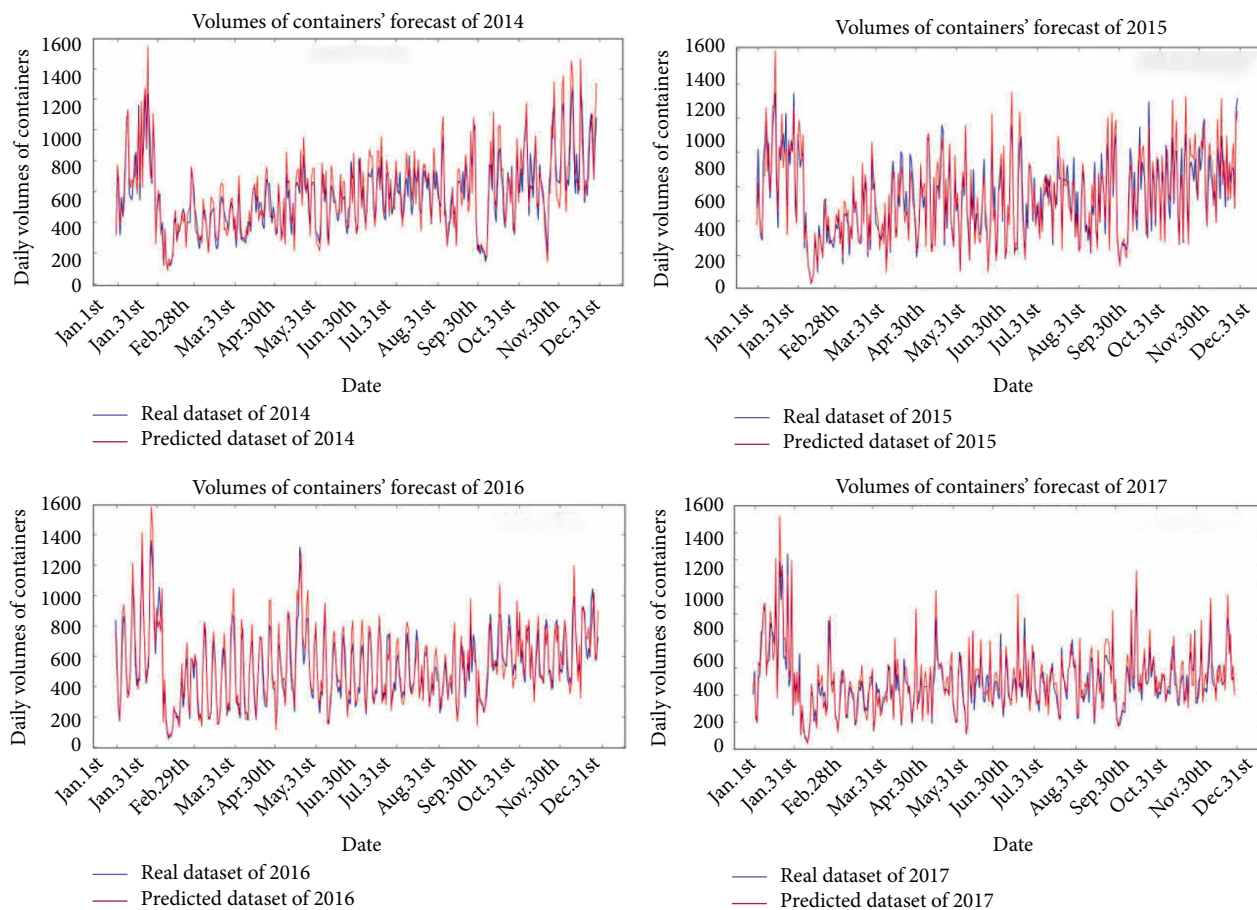


FIGURE 9: The prediction results with BP neural network.

the certain port. The comparison between the real data and the people-predicted data of 2017 is also provided in Figure 7.

Compared with the people-predicted dataset of 2017, the prediction error of daily container volumes between real and people-predicted dataset is 21.22%. If without the scientific prediction of daily container volumes, the employees in the port would make predictions on the basis of their experience, and subjective judgment, leading to a much larger prediction gap and affecting the efficiency of the storage yard. However, the prediction made by the proposed LSTM RNN model is obtained through studying and training of the historical dataset, which is extremely essential to provide decision support for the port operation.

5.3. Comparison between LSTM and Other Prediction Methods. In order to further verify the effectiveness and efficiency of the LSTM method, the Autoregressive Integrated Moving Average (ARIMA) model and the BP neural network are used to predict the daily container volumes of 2017, which can be compared with the prediction result of LSTM. ARIMA is a method of time series prediction and is represented as ARIMA (p, d, q) . In the ARIMA model, p is the autoregressive part of the model and allows us to incorporate the influence of past values into our model, d is the integrated part of the model, and q is the moving average part of the model which allows us to set the error

to a linear combination of the error values observed at previous time points. When applying the ARIMA model to make prediction, the grid search is used in this study to iteratively explore different combinations of parameters, so as to find the value of ARIMA (p, d, q) . For each combination of parameters, the new ARIMA model using the SARIMA function of the stats models module in Python was established and its overall quality was evaluated. Then, $(p, d, q) = (1, 1, 1)$ was used in the ARIMA model to make predictions of daily container volumes in the port, and the prediction result is given in Figure 8. With regard to the BP neural model, the similar parameters with the LSTM model are also provided. Its time step is set to be 7 and it has two hidden layers with 30 hidden neurons. The batch size is 50 and the number of model iterations is set at 5000. The prediction result is given in Figure 9.

The prediction gap between the prediction values and real values of three methods is also provided in Table 3. It can be found that the prediction error of ARIMA model is becoming less mainly because the ARIMA model is based on historical data, and the more historical data collected, the more accurate the model. Compared with the ARIMA model, the RMSE and prediction error percentage of BP neural network are both lower according to the experiments. That is to say, the prediction accuracy of BP neural network is higher than the ARIMA model, which can solve the

TABLE 3: The prediction gap of three prediction methods.

Prediction gap Year	RMSE				Prediction error percentage			
	2014	2015	2016	2017	2014	2015	2016	2017
ARIMA model	89.67	84.85	83.11	78.09	13.00%	12.66%	12.59%	12.37%
BP neural network	87.94	83.38	80.48	77.91	12.63%	12.44%	12.31%	11.94%
LSTM	69.68	53.62	38.70	23.83	9.37%	7.98%	5.69%	3.77%

problems of traditional prediction methods. In addition, the prediction gap of the LSTM model is the lowest which means the LSTM RNN model is superior to other prediction method in predicting the daily container volumes. The LSTM RNN can solve nonlinear and local minima problems, and has stronger data learning ability and generalization ability, which can predict the trend of daily container volumes more accurately and can provide a basis for the decision-maker.

The comparison shows that the prediction of daily container volumes which used LSTM RNN model is superior to other prediction method, and has less prediction error. Therefore, the proposed LSTM RNN model can be applied to make predictions of daily container volumes, so as to provide decision support when managing storage yard.

6. Conclusions

This paper makes prediction of daily volumes of containers which will enter the storage yard in the future days. The historical dataset of daily container volumes is preprocessed by box plot, and the outliers identified are replaced with means. Then the preprocessed dataset is used to train the LSTM RNN model and predict the daily container volumes. In addition, the historical dataset was chosen as the training set and used to make predictions, which also was compared with the real values of each year. More precisely, the ARIMA model and BP neural network were applied to make predictions of daily container volumes and make comparison with the prediction results of the LSTM model. The results show that the structure of the LSTM RNN can be applied to predict the daily container volumes of storage yard, and the prediction gap is lower than other prediction methods. With the prediction of container volumes using the LSTM model, it provides the data support when designing the storage plan including the location and number of containers. Then the space utilization of storage yard is given an opportunity to be maximized, and the rate of turning over containers is going to be reduced as a more precise and detailed plan than before making predictions. Consequently, the ship berthing at the container terminal can be driven away after loading containers in a fast and efficient way, so as to improve the operation efficiency.

Data Availability

The authors have no right to publish the raw data of the port.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Funding

This work was supported by the National Natural Science Foundation of China (71602114, 71631007), Shanghai Science & Technology Committee Research Project (16040501500, 17595810300).

References

- [1] Robert H. Shumway and David S. Stoffer, *Time Series Analysis and Its Applications*, Springer Texts in Statistics, 2000.
- [2] J. Z. Wang, R. L. Jia, W. G. Zhao, J. Wu, and Y. Dong, "Application of the largest Lyapunov exponent and non-linear fractal extrapolation algorithm to short-term load forecasting," *Chaos, Solitons & Fractals*, vol. 45, pp. 1277–1287, 2012.
- [3] J. Z. Wang, S. L. Zhu, W. Y. Zhang, and H. Y. Lu, "Combined modeling for electric load forecasting with adaptive particles warm optimization," *Energy*, vol. 35, pp. 1671–1678, 2010.
- [4] G. Sudheer and A. Suseelatha, "Short term load forecasting using wavelet transform combined with Holt–Winters and weighted nearest neighbor models," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 340–346, 2015.
- [5] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer, Berlin, 1991.
- [6] K.-B. Song, Y.-S. Baek, D. H. Hong, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 96–101, 2005.
- [7] Y. Guo, E. Nazarian, J. Ko, and K. Rajurkar, "Hourly cooling load forecasting using time-indexed ARX models with two-stage weighted leastsquares regression," *Energy Conversion and Management*, vol. 80, pp. 46–53, 2014.
- [8] P. Zhou, B. W. Ang, and K. L. Poh, "A trigonometric grey prediction approach to forecasting electricity demand," *Energy*, vol. 31, pp. 2839–2847, 2006.
- [9] D. Akay and M. Atak, "Grey prediction with rolling mechanism for electricity demand forecasting of Turkey," *Energy*, vol. 32, pp. 1670–1675, 2007.
- [10] T. R. García, N. G. Cancelas, and F. Soler-Flores, "The artificial neural networks to obtain port planning parameters," *Procedia – Social and Behavioral Sciences*, vol. 162, pp. 168–177, 2014.
- [11] G. R. Gunnu and T. Moan, "An assessment of anchor handling vessel stability during anchor handling operations using the

- method of artificial neural networks,” *Ocean Engineering*, vol. 140, pp. 292–308, 2017.
- [12] A. P. IoannaKourouniotti and C. Tsiklidis, “Development of models predicting dwell time of import containers in port container terminals – an artificial neural networks application,” *Transportation Research Procedia*, vol. 14, pp. 243–252, 2016.
- [13] Yu Deng Li and Dong., “Deep Learning: Methods and Applications,” *Foundations & Trends in Signal Processing*, vol. 7, no. 3-4, pp. 197–387, 2014.
- [14] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, “Learning precise timing with LSTM recurrent networks,” *Journal of Machine Learning Research*, vol. 3, pp. 115–143, 2002.
- [15] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: a review and new perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [16] M. Niu, Y. Hu, S. Sun, and Y. Liu, “A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting,” *Applied Mathematical Modelling*, vol. 57, pp. 163–178, 2018.
- [17] S. H. Chen and J. N. Chen, “Forecasting container throughputs at ports using genetic programming,” *Expert Systems with Applications*, vol. 37, no. 3, pp. 2054–2058, 2010.
- [18] C. Zhang, L. Huang, and Z. Zhao, “Research on combination forecast of port cargo throughput based on time series and causality analysis,” *Journal of Industrial Engineering and Management*, vol. 6, no. 1, pp. 124–134, 2013.
- [19] R. Weron, “Electricity price forecasting: a review of the state-of-the-art with a look into the future,” *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.
- [20] S. Cincotti, G. Gallo, L. Ponta, and M. Raberto, “Modeling and forecasting of electricity spot-prices: computational intelligence vs. classical econometrics,” *Ai Communications*, vol. 27, no. 3, pp. 301–314, 2014.
- [21] E. Twrdy and M. Batista, “Modeling of container throughput in Northern Adriatic ports over the period 1990–2013,” *Journal of Transport Geography*, vol. 52, pp. 131–142, 2016.
- [22] B. Cortez, B. Carrera, Y. J. Kim, and J. Y. Jung, “An architecture for emergency event prediction using LSTM recurrent neural networks,” *Expert Systems With Applications*, vol. 97, pp. 315–324, 2018.
- [23] T. Fischer and K. Christopher, “Deep learning with long short-term memory networks for financial market predictions,” *European Journal of Operational Research*, vol. 270, pp. 654–669, 2018.
- [24] J. Kumar, G. Rimsha, and K. S. Ashutosh, “Long short term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters,” *Procedia Computer Science*, vol. 125, pp. 676–682, 2018.
- [25] S. Oehmcke, O. Zielinski, and O. Kramer, “Recurrent neural networks and exponential PAA for virtual marine sensors,” in *International Joint Conference on Neural Networks, IJCNN*, IEEEAnchorage, AK, 14–19 May 2017.
- [26] A. Nigri, S. Levantesi, M. Marino, S. Scognamiglio, and F. Perla, “A deep learning integrated Lee–Carter model,” *Risks*, vol. 7, no. 1, p. 33, 2019.
- [27] S. Srivastava and S. Lessmann, “A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data,” *Solar Energy*, vol. 162, pp. 232–247, 2018.
- [28] K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: a case study of China stock market,” in *Proceedings of the 2015 IEEE International Conference on Big Data (Big Data)*, pp. 2823–2824, Santa Clara, CA, 29 October–1 November 2015.
- [29] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, “LSTM-based traffic flow prediction with missing data,” *Neurocomputing*, vol. 318, pp. 297–305, 2018.
- [30] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

