

Multi-dimensional prediction method based on Bi-LSTM for ship roll

Yuchao Wang, Hui Wang, Bin Zhou, Huixuan Fu ^{*}

College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, 150001, China

ARTICLE INFO

Keywords:

Ship motion prediction
Deep learning
CNN
LSTM
Bi-LSTM

ABSTRACT

When ships sail in the sea, they will move irregularly due to the influence of strong wind, waves and other complex marine environment. Among them, the ship roll is very important to the safety of ship navigation. In order to ensure navigation safety, it is necessary to predict the ship roll effectively and accurately in advance to provide a basis for ship advance control. Aiming at the problem that traditional methods and machine learning methods are not accurate for ship roll prediction, a single input single output (SISO) and multiple input single output (MISO) ship roll prediction methods based on deep learning are proposed, and the influence of input variables on the ship roll prediction model is studied. Firstly, a single input Bi-LSTM for ship roll prediction method is proposed. The network takes the advantage of LSTM time series prediction and combines convolution kernel to extract cross time features. Then the analysis and test results of six different single input prediction models are given based on real ship data. Secondly, a deep bidirectional feature network with multiple inputs of ship roll angle, roll angular velocity, relative wind speed, relative wind direction, turning angle and rudder angle is proposed for ship roll prediction. The network uses Bi-LSTM to extract forward and reverse information respectively, and introduces two Bi-LSTM branch structures to explore the deep features between multiple input data to improve the accuracy of ship roll prediction. Finally, four error indexes were used to evaluate the proposed single input and multiple input ship roll prediction algorithms on real ship data. The applicable conditions of the single input and multiple input ship roll prediction algorithms are obtained, and the effectiveness of the proposed algorithms are also verified.

1. Introduction

Due to the coupling effect of uncertain external forces such as sea wind, wave and ocean current, when the ship is sailing at sea, the overall force of the ship is very complicated, and the ship will swing irregularly. The motion state of a ship sailing on the sea includes rolling, pitching, heaving, turning, yawing, etc. Among the various posture parameters of the ship, the excessive roll angle of the ship poses the greatest threat to the safety of the ship. It will not only affect the comfort of the crew, but also cause great safety hazards to the navigation of the ship. On the basis of mastering the ship's motions, if it is possible to predict the ship's motion changes in the future, more time can be given to take remedial measures, pre-control the ship in advance, and greatly improve the safety and stability of the ship. In addition, when the shipboard radar tracking system, precision instruments and other stable platforms perform real-time control, short-term prediction of the motion becomes particularly important, and the stable platform needs to be controlled according to the predicted yaw angle. However, the precondition of pre-

control is to predict the change of ship's motion in advance. The prediction problem of ship roll can be regarded as a time series prediction problem. The introduction shows in detail the methods used by researchers in different periods to solve the time series prediction problem before the deep learning model methods was proposed in the field of time series prediction.

At present, ship's motion prediction methods are mainly divided into three categories: mathematical model, statistical model and machine learning model. The ship has six degrees of freedom in the waves: roll, pitch, yaw, surge, sway and heave. The mathematical model method is to establish a mathematical model of ship motion and a mathematical model of the marine environment. Use the mathematic model established by the information of the ship's available posture parameters in the current coordinate system, ship design parameters, wind speed, and wind direction etc. to make predictions.

Jiang H. et al. used an AR model to research the effect of spectrum band-width, peak frequency and hull scale on ship motion prediction (Jiang et al., 2020). Yin J. C. et al. used wavelet transform and RBF

* Corresponding author.

E-mail address: fuhuixuan@hrbeu.edu.cn (H. Fu).

network, proposing a combined roll prediction model (Yin et al., 2018). Bu S. X. et al. firstly studied the influence of radiation force and diffraction force on the rolling recovery moment under different roll angle and wave amplitude based on the body exact method. Secondly, the heave-roll-pitch coupling model based on three-dimensional mixed source method is used to study the effects of radiation force and diffraction force on parametric roll angle through simulation experiments (Bu et al., 2019). Yu L. et al. used a five-degree-of-freedom nonlinear time-domain model based on potential flow to quantitatively predict the roll angle of a KCS container ship (Yu et al., 2019). Although mathematical model prediction does not require huge history data, the establishment of mathematical model requires a more in-depth study of the modeling system. In addition, due to the inevitable need to add random interference and select empirical parameters to the model, it leads to problems such as large prediction errors and poor anti-interference performance.

Commonly used statistical prediction methods include regression analysis method (Li et al., 2016), grey theory (Huang and Huang, 2000), fuzzy theory (Sharma et al., 2021), time series method (Cai et al., 2019), etc. Statistical methods generally rely on a large amount of accurate input-output data, and use curve fitting, parameter estimation and other methods to solve the mapping relationship between input-output data. Gong X. et al. proposed a bottom-up forecasting with Markov-based error reduction method to predict power consumption of aggregated DEWHs for multiple forecast horizons (Gong et al., 2019). Jiménez J. et al. used the multivariate statistical analysis method to successfully make a long-term forecast of energy demand (Jiménez et al., 2019). De Lima Silva P.C. et al. proposed a prediction method based on fuzzy time series, which uses fuzzy and random patterns of data to predict points, intervals and distributions (De Lima Silva et al., 2020). Although statistical model prediction does not require precise modeling of the system, statistical models need to process huge and accurate history data, which is costly and requires a long time to obtain prediction results.

The machine learning algorithms used in the field of time series prediction mainly include support vector machines, decision trees, artificial neural networks (ANN) and other methods. Its advantage lies in the powerful ability of complex nonlinear fitting. Many researchers have proven that machine learning methods have significant advantages over the other two methods, and machine learning methods have become one of the most commonly used methods for time series prediction. Zhang P. C. et al. proposed a short-term rainfall prediction method based on multilayer perceptrons (Zhang, 2018). Gilanifar M. et al. used a multi-task learning framework to predict household electricity load. The Bayesian Spatiotemporal Gaussian Process Model (BSGP) within the framework was used to describe the correlation between different residential communities, a low-ranked dirty model and iterative algorithm are used to update model parameters under the multi-task learning framework (Gilanifar et al., 2020). Wu D. et al. used multi-kernel learning in residential power load forecasting, which has more flexibility than traditional nuclear method (Wu et al., 2020). Based on SVM and a two-step hybrid parameter optimization method, Jiang H. et al. completed a high-precision, high-resolution short-term load forecasting (Jiang et al., 2018). Inspired by kernel methods and support vector machines, Yuan K. et al. proposed a time series prediction framework based on kernel mapping and high-order HFCM filtering (Yuan et al., 2020). Among the machine learning methods, artificial neural networks can learn network weights from data sets, and have a stronger ability to express complex nonlinearities. Almonacid F. et al. used the neural network model method to compare three classical prediction methods on the real photovoltaic power generation data set, and the results showed that the neural network method has an error of 6%–8% lower than other methods (Almonacid et al., 2011). Raza M. Q. et al. proposed an integrated prediction framework, which consists of three multiple predictors, namely, Elman neural network, feedforward neural network and radial basis function neural network. Compared with ARIMA and reverse-propagation neural network models, the prediction accuracy of

the integrated prediction framework is significantly improved (Raza et al., 2020). Li C. et al. proposed a novel neural fuzzy method for the hourly wind speed prediction (Li et al., 2017). Faraji J. et al. used adaptive network based Fuzzy Inference System, multilayer perceptron and radial basis function neural network to accurately predict load and weather data (Faraji et al., 2020). Sun N. et al. proposed an uncertainty prediction system for predicting unstable time series. In the designed architecture, the adaptive variational mode decomposition was first used to extract hidden information from the original time series. Then used random forest to select the appropriate input for each mode. Finally, various neural networks (extreme learning machine, back propagation neural network, generalized regression neural network and radial basis function neural network) produced prediction results in a non-linear manner (Sun et al., 2020). Liu H. et al. proposed the MLP wind speed prediction algorithm and the adaptive fuzzy network wind speed prediction algorithm. On the real wind speed data set, the MLP method performed better than the adaptive neural network (Liu et al., 2015). Obviously, the artificial neural network method has achieved better results than other methods of machine learning, but the neural network model used in the above literature has a small number of hidden layers, which cannot completely extract the features in the time series, and blindly increases the network layer will not only increase the amount of calculation, but even cause the network to have problems such as vanishing gradient, exploding gradient, and degrading prediction effects.

Until 2006, Hinton G.E. proposed Deep Belief Network (DBN) (Hinton et al., 2006). It alleviates the problem that deep neural networks are difficult to train, and promotes the development of artificial neural networks. In the following years, deep learning (deep neural network model algorithm) was proposed and successfully applied in various fields. So far, the classic neural network models include Convolutional Neural Networks, Recurrent Neural Networks, and Deep Belief Networks, etc. Kuremoto T. et al. applied deep neural network to the field of time series prediction, and compared with MLP model, the results showed that the deep neural network has stronger nonlinear expression ability and better prediction effect (Kuremoto et al., 2014). Ouyang T. et al. put forward a deep learning framework based on DBN, and compared with SVM and ELM in the power load forecasting data set. The results showed that the prediction accuracy of deep neural network is better than that of shallow neural network and SVM, and the effectiveness of the framework was verified (Ouyang et al., 2019). Since the training process of the deep belief network is to fully train the parameters of the first layer network and fix the parameters of the first layer network, then fully train the parameters of the second layer network and fix it until the last layer. The way will lead to a slower training process.

As the neural network layer continues to deepen, the amount of network parameters and the difficulty of training also continue to increase. Convolutional neural networks use the idea of local correlation and weight sharing, which greatly reduces the amount of network parameters, improves training efficiency, and makes it easier to implement ultra-large deep networks. Ma X. et al. proposed an interval prediction method based on Gaussian distribution and improved convolution neural network (CNN) (Ma and Dong, 2020). Barra S. et al. used convolution network to predict the financial situation (Barra et al., 2020). Sezer O.B. et al. used a deep convolutional neural network to predict transaction volume in the field of financial transactions, and has achieved good prediction results (Sezer and Ozbayoglu, 2018). Although the convolutional neural network has a strong ability to extract features, it can only extract the time features of local data, and cannot make predictions based on all history data of the time series.

Long Short-Term Memory (LSTM) is a variant of recurrent neural network proposed by Hochreiter S., which can generate prediction results based on the contextual information of time series data (Hochreiter, 1997). The memory unit and gating mechanism of LSTM alleviate the problems of gradient disappearance and gradient dispersion that are prone to traditional recurrent neural network (RNN).

Sagheer A. et al. predicted oil production by stacking multiple LSTM units (Sagheer and Kotb, 2019). By comparing the LSTM model with the ARIMA model, Wang Z. et al. found that the prediction error of the LSTM model is about 4% smaller than that of the ARIMA model (Wang et al., 2019a). Tan M. et al. not only proposed an integrated learning prediction model based on LSTM networks, but also proposed a loss function for comprehensive peak demand prediction errors based on the principle of Bias-Variance Tradeoff. Achieved high-precision power demand forecasting (Tan et al., 2020). Hossain M.S. et al. proposed a prediction algorithm that used the LSTM model to predict photovoltaic power generation. The proposed prediction model has achieved superior prediction effects in the field of photovoltaic power generation prediction (Hossain and Mahmood, 2020). Sun Z. et al. proposed a hybrid short-term wind power prediction model composed of variational mode decomposition (VMD), k-means and LSTM, which solved the problem of wind power generation prediction (Sun et al., 2019).

The feature extraction ability of a single neural network is limited. Mixing different neural networks can give play to the advantages of respective networks to obtain better results. Wang K. et al. used the LSTM-CNN hybrid network to predict photovoltaic power generation, and used 8 error indexes to evaluate the LSTM, CNN, and LSTM-CNN networks. The results showed that the hybrid prediction model has better prediction effects than the single prediction model (Wang et al., 2019b). Shi X. et al. used the CNN-LSTM network to predict rainfall, and the CNN-LSTM model reduced the mean square error by 23.8% compared with the FC-LSTM model (Shi et al., 2015). Sajjad M. et al. developed an energy prediction model based on hybrid sequential learning, which applies convolutional neural network and gate recurrent unit (GRU) to an accurate energy consumption prediction framework, and achieved small error rate prediction on electrical energy prediction and personal household power consumption data set (Sajjad et al., 2020). Sun Z. et al. used the method of combining the variational mode decomposition (VMD), Conv-LSTM and error analysis to predict the short-term wind power. The experimental results showed that the model has excellent prediction performance for the wind power sequences that are difficult to capture (Sun and Zhao, 2020). Alhussein M. et al. proposed a deep learning framework based on convolutional neural networks and LSTM. The average absolute percentage error on the household public electricity load data set is 40.38% (Alhussein et al., 2020). He K. et al. proposed the concept of residual structure in 2015, which solved the difficult problem of deep neural network training (He et al., 2016). Chen K. J. et al. applied the concept of residual structure to the field of power load forecasting, and compared with the ELM and ARIMA models. The experimental results showed that the proposed framework performs well in forecasting (Chen et al., 2018). Lai G. et al. used convolution neural network and recurrent neural network to extract the variation rules between variables. On the seasonal time series data set, the prediction results were better than the traditional linear model and GRU neural network (Lai et al., 2018). Qin Y. et al. used CNN-LSTM network to predict power generation and power demand using wind power signal data (Qin et al., 2019). Wang K. et al. Compared CLSTM, LSTM and CNN on real photovoltaic power data set, and the results showed that CLSTM model had the best effect (Wang et al., 2019c). Wang et al. used the attention mechanism and recurrent neural network technology to achieve the prediction of ship roll, but the attention mechanism increased the calculation amount (Wang et al., 2021).

A method of ship roll prediction using deep learning technology is proposed in this paper. At present, the deep learning method has not been applied in the field of ship roll prediction, but it shows strong ability to deal with the time series prediction problems in many other fields, such as photovoltaic power prediction, wind speed prediction, rainfall prediction, etc. Therefore, the deep learning technology is introduced to solve the problem of ship roll prediction. The contribution of this paper is to provide a method of using deep learning technology to solve the problem of ship roll prediction.

The main structure of the paper is described as follows: Section 2 introduces the neural network structure; Section 3 describes the proposed model structure; Section 4 shows the data description and evaluation indexes; Section 5 shows the simulation results and analysis; Section 6 summarizes the conclusion and discusses future work.

2. Neural network feature extraction structure

2.1. Time series feature extraction

The LSTM network can extract contextual information of data and is widely used in the field of time series. The memory unit of the LSTM is the long-term and short-term information memory vector inside the LSTM, and the gating mechanism is the forgetting gate, input gate, and output gate that control information forgetting, input, and output. The principle of LSTM can be summarized as equations (1)–(5):

$$f_t = \sigma(W_{xf}X_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \quad (2)$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o) \quad (3)$$

$$C_t = f_t C_{t-1} + i_t \tanh(W_{xc}X_t + W_{hc}h_{t-1} + b_c) \quad (4)$$

$$h_t = O_t \tanh(C_t) \quad (5)$$

Where, h_t and C_t are the short-term memory vector (output vector) and long-term memory vector at time t respectively. i_t , O_t , and f_t are the control vectors of input gate, output gate and forgetting gate, C_{t-1} and h_{t-1} are the long-term memory vector and short-term memory vector (output vector) at time $t-1$, respectively, and X_t is the input at time t . W_{xf} , W_{hf} , W_{cf} , W_{xi} , W_{hi} , W_{ci} , W_{xo} , W_{ho} , W_{co} , W_{xc} , and W_{hc} are the corresponding weight, and b_c is the bias. The detailed structure of LSTM is shown in Fig. 1.

In Fig. 1, the green part represents the forgetting gate, the red part represents the input gate, the blue part represents the output gate, and the orange part represents the update information. Both RNN and LSTM can only rely on single direction time sequence information to predict the output of the next time, but in some problems, the output of the current time is not only related to the state of forward time feature, but also related to the state of reverse time feature. In order to extract the two directions time characteristics of the time series separately, Bi-LSTM is stacked by two LSTM layers (Schuster and Paliwal, 1997). The first layer of LSTM inputs the data forward, the second layer of LSTM inputs the data in the reverse direction, and the output is composed of the two layers of LSTM. The structure of Bi-LSTM is shown in Fig. 2.

In Fig. 2, LSTM is the structure shown in Fig. 1, $X = [x_1, x_2, x_3, \dots, x_{n-1}, x_n]$ represents input data, $O = [o_1, o_2, o_3, \dots, o_{n-1}, o_n]$ represents output characteristics. The first LSTM layer represents the forward

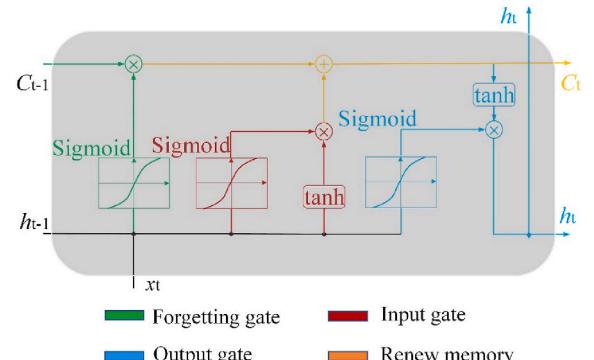


Fig. 1. The structure of the LSTM.

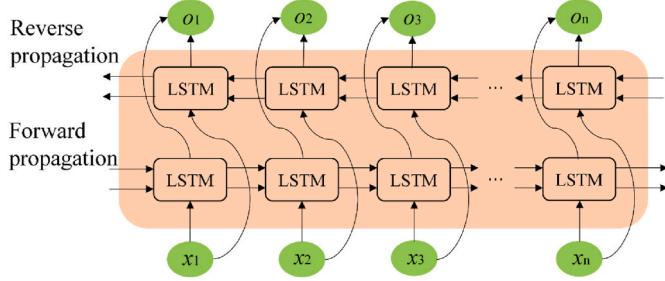


Fig. 2. The structure of Bi-LSTM.

propagation layer, and the second LSTM represents the reverse propagation layer.

2.2. Time pattern feature extraction

Convolutional neural networks use the idea of weight sharing and local correlation to solve the problem of large amount of model parameters and difficult training. The feature extraction of the convolution layer mainly relies on the convolution kernel, and each convolution kernel and the feature are operated to generate a feature mapping channel. The operation of the convolutional layer is shown in (6)-(7):

$$y_j = \left(\sum_{i \in C_j} t_i w_{i,j} \right) + b_j \quad (6)$$

$$t_i = f(y_j) \quad (7)$$

Where, t_i is the feature map of the convolutional layer, C_j is the input feature map, b_j is the bias, y_j is the output of the convolution, $w_{i,j}$ is the convolution kernel, and f is the activation function.

Fig. 3 shows the calculation process of the convolutional layer.

In Fig. 3, The green area represents the calculation area with convolution kernel, the red box represents convolution kernel, and the orange area represents convolution calculation result.

3. Single input Bi-LSTMC model and multiple-input deep bidirectional feature model for ship roll angle prediction

3.1. Single input Bi-LSTMC ship roll angle prediction model

In the case of single input, the input data is the ship roll angle. A hybrid network prediction model Bi-LSTMC for ship roll angle is proposed. The Bi-LSTM network is used to extract the bidirectional features of the ship roll angle history information, and then the convolutional neural network is used to extract the deep information of a single feature at different sampling times. Fig. 4 shows the structure of the Bi-LSTMC ship roll angle prediction model.

In Fig. 4, the 9 sampling times of the history data of the ship roll angle are input into the Bi-LSTM network as a group, the characteristics of the history data are extracted, and the data with the format of (9,

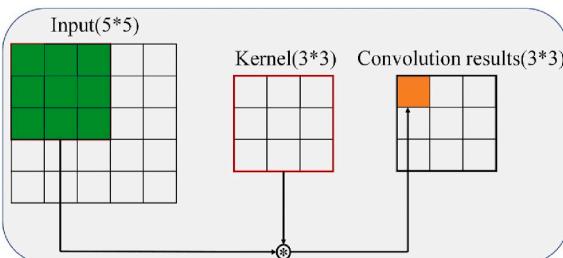


Fig. 3. The calculation process of convolution layer.

200,1) is output. The first nine sampling times are used for model training, and the data at the tenth sampling time is the real value of the predicted data. The feature dimensions of the two directions extracted by Bi-LSTM are set to 200, and the unidirectional feature memory vector is 100 dimensions. The data in the last dimension of (9, 200, 1) represents the number of input variables. The data with the shape of (9,200,1) is used as the input of the convolutional layer to extract the deep information of single feature at different sampling times. The model uses a Dropout layer to prevent overfitting, and the parameter is set to 0.5.

3.2. Multi-input deep bidirectional feature network ship roll angle prediction model

There is a certain potential mapping relationship between the change of ship roll angle and relative wind speed, relative wind direction, ship roll angle velocity, rudder angle, turning angle and other data. If the network can fully extract the characteristics of each type of data, and extract the relationship and change rules between the characteristics, it will improve the accuracy of ship roll angle prediction. The input data of the Model B includes ship roll angle, relative wind speed, relative wind direction, roll acceleration, trim, and rudder angle. The output data of the model is ship roll angle. Every ten sampling times is set as a group. The first nine sampling times are used for model training, the tenth sampling time is used for model prediction. The dimension of the input variable is 6. Fig. 5 shows the multi-input deep bidirectional feature network structure.

In Fig. 5, the red dashed box represents the Bi-LSTM layer, the blue dashed box represents the Dropout layer, the green dashed box represents the LSTM layer, and the purple dashed box represents the convolutional layer. History data includes ship roll angle, relative wind speed, relative wind direction, ship roll angle velocity, ship rudder angle, and ship rotation angle. The output of the network is the ship roll angle. The shape of the output data of Bi-LSTM layer is (9,200,6). In (9, 200, 6), the numeric of the last dimension represents the number of input variables, and the network has six input variables. The first nine sampling time times are used for model training, and the data at the tenth sampling time is the real value of the predicted data. The Bi-LSTM network can extract the forward characteristics and reverse characteristics of history data through forward LSTM layer and reverse LSTM layer, respectively. Then through two Bi-LSTMC branch networks, the forward feature and the reverse feature are extracted respectively. Finally, the deep forward and reverse features are stacked, and the predicted values are output through one LSTM layer and two fully connected layers.

4. Data description and evaluation indexes

4.1. Dataset and experimental environment

The dataset used in the experiment is the real data of the ship sailing on the sea, which mainly includes ship roll angle, roll angular velocity, relative wind speed, relative wind direction, turning angle and rudder angle. Fig. 6 is part of the raw data of the ship's motion dataset. The data sampling interval is 1s, and there are total of 32371 samples. The training data is the first 80% of the total samples, and the test data is the last 20% of the total samples. The division of the dataset is shown in Fig. 7.

In the sampling process, the collected data will not only be mixed with noise, but also have null data points. Therefore, it is necessary to preprocess the data before it is input into the network model. The data preprocessing process mainly includes data interpolation and filling in missing data. Fig. 8 is the result of the interpolation processing for part of the ship's motion dataset.

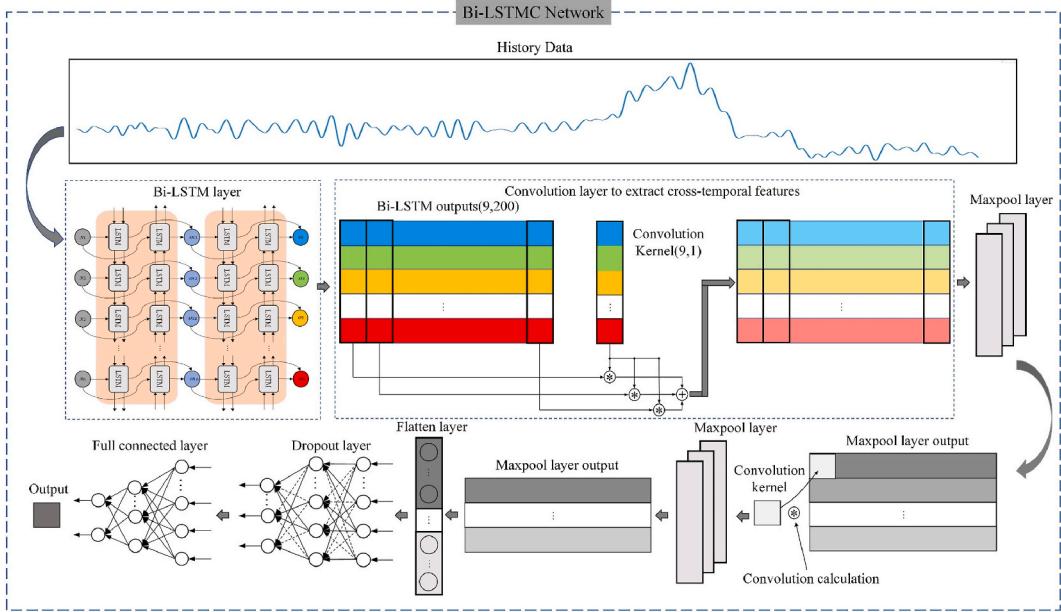


Fig. 4. The structure of the single input Bi-LSTM-C.

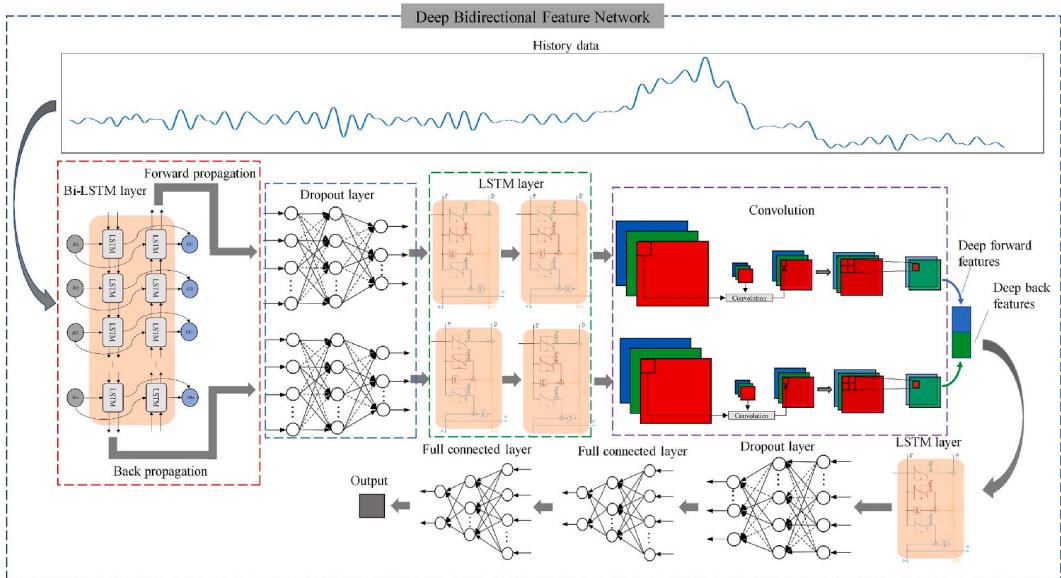


Fig. 5. The structure of multi-input deep bidirectional feature network.

4.2. Evaluation indexes

The performance of the model needs to be evaluated by evaluation indexes. The four error evaluation indexes selected in the experiment include mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE). MSE reflects the average difference between the real value and the predicted value, MAPE represents the average degree of deviation of the predicted value from the real value. Because the magnitude of the value in the data set is very small, if MSE is used as the loss function, the loss function will be close to 0, but the prediction effect is very poor. Therefore, the pivotal index MAPE is used as the loss function in the simulation experiment. The MAPE is the most key error index in the experiment and can best reflect the prediction performance of ship roll angle. The simulation experiments also use four promotion percentage indexes, promoting mean absolute error (PMAE), promoting mean

square error (PMSE), promoting root mean square error (PRMSE) and promoting mean absolute percentage error (PMAPE) to characterize the performance improvement of the proposed model compared with other models, the formulas are shown in (8)–(15):

$$\text{MAE} = \frac{(\sum_{i=1}^N |y(i) - \hat{y}(i)|)}{N} \quad (8)$$

$$\text{MSE} = \frac{\left(\sum_{i=1}^N |y(i) - \hat{y}(i)|^2 \right)}{N} \quad (9)$$

$$\text{MAPE} = \frac{\left(\sum_{i=1}^N |y(i) - \hat{y}(i)| / y(i) \right)}{N} \quad (10)$$

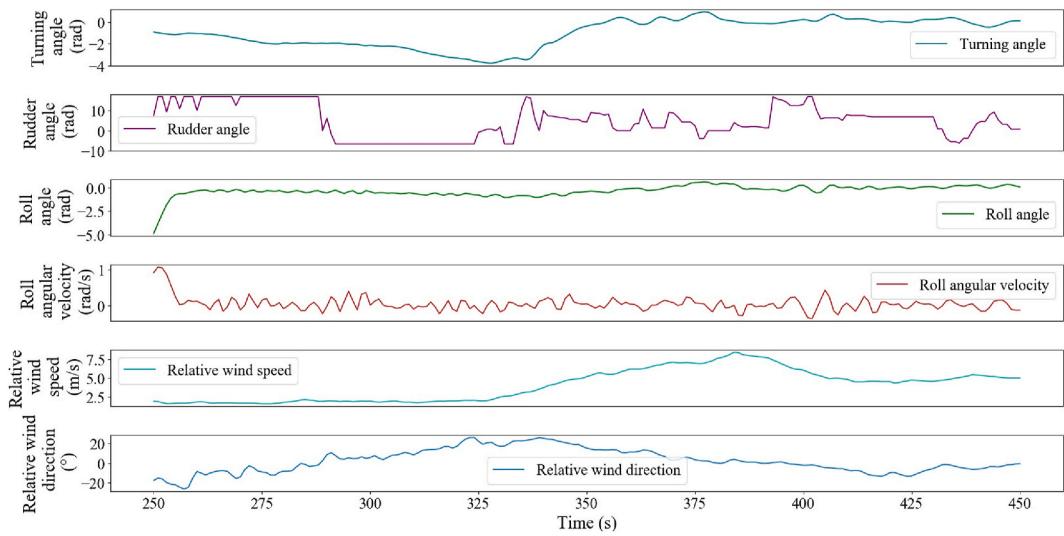


Fig. 6. Partial raw data of the ship's motion dataset.

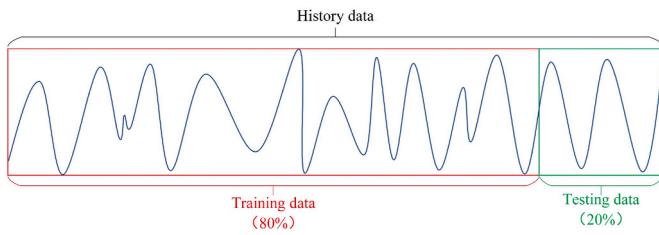


Fig. 7. Division of dataset.

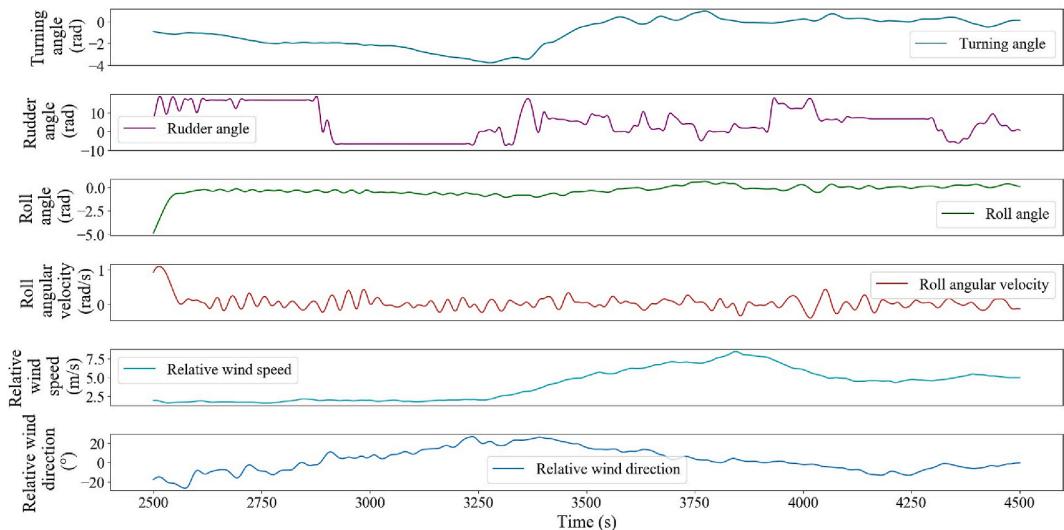


Fig. 8. Partial data of the ship's motion dataset by data interpolation. In Fig. 6, the change rate of the data at the adjacent sampling time is large. Fig. 8 shows the data after interpolation, and the change rate of the data at the adjacent sampling time is small. The experiment was carried out on the computer installed with Python 3.6. The main software versions are shown in Table 1.

$$\text{RMSE} = \sqrt{\frac{\left(\sum_{i=1}^N |y(i) - \hat{y}(i)|^2\right)}{N}} \quad (11)$$

$$\text{PMAE} = \frac{\text{MAE}_1 - \text{MAE}_2}{\text{MAE}_1} \quad (12)$$

$$\text{PMSE} = \frac{\text{MSE}_1 - \text{MSE}_2}{\text{RMSE}_1} \quad (13)$$

$$\text{PMAPE} = \frac{\text{MAPE}_1 - \text{MAPE}_2}{\text{MAPE}_1} \quad (14)$$

$$\text{PRMSE} = \frac{\text{RMSE}_1 - \text{RMSE}_2}{\text{RMSE}_1} \quad (15)$$

Where, $y(i)$ is the real value, $\hat{y}(i)$ is the predicted value, and N is the total number of sample points.

5. Simulation results and analysis

5.1. Experimental results of ship roll angle prediction with single input

Under the condition of single input, Bi-LSTM-C is proposed in Section 3.1, hereinafter referred to as Model A. Model A contains a layer of Bi-LSTM. The memory vector in the forward propagation layer is 100 dimensions, and the memory vector in the reverse propagation layer is also 100 dimensions, and each sampling time outputs a state vector. The convolution part uses two layers 2-D convolution operation, and the number of convolution kernels is set to 1. More detailed parameters in Model A are shown in Table 2.

In Table 2, Units represents the number of neurons, Filters represents the number of convolution kernels, go backwards represents whether the LSTM layer is a reverse propagation layer, padding represents the filling method, and Kernel size represents the size of convolution kernels. The size of the first layer convolution kernel is set to (9, 1). Fig. 9 shows the complete roll angle prediction process of single input Bi-LSTM-C model.

Fig. 9 shows the prediction process of the Bi-LSTM-C model for ship roll angle. Firstly, the original roll data is preprocessed and input the Bi-LSTM network to extract the bidirectional time features of the data. Then extract deep features across time through the convolution part. Finally, the predicted values are output from the fully connected layer, and the evaluation indexes are used to evaluate the performance of the model.

The simulation experiments used six models, including the neural network model and the traditional model. They are LSTM, GRU, Conv-LSTM, SVM, ARIMA and Model A respectively. Figs. 10–11 shows the prediction results of the six models.

Table 3 shows the prediction errors of the six models on the ship roll angle data, including MSE, RMSE, MAPE, and MAE. The bold font indicates the maximum or minimum value.

It can be seen from Table 3 that the performance of the hybrid network is generally better than that of the GRU neural network model and the traditional model. The MSE of the Conv-LSTM network is 0.012, the MAPE is 14.0%, the MAE is 0.088, and the RMSE is 0.110, the most

Table 1
Hardware and software versions.

Hardware and software	Configuration
Desktop computer	Operating System: Ubuntu 18.04.3 LTS CPU: Intel® Core™ i7-10750H CPU@2.60GHz × 6 RAM: 16 G GPU: GeForce GTX 2060 (6G memory)
Software version	Pycharm-2020 + Tensorflow-2.0.0 + Keras-2.3.1 + NVIDIA Driver-440.1 + CUDNN-7.6.4 + CUDA-10.2

Table 2
Detailed configuration of Model A.

Network layer	Layer parameter setting	
Bi-LSTM-layer1	Units = 100, go backwards = F	Epoch = 50
Bi-LSTM-layer2	Units = 100, go backwards = T	Batch size = 64
Conv2D	Filters = 1, Kernel_size=(9,1) s = 1, Padding = 'same'	
MaxPool	s = 1	
Conv2D	Filters = 1, Kernel size = 1 s = 1, Padding = 'same'	
MaxPool	s = 1	
Dropout	Rate = 0.5	
Fully connected	Units = 100	
Fully connected	Units = 1	

important predictor, MAPE, is only slightly higher than Model A's MAPE by 0.002. The MSE of the LSTM network is slightly higher than that of Conv-LSTM and Model A by 0.004 and 0.006, respectively. The MAPE of the LSTM network is 10.83% and 17.20% higher than that of Conv-LSTM and Model A, respectively. The performance of single GRU network is relatively poor, and various error indicators are higher than other similar neural networks. The MSE, MAE, MAPE, and RMSE of the GRU network are 0.065, 0.225, 29.8%, and 0.255, respectively. The prediction accuracy of SVM is significantly higher than that of GRU network, but the prediction accuracy is lower than that of LSTM, Conv-LSTM, and Model A networks. The MSE of SVM is 0.032, MAPE is 27.9%, MAE is 0.168, and RMSE is 0.179. The prediction accuracy of the ARIMA model is higher than that of SVM and GRU. The MSE of ARIMA is 0.019, the MAPE is 20.9%, the MAE is 0.111, and the RMSE is 0.138. It can be seen from Table 2 and the above analysis that the performance of the hybrid model will be slightly higher than that of the single model, which rely on the ability of convolutional networks in extracting features between variables. In order to conveniently show the difference between different models and Model A, the promotion percentage indexes are used for evaluation, as shown in Table 4.

The promotion percentage indexes describe the difference between the two models on the same error indexes. Table 4 shows the promotion percentage improvement of Model A compared to other models, and the results are as follows: 1) In terms of PMAPE, the MAPE of Model A is 7.14% lower than that of Conv-LSTM model, and 17.19%, 56.38%, 53.40% and 37.80% lower than that of LSTM, GRU, SVM and ARIMA model, respectively, the prediction effect of Conv-LSTM model is second only to Model A in the case of single input. 2) In terms of PMSE, Model A compared with LSTM, GRU, Conv-LSTM, SVM, ARIMA, the MSE of Model A is reduced by 37.50%, 84.61%, 16.67%, 68.75%, 47.37%, respectively. 3) In terms of PMAE, the best performer is also Model A. Compared with the LSTM, GRU, Conv-LSTM, SVM, and ARIMA, the MAE of Model A has been reduced by 20.39%, 63.56%, 6.81%, 51.19%, 26.13%, respectively. 4) PRMSE has the same performance as PMSE. The PRMSE of LSTM, GRU, Conv-LSTM, SVM, and ARIMA are 20.63%, 60.78%, 9.10%, 44.13%, 27.53%, respectively.

In order to show in detail, the deviation of the predicted value of roll angle from the real value after 1000s, Figs. 12–17 shows the absolute value of the deviation between the predicted ship roll value and the real ship roll value of the six models at each sampling time, denoted by Error = $|y_{\text{real}} - y_{\text{pre}}|$.

Figs. 12–17 shows the ship roll prediction error curves of ARIMA, SVM, GRU, LSTM, Conv-LSTM and Model A respectively. The error curve is represented by green solid line. In Fig. 12, the Error curve of ARIMA is mainly in the range of (0, 0.2). In Fig. 13, the Error curve of SVM is mainly in the range of (0.1, 0.3), the Error curve of the SVM model is generally higher than that of the ARIMA model, indicating that the prediction effect of the ARIMA model is better than that of the SVM

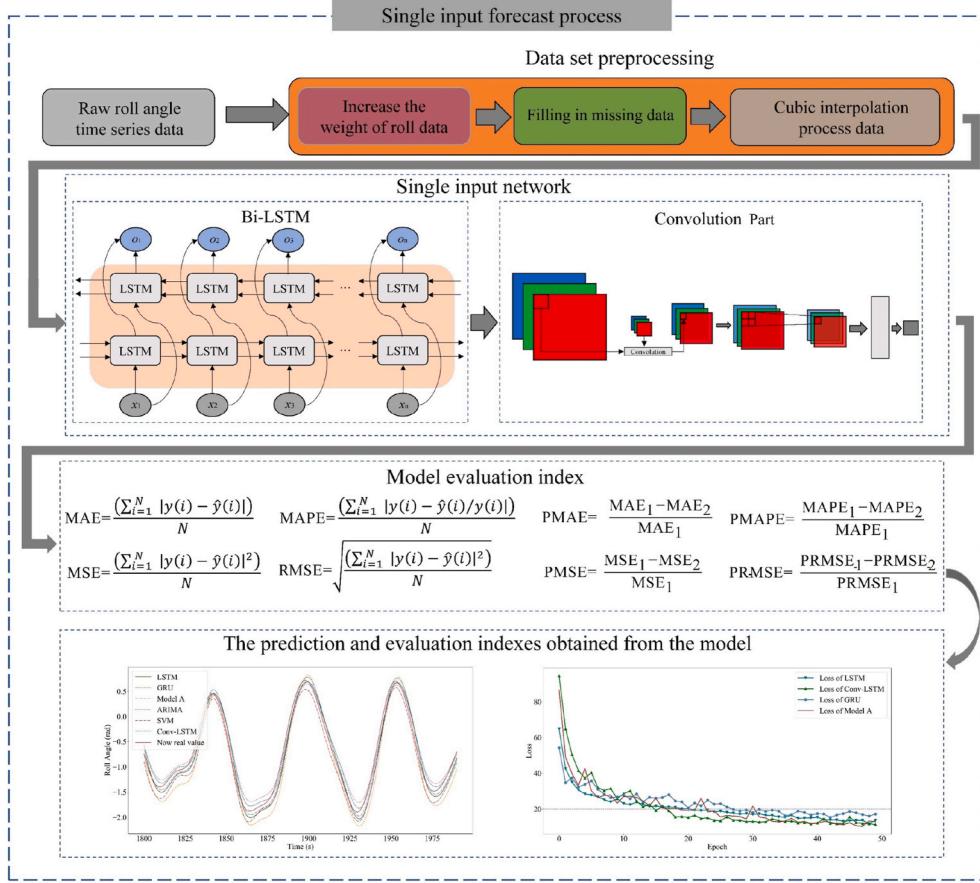


Fig. 9. Single input Bi-LSTM model ship roll angle prediction process.

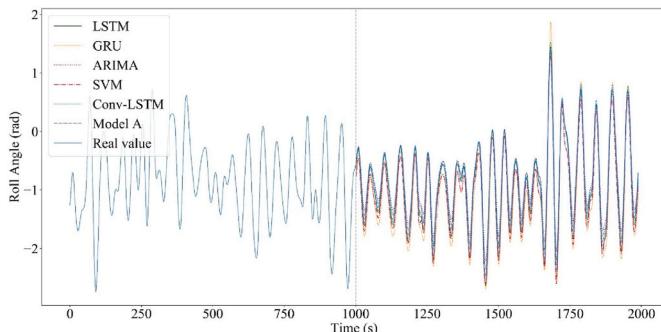


Fig. 10. Ship roll angle prediction results of six models under single input condition.

model. In Fig. 14, the Error curve of GRU is mainly in the range of (0.1, 0.4), and it is generally higher than the Error curves of ARIMA and SVM models, indicating that the prediction performance of the ARIMA and SVM models are better than that of the GRU model. In Fig. 15, the Error curve of LSTM is mainly in the range of (0, 0.2), which shows the overall similarity compared with the Error curve of ARIMA model. The results are consistent with the performance of the MAE in Table 3, the MAE of LSTM and ARIMA models are 0.103 and 0.111 respectively.

In Fig. 16, the Error curve of the Conv-LSTM model is mainly within the range of (0, 0.2), the Error curve of the Conv-LSTM model is similar to the Error curve of the ARIMA and LSTM generally. The MAE of the Conv-LSTM model is 0.088. In Fig. 17, the Error curve of Model A is mainly in the range of (0, 0.15). The Error curve of Model A is generally lower than the Error curve of the other five models as a whole.

According to Table 3, the MAE of Model A is 0.082. By analyzing the prediction results of six models (Figs. 10–11), the error indexes of the six models (Table 3), the promotion percentage indexes of Model A and other five models (Table 4) and the Error curves of the six models (Figs. 12–17), it can be concluded that, the prediction performance of Model A is better than the prediction performance of the other five models.

5.2. Experimental results of ship roll angle prediction under multiple input condition

The Multi-input deep bidirectional feature network is proposed in Section 3.2, hereinafter referred to as Model B. In Model B, the first layer is the Bi-LSTM layer, which is used to extract the deep features of the input data in two directions. The Dropout layer is used to reduce the overfitting, and the parameter is set to 0.2. The features in the two directions are further extracted using a branch network similar to the Bi-LSTM (Model A). The number of neurons is set to 100. The first layer returns the output vector at each sampling time, and the second layer only returns the output vector of the last sampling time. The convolution layer uses a convolution kernel with a format of (9,1) to extract the deep information of the same feature at different sampling times. The deep forward feature and the deep reverse feature are spliced together, and using LSTM to extract stacked deep features. The number of neurons is set to 100, and finally the predicted value is output through two fully connected layers, the number of neuron nodes is 50 and 1, respectively. The detailed settings of Model B are shown in Table 5.

In Table 5, Units represents the number of neurons, Filters represents the number of convolution kernels, return sequence represents whether to return the output of each time, Padding represents the filling method, and Kernel size represents the size of the convolution kernel. Fig. 18

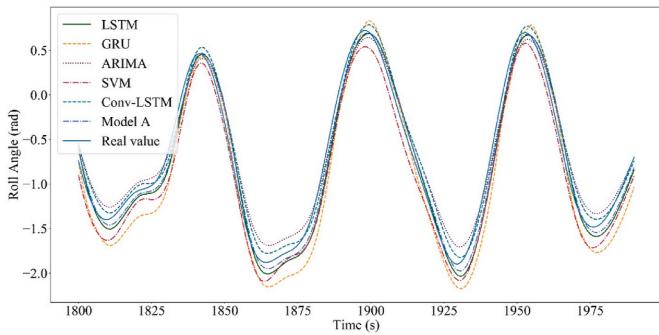


Fig. 11. Local data (1800s–1990s) prediction results of six models for ship roll angle under single input condition. **Fig. 10** shows the prediction results of the ship roll angle within 2000s of the six models under the condition of single input. The blue solid line represents the real data. After 1000s, the prediction results of each model are displayed with different color lines and different types of line. **Fig. 11** shows the prediction results of ship roll angle of the six models in 1800–1990s. When the real roll angle curve is in a uniformly rising or falling area, such as 1880–1895s, the change trend of the prediction curve of the six models is consistent with the real curve. In the area where the real roll angle curve is near the extreme value, the prediction effect is reduced due to the sharp change of the roll angle curve. 1800–1825s is the area near the minimum value of the real roll angle curve, the prediction curves of the six models all show an overall upward trend, the prediction curves of the GRU and SVM do not follow the real data curve in time. In 1800–1825s, the MAE between the real value and the predicted value of the GRU and SVM are 0.26 and 0.19, respectively, and the MAE between the real value and the predicted value of the ARIMA and Conv-LSTM are 0.16 and 0.09 respectively. The MAE between the real value and the predicted value of Model A and LSTM are 0.06 and 0.07, respectively. 1860–1875s is the area near the minimum value of the real roll angle curve, the change trend of the predicted curves of the six models is basically consistent with the change trend of the real roll angle curve. The MAE between the predicted value and the real value of GRU, ARIMA, SVM, LSTM, Conv-LSTM, and Model A in 1860–1875s are 0.24, 0.18, 0.15, 0.13, 0.11, 0.09, respectively. 1880–1910s is the area near the maximum value of the real roll angle curve. The tracking effect of the six models for the real roll angle curve is significantly higher than that of the area near the minimum value. The MAE between the predicted value and the real value of SVM, GRU, ARIMA, Conv-LSTM, LSTM, and Model A in 1880–1910 s is 0.22, 0.19, 0.15, 0.14, 0.06, 0.05, respectively. From the above results, it can be seen that the Model A has the best prediction effect, and the tracking ability near the extremum is the strongest, the prediction error is the smallest, and the prediction effect is the best.

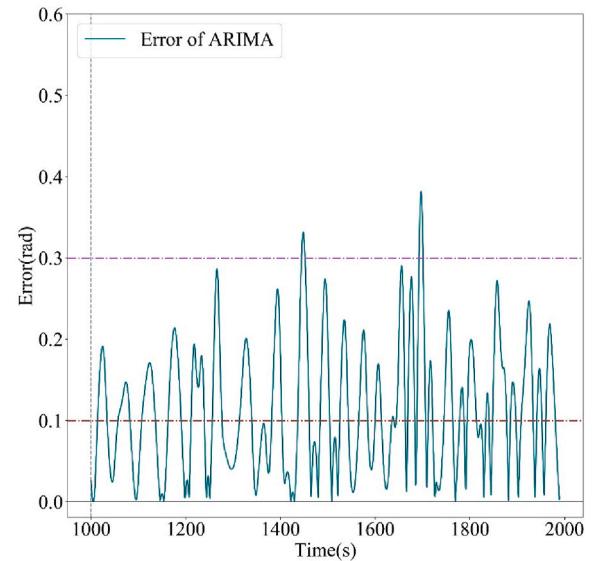


Fig. 12. Error curve of ARIMA model in 1000–2000s.

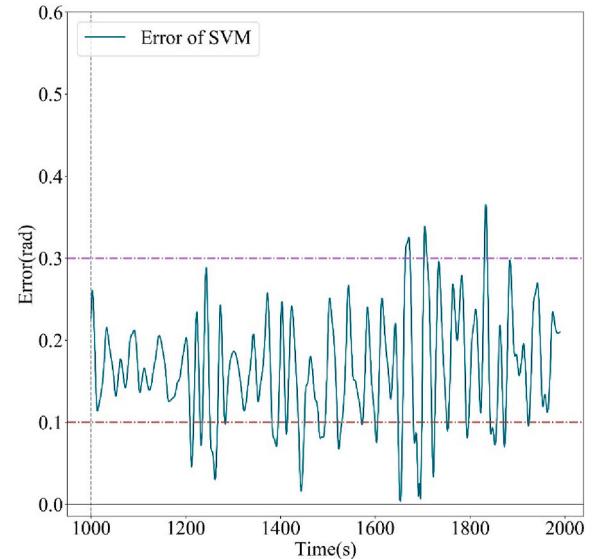


Fig. 13. Error curve of SVM model in 1000–2000s.

weight of the roll angle, interpolation and filling in missing data, etc., and passed to the Bi-LSTM network to generate the forward time features and the reverse time features. Then, the forward and reverse features are extracted by two branch structures similar to Model A, and finally output the predicted value, and evaluate the model through the evaluation indexes.

The simulation experiment uses Model B, GRU, LSTM, and Conv-LSTM to predict the ship roll angle. **Figs. 19–20** shows the prediction results of the four models of GRU, LSTM, Conv-LSTM, and Model B, respectively.

Fig. 19 shows the prediction results of the four models under multi-input condition, showing the change of roll angle in 2000s. The blue solid line represents the real data, the data after 1000s is the predicted data. In order to show the predicted results more clearly, **Fig. 20** shows roll angle prediction of the four models during 1800–1990s. As can be seen from **Fig. 20**, when the real roll angle curve is in a uniformly rising or falling area, such as 1880–1895 s, the curve change trend predicted by the four models is consistent with the real curve. In the area where the real roll angle curve is near the extreme value, the prediction effect is

Table 3
Indexes comparison of six models.

Model	MSE	MAE	MAPE(%)	RMSE
LSTM	0.016	0.103	15.7	0.126
GRU	0.065	0.225	29.8	0.255
CONV-LSTM	0.012	0.088	14.0	0.110
SVM	0.032	0.168	27.9	0.179
ARIMA	0.019	0.111	20.9	0.138
Model A	0.010	0.082	13.0	0.100

Table 4
Promotion percentage of Model A compared with other models.

	LSTM	GRU	Conv-LSTM	SVM	ARIMA
PMAPE	17.19%	56.38%	7.14%	53.40%	37.80%
PMSE	37.50%	84.61%	16.67%	68.75%	47.37%
PMSE	20.39%	63.56%	6.81%	51.19%	26.13%
PRMSE	20.63%	60.78%	9.10%	44.13%	27.53%

shows the complete ship roll angle prediction process of the multi-input deep bidirectional feature network.

In **Fig. 18**, firstly, data preprocessing is performed by increasing the

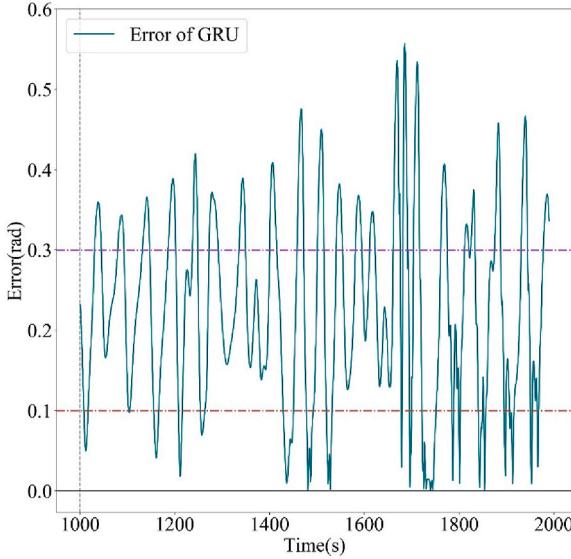


Fig. 14. Error curve of GRU model in 1000–2000s.

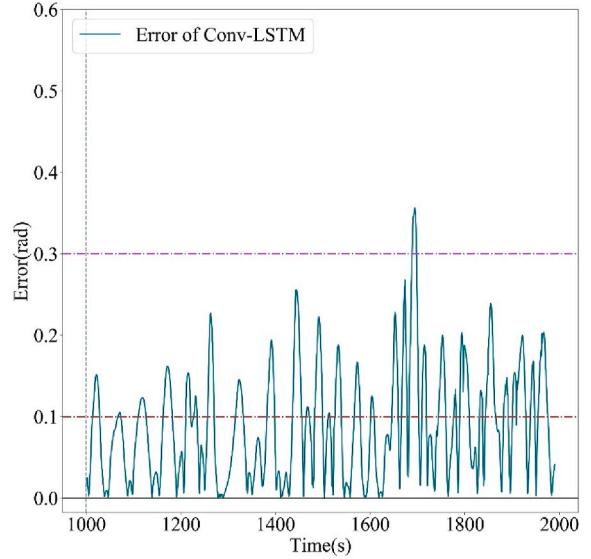


Fig. 16. Error curve of Conv-LSTM model in 1000–2000s.

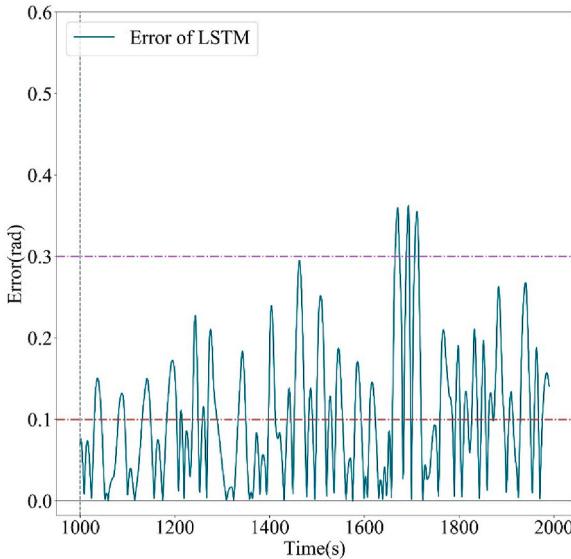


Fig. 15. Error curve of LSTM model in 1000–2000s.

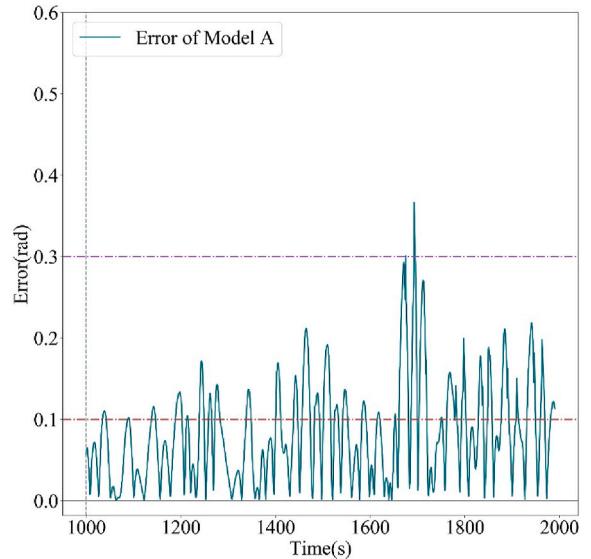


Fig. 17. Error curve of Model A in 1000–2000s.

significantly reduced due to the sharp change of the roll angle curve. The GRU model basically loses its tracking ability near the peak of the real roll angle curve. Near the minimum value, although the change trend of the prediction curve of the Conv-LSTM model is basically consistent with the change trend of the real curve, the prediction error of the Conv-LSTM model near the extreme value is large, indicating that the Conv-LSTM model has good feature extraction ability for single input variable, but poor feature extraction ability for multi-input variables. 1800–1825s is the area near the minimum value of the real roll angle curve, the prediction curves of the four models all show an upward trend, the MAE of Model B and LSTM between the predicted value and the real value in 1800–1825s is 0.065 and 0.071 respectively. The LSTM model has a certain improvement in prediction performance compared with the situation with single input. 1860–1875s is the area near the minimum value of the real roll angle curve. The MAE of Model A and LSTM between the predicted value and the real value in 1860–1875s is 0.069 and 0.073, respectively. 1880–1910s is the area near the maximum value of the real roll angle curve, the MAE of the LSTM, Conv-LSTM, and Model B between the predicted value and the real value in

the 1880–1910s is 0.09, 0.21, and 0.08, respectively. The Model B model has the best prediction effect, and has the strongest tracking ability near the extremum, the best prediction effect, and the smallest prediction error.

In order to more clearly show the decline of MAPE, Fig. 21 shows the decline of the loss function of the four models.

In Fig. 21, the red solid line indicates the decrease of the loss function of Model B. Almost every training period is below the other three curves, indicating that the Model B has better performance in the case of multiple inputs.

Table 6 shows the prediction errors of the four models on the roll angle data under the multi-input situation, including MSE, RMSE, MAPE, MAE, and the bold font indicates the maximum or minimum value.

From the data in Table 6, it can be obtained that single LSTM performs best on the MSE, MAE, and RMSE, which is slightly lower than the related error indexes of the Model B. But on the pivotal index MAPE, the single LSTM model is 13.9%, however Model B is 10.4%. The MAPE of Model B is 25.18% lower than the MAPE of LSTM. The MSE, MAE, and

Table 5
Detailed configuration of Model B.

Layer	Layer parameter setting	Layer	Layer parameter setting
Bi-LSTM Units = 100			
Dropout Rate = 0.2	Dropout Rate = 0.2	LSTM	Units = 100,return sequence = True
LSTM Units = 100, return sequence = True		LSTM	Units = 100,return sequence = False
LSTM Units = 100, return sequence = False		Conv2D	Filters = 1, Kernel size=(9,1),s = 1
Conv2D Filters = 1, Kernel size=(9,1),s = 1			

Concat Layer New shape=(batch, 2, 100).

LSTM Units = 100,return sequence = False.

Dropout Rate = 0.1 Padding = 'same'.

Fully connected Units = 50.

Fully connected Units = 1.

Epoch = 100,Batch size = 64.

RMSE of Model B is 0.0089, 0.0756, and 0.0943, respectively, and MAPE is only 10.4%. The performance of LSTM and GRU in the multi-input case is better than that in the single-input case, however the Conv-LSTM has higher prediction accuracy in the single-input case, indicating that the model has insufficient feature extraction capabilities for multi-input features. Therefore, in terms of the key error index MAPE, Model B has the best prediction effect.

Based on the above analysis, in the case of multiple inputs, the key error index MAPE of Model B is the lowest among all models, the performance of other error indexes is better than most models. Therefore, Model B has the best comprehensive performance of roll angle

prediction under the multi-input situation. In order to conveniently show the differences between different models and Model B, the promotion percentage indexes are used for evaluation, as shown in Table 7.

Table 7 shows the promotion percentage of Model B and other models. The results are as follows: 1) In terms of PMAPE, Model B compares with LSTM, GRU, and Conv-LSTM, the MAPE of Model B is reduced by 25.18%, 59.38%, and 58.73% respectively, and the Model B performs best. 2) In terms of PMSE, the MSE of Model B model is 5.95% higher than that of LSTM model, but it is 74.64% and 85.80% lower than the GRU and Conv-LSTM, respectively. 3) In terms of PMAE, the MAE of Model B is 3.43% higher than that of LSTM, but it is 47.32% and 64.30% lower than the GRU and Conv-LSTM, respectively. 4) In terms of PRMSE, the RMSE of Model B is 2.83% higher than that of LSTM, but it is 49.65% and 62.32% lower than the MSE of GRU and Conv-LSTM models, respectively. The MSE, MAE, and RMSE of Model B are 5.95%, 3.43%, and 2.83% higher than those of LSTM, but the key error index MAPE of Model B is 25.18% lower than that of LSTM. In order to specify the deviation of the predicted value of roll angle from the real value after 1000s, Figs. 22–25 shows the absolute value (Error) of the deviation between the predicted value of roll and the real value of the four models at every sampling time.

In Fig. 22, the Error curve of the GRU model is mainly in the range of (0, 0.3). In Fig. 23, the Error curve of the LSTM is mainly in the range of (0, 0.2). The Error curve of the GRU model is generally higher than the LSTM model. The Error curve shows that the prediction effect of the LSTM model is better than the GRU model. In Fig. 24, the Error curve of the Conv-LSTM model is mainly in the range of (0.1, 0.3). The Error curve of the Conv-LSTM is higher than the LSTM model generally, indicating that the prediction performance of the LSTM model is better than the Conv-LSTM model. In Fig. 25, the Error curve of the Model B is mainly in the range of (0, 0.2). The Error curve of the Model B and the Error curve of the LSTM model are similar in general. The result analysis conforms to the MAE in Table 6. The MAE of the LSTM and Model B are

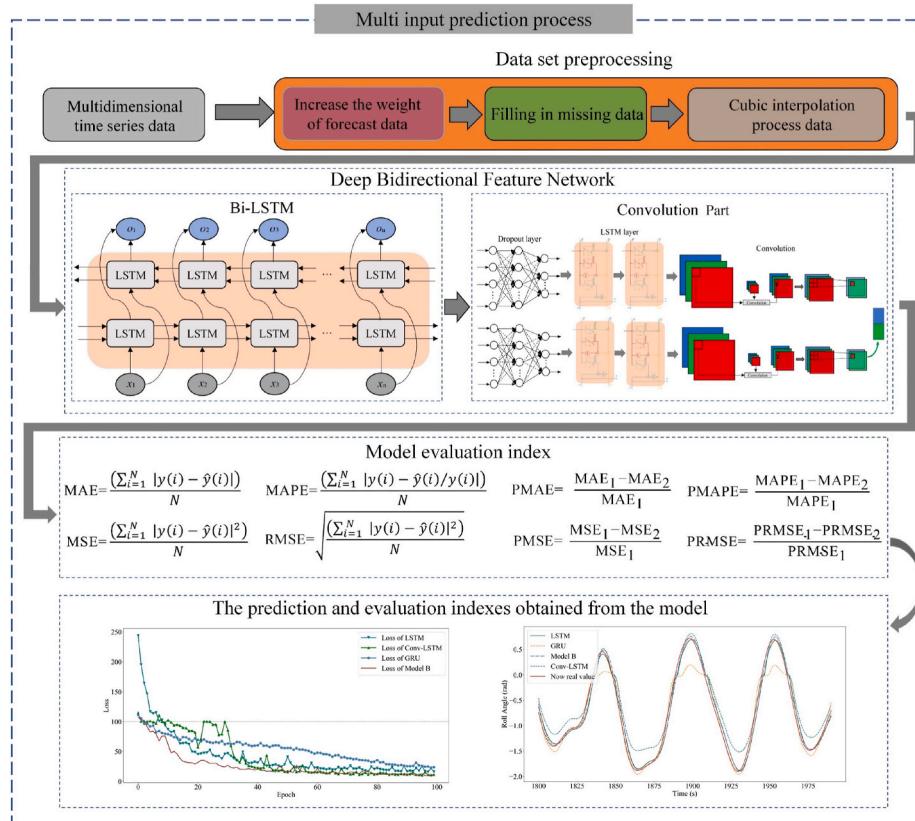


Fig. 18. Prediction process of ship roll angle based on deep bidirectional feature network model under multiple input condition.

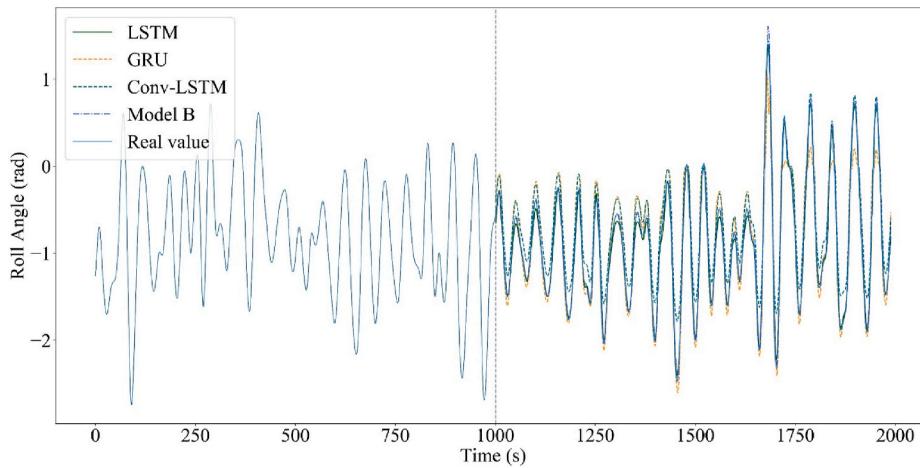


Fig. 19. Ship roll angle prediction results of four models under multi-input condition.

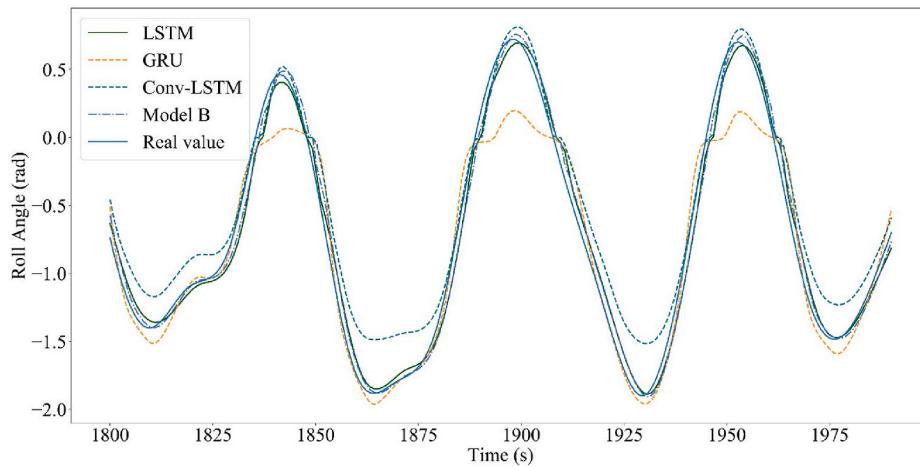


Fig. 20. Ship roll angle prediction results of four models under multi-input condition during 1800–1990s.

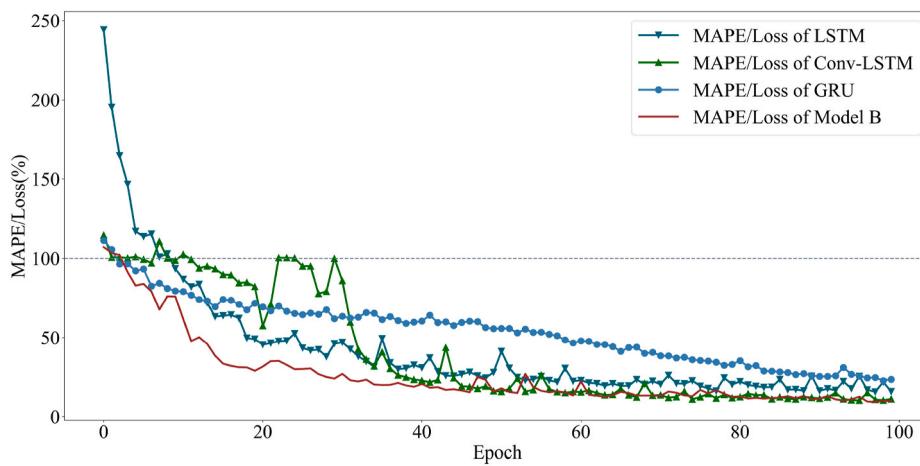


Fig. 21. Under the condition of Epoch = 100, the change of MAPE of the four models.

0.0730 and 0.0756 respectively. By analyzing the prediction results of the four models (Figs. 19–20), the attenuation of the MAPE of the four neural network models (Fig. 21), the error indexes of the four models (Table 6), the promotion percentage indexes of Model B and the other three models (Table 7), and the Error curve of the four models (Figs. 22–25), it can be concluded that the prediction performance of

Model B is better than the other three models.

6. Conclusion

Aiming at the characteristics of nonlinearity and high randomness of the ship's motion when ships sail on the sea, firstly, Bi-LSTM model

Table 6

Comparison of prediction errors under each index of the 4 models.

Model	MSE	MAE	MAPE(%)	RMSE
LSTM	0.0084	0.0730	13.9	0.0917
GRU	0.0351	0.1435	25.6	0.1873
Conv-LSTM	0.0627	0.2118	25.2	0.2503
Model B	0.0089	0.0756	10.4	0.0943

Table 7

Promotion percentage of Model B compared with other models.

	LSTM	GRU	Conv-LSTM
PMAPE	25.18%	59.38%	58.73%
PMSE	-5.95%	74.64%	85.80%
PMAE	-3.43%	47.32%	64.30%
PRMSE	-2.83%	49.65%	62.32%

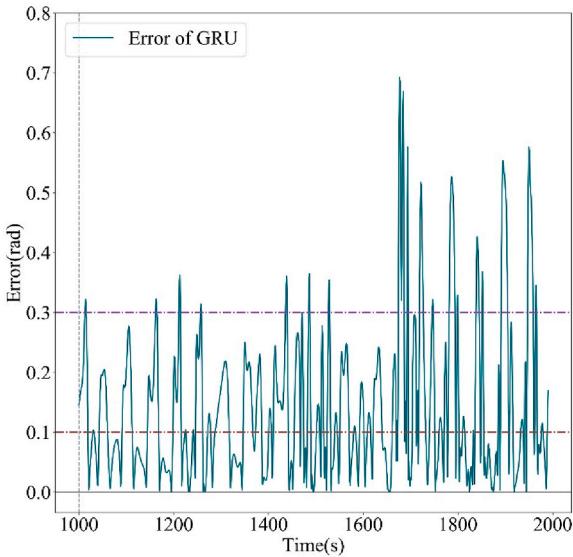


Fig. 22. Error curve of GRU model in 1000–2000s.

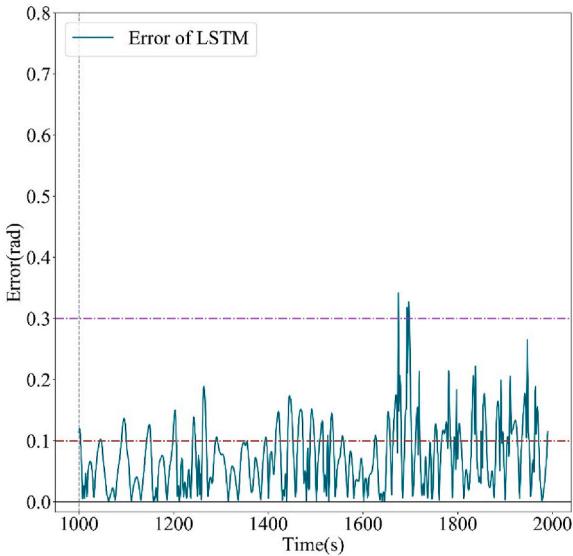


Fig. 23. Error curve of LSTM model in 1000–2000s.

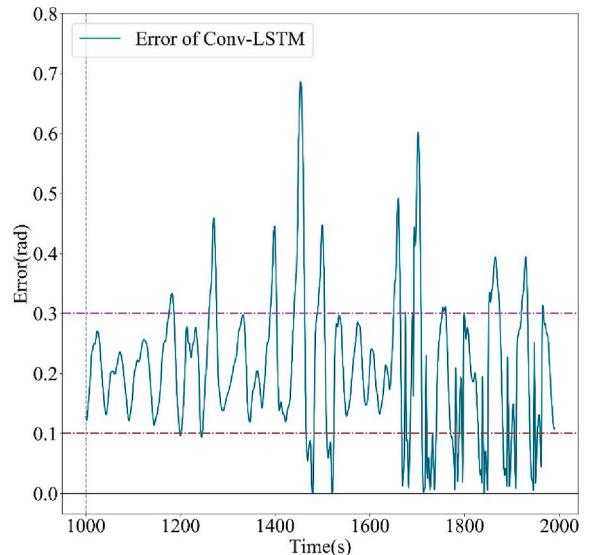


Fig. 24. Error curve of Conv-LSTM model in 1000–2000s.

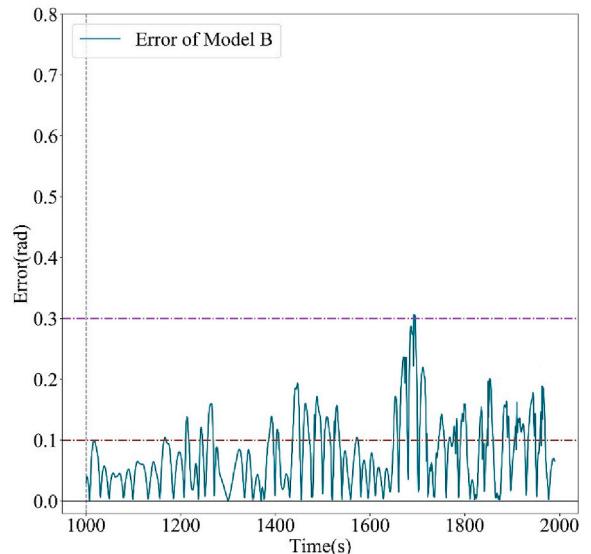


Fig. 25. Error curve of Model B in 1000–2000s.

(Model A) that can extract the temporal pattern is proposed under the condition of single input. Secondly, a deep bidirectional feature network (Model B) in the case of multiple inputs is proposed. Using real ship motion data, several commonly used models in the prediction field are compared with the proposed model under the conditions of single input and multiple inputs respectively, and four error evaluation indexes are used to evaluate the models. Under the single input condition, Model A uses the convolution kernel to extract the deep information contained in the state of single feature at all sampling times, and the single feature comes from the output feature of the Bi-LSTM network. The ship roll angle prediction performance of Model A is better than the single LSTM network. Under multi-input condition, Model B uses Bi-LSTM and two branch networks containing Model A to extract the depth features of forward and reverse information respectively, and compare the prediction results of LSTM, GRU, and Conv-LSTM networks, the prediction performance of Model B has improved significantly. According to the comparison of four error evaluation indexes and the visual display of the prediction results, Model An under the single-input condition and the Model B under the multi-input condition have the best ship roll angle

prediction performance. In addition, from the analysis of the prediction results of the LSTM, GRU and Conv-LSTM models under single-input and multiple-input conditions, it can be seen that simply adding input features does not necessarily improve the prediction performance of the model. Finally, the experimental results with real ship data show that the Bi-LSTM model can obtain the highest prediction accuracy under the single input condition when the data of ship roll angle is insufficient, and the key error evaluation index MAPE is 13.0%. When the data of ship roll is sufficient (obtain the data of roll angular velocity, relative wind speed, relative wind direction, turning angle and rudder angle), the deep bidirectional feature network can obtain a prediction performance of MAPE = 10.4%, which has a higher prediction accuracy than the single input case. This study provides a new solution for ship roll prediction. In order to improve the generalization performance of the model and make the model applicable with different sea conditions, it is required to train the model with a large number of different sea condition data sets, which is not friendly to scientific research and engineering applications. In future work, small sample data will be applied for ship model, or cross domain tasks to improve the generalization performance of the model will be conducted. In addition, the prediction time of the model is also a problem that needs to be solved.

CRediT authorship contribution statement

Yuchao Wang: Conceptualization, Methodology, All authors have read and agreed to the published version of the manuscript. **Hui Wang:** Methodology, Software, Validation, Writing – original draft, Writing – review & editing, All authors have read and agreed to the published version of the manuscript. **Bin Zhou:** Software, Validation. **Huixuan Fu:** Conceptualization, Methodology, Writing – original draft, writing—original draft preparation, Writing – review & editing, All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their valuable comments and constructive suggestions. This research was funded by National Natural Science Foundation of China, grant number 52071112, Fundamental Research Funds for the Central Universities, grant number 3072021CFJ0408.

References

- Alhussein, M., Aurangzeb, K., Haider, S.I., 2020. Hybrid CNN-LSTM model for short-term individual household load forecasting. *IEEE Access* 8, 180544–180557.
- Almonacid, F., Rus, C., Pérez-Higueras, P., Hontoria, L., 2011. Calculation of the energy provided by a PV generator. Comparative study: conventional methods vs. artificial neural networks. *Energy* 36, 375–384. <https://doi.org/10.1016/j.energy.2010.10.028>.
- Barra, S., Carta, S.M., Corriga, A., Podda, A.S., Recupero, D.R., 2020. Deep learning and time series-To-image encoding for financial forecasting. *IEEE/CAA J. Autom. Sin.* 7 (3), 683–692.
- Bu, S.X., Gu, M., Lu, J., Abdel-Maksoud, M., 2019. Effects of radiation and diffraction forces on the prediction of parametric roll. *Ocean. Eng.* 175, 262–272.
- Cai, M., Pipattanasomporn, M., Rahman, S., 2019. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Appl. Energy* 236, 1078–1088.
- Chen, K.J., Chen, K.L., Wang, Q., He, Z., Hu, J., He, J., 2018. Short-term load forecasting with deep residual networks. *IEEE Trans. Big Data* 10 (4), 3943–3952.
- De Lima Silva, P.C., Sadaei, H.J., Ballini, R., Guimaraes, F.G., 2020. Probabilistic forecasting with fuzzy time series. *IEEE Trans. Fuzzy Syst.* 28, 1771–1784.
- Faraji, J., Ketabi, A., Hashemi-Dezaki, H., Shafie-Khah, M., Catalao, J.P.S., 2020. Optimal day-ahead self-scheduling and operation of prosumer microgrids using hybrid machine learning-based weather and load forecasting. *IEEE Access* 8, 157284–157305. <https://doi.org/10.1109/ACCESS.2020.3019562>.
- Gilanifar, M., Wang, H., Sriram, L.M.K., Ozguven, E.E., Arghandeh, R., 2020. Multitask bayesian spatiotemporal Gaussian processes for short-term load forecasting. *IEEE Trans. Ind. Electron.* 67 (6), 5132–5143. <https://doi.org/10.1109/TIE.2019.2928275>.
- Gong, X., Cardenas-Barrera, J.L., Castillo-Guerra, E., Cao, B., Saleh, S.A., Chang, L., 2019. Bottom-up load forecasting with markov-based error reduction method for aggregated domestic electric water heaters. *IEEE Trans. Ind. Appl.* 55 (6), 6401–6413.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016–December 770–778.
- Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554.
- Hochreiter, S., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Hossain, M.S., Mahmood, H., 2020. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* 8, 172524–172533.
- Huang, S.J., Huang, C. Lo, 2000. Control of an inverted pendulum using grey prediction model. *IEEE Trans. Ind. Appl.* 36 (2), 452–458. <https://doi.org/10.1109/28.833761>.
- Jiang, H., Zhang, Y., Muljadi, E., Zhang, J.J., Gao, D.W., 2018. A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. *IEEE Trans. Smart Grid* 9 (4), 3331–3350.
- Jiang, H., Duan, S.L., Huang, L., Han, Y., Yang, H., Ma, Q., 2020. Scale effects in AR model real-time ship motion prediction. *Ocean. Eng.* 203, 107202. <https://doi.org/10.1016/j.oceaneng.2020.107202>.
- Jiménez, J., Pertuz, A., Quintero, C.G., Montaña, J., 2019. Multivariate statistical analysis based methodology for long-term demand forecasting. *IEEE Latin America Transactions* 17 (1), 9.
- Kuremoto, T., Kimura, S., Kobayashi, K., Obayashi, M., 2014. Time series forecasting using a deep belief network with restricted Boltzmann machines. *Neurocomputing* 137, 47–56.
- Lai, G., Chang, W.C., Yang, Y., Liu, H., 2018. Modeling long- and short-term temporal patterns with deep neural networks. *41st Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR* 95–104. <https://doi.org/10.1145/3209978.3210006>, 2018.
- Li, Y., He, Y., Su, Y., Shu, L., 2016. Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines. *Appl. Energy* 180, 392–401. <https://doi.org/10.1016/j.apenergy.2016.07.052>.
- Li, C., Wang, L., Zhang, G., Wang, H., Shang, F., 2017. Functional-type single-input-rule-modules connected neural fuzzy system for wind speed prediction. *IEEE/CAA J. Autom. Sin.* 4 (4), 751–762. <https://doi.org/10.1109/JAS.2017.7510640>.
- Liu, H., Tian, H.Q., Li, Y.F., 2015. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions. *Energy Convers. Manag.* 89, 1–11. <https://doi.org/10.1016/j.enconman.2014.09.060>.
- Ma, X., Dong, Y., 2020. An estimating combination method for interval forecasting of electrical load time series. *Expert Syst. Appl.* 158, 113498. <https://doi.org/10.1016/j.eswa.2020.113498>.
- Ouyang, T., He, Y., Li, H., Sun, Z., Baek, S., 2019. Modeling and forecasting short-term power load with copula model and deep belief network. *IEEE Trans. Emerg. Top. Comput. Intell.* 3 (2), 127–136. <https://doi.org/10.1109/TETCI.2018.2880511>.
- Qin, Y., Li, K., Liang, Z., Lee, B., Zhang, F., Gu, Y., Zhang, L., Wu, F., Rodriguez, D., 2019. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. *Appl. Energy* 236, 262–272.
- Raza, M.Q., Mithulanthan, N., Li, J., Lee, K.Y., 2020. Multivariate ensemble forecast framework for demand prediction of anomalous days. *IEEE Trans. Sustain. Energy* 11 (1), 27–36.
- Sagheer, A., Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* 323, 203–213. <https://doi.org/10.1016/j.neucom.2018.09.082>.
- Sajjad, M., Khan, Z.A., Ullah, A., Hussain, T., Ullah, W., Lee, M.Y., Baik, S.W., 2020. A novel CNN-GRU-Based hybrid approach for short-term residential load forecasting. *IEEE Access* 8, 143759–143768. <https://doi.org/10.1109/ACCESS.2020.3009537>.
- Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45, 2673–2681. <https://doi.org/10.1109/78.650093>.
- Sezer, O.B., Ozbayoglu, A.M., 2018. Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach. *Appl. Soft Comput.* 70, 525–538.
- Sharma, M.K., Dhiman, N., Vandana, Mishra, V.N., 2021. Mediative fuzzy logic mathematical model: a contradictory management prediction in COVID-19 pandemic. *Appl. Soft Comput.* 105, 107285.
- Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 2015–January 802–810.
- Sun, Z., Zhao, M., 2020. Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis. *IEEE Access* 8, 134422–134434.
- Sun, Z., Zhao, S., Zhang, J., 2019. Short-term wind power forecasting on multiple scales using VMD decomposition, K-means clustering and LSTM principal computing. *IEEE Access* 7, 166917–166929.
- Sun, N., Zhang, S., Peng, T., Zhou, J., Sun, X., 2020. A composite uncertainty forecasting model for unstable time series: application of wind speed and streamflow forecasting. *IEEE Access* 8, 209251–209266. <https://doi.org/10.1109/ACCESS.2020.3034127>.

- Tan, M., Yuan, S., Li, S., Su, Y., Li, H., He, F.H., 2020. Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning. *IEEE Trans. Power Syst.* 35 (4), 2937–2948. <https://doi.org/10.1109/TPWRS.2019.2963109>.
- Wang, Z., Hong, T., Piette, M.A., 2019a. Predicting plug loads with occupant count data through a deep learning approach. *Energy* 181, 29–42. <https://doi.org/10.1016/j.energy.2019.05.138>.
- Wang, K., Qi, X., Liu, H., 2019b. Photovoltaic power forecasting based LSTM-Convolutional Network. *Energy* 189, 116–225.
- Wang, K., Qi, X., Liu, H., 2019c. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl. Energy* 251, 113315.
- Wang, Y., Wang, H., Zou, D., Fu, H., 2021. Ship roll prediction algorithm based on Bi-LSTM-TPA combined model. *J. Mar. Sci. Eng.* 9, 387.
- Wu, D., Wang, B., Precup, D., Boulet, B., 2020. Multiple kernel learning-based transfer regression for electric load forecasting. *IEEE Trans. Smart Grid* 11 (2), 1183–1192.
- Yin, J.C., Perakis, A.N., Wang, N., 2018. A real-time ship roll motion prediction using wavelet transform and variable RBF network. *Ocean. Eng.* 160, 10–19.
- Yu, L., Ma, N., Wang, S., 2019. Parametric roll prediction of the KCS containership in head waves with emphasis on the roll damping and nonlinear restoring moment. *Ocean. Eng.* 188, 106298. <https://doi.org/10.1016/j.oceaneng.2019.106298>.
- Yuan, K., Liu, J., Yang, S., Wu, K., Shen, F., 2020. Time series forecasting based on kernel mapping and high-order fuzzy cognitive maps. *Knowl. Base Syst.* 206, 106359.
- Zhang, P.C., 2018. Short-term rainfall forecasting using multi-layer perceptron. *IEEE Trans. Big Data* 6 (1), 93–106.