LSTM-PSO: Long Short-Term Memory Ship Motion Prediction Based on Particle Swarm Optimization

Yuxin Yao[1], Liang Han[2], Jiangyun Wang[3]

1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
E-mail: yyxdawang78@buaa.edu.cn

2. E-mail: liang@dept3.buaa.edu.cn

3. E-mail: wangjiangyun@buaa.edu.cn

Abstract— **According to the nonlinear characteristic of ship motion, the ship motion pose will be disturbed by coupling, indefinite period, noise signals, chaotic and some other factors, which leads that it is hard to predict ship motion in the future precisely. Based on the above, and considering the sequence of ship movement, many neural networks have been applied in ship motion prediction, such as LSTM (Long Short-Term Memory, LSTM) and ESN (Echo State Network, ESN). However, there are problems in the parameter setting of ANN (Artificial Neural Network) algorithm, that how to update network parameters during training iterations of the network to avoid iterates getting into local optimum. LSTM with PSO optimization is proposed in this paper. Testing simulation results show that the combination LSTM and PSO improves the accuracy of ship motion prediction.**

I. Introduction

Ship motion prediction has been widely used in optimization of ship control, navigation control under large wave, danger alarm, assessment of manipulation danger, and assist of weapon lunch. Moreover, ship motion prediction can help to improve the accuracy of ship manipulation, so to reduce the chance of accident. Therefore, the research of ship motion prediction has great civilian and military value [1,2].

At present, ship model theory is still widely used as a common method to predict ship motion. Ship is divided into number of rigid modules, and described in the form of differential equations in the ship motion theory. There are certain shortages, for instance that many parameters of ship and wave are hard to get, and this method lacks generality because of different wave conditions requiring different models. According to ship motion's stochastic process, the regression method can be considered to be used to model ship motion. Linear and nonlinear can predict ship movement simply, but the accuracy and time of forecast are limited. Time series, one of regression method, is considered to forecast ship motion usually. AR (autoregressive model, AR) and ARMA (autoregressive and moving average model, ARMA) are used to construct time series, which have been applied in many fields[7]. But there exists certain problem of applying AR and ARMA in ship motion prediction, because AR and ARMA are linear methods and ship motion is nonlinear, which leads to the great decreasing of prediction accuracy.

LSTM was presented by Sepp Hochreiter and Jürgen Schmidhuber in 1997, one more mature RNN (Recurrent Neural Network, RNN)[5]. RNN is known for its ability of handling time series data, which can connect the output data of the previous time node to the next time node, so that the extraction of data features is more accurate and the prediction accuracy is greatly improved[6,11]. However, RNN encounters great difficulties when dealing with long-term dependencies data, because multiple multiplications of the Jacobian matrix during the calculation of connection of nodes far apart with each other leads more gradient disappearance or less gradient swelling[8]. To solve this problem, researchers proposed many solutions, such as ESN, the addition of Leaky Units, and so on[3]. Gate RNN is the successful and wide applications in these solutions, and LSTM is the most famous and successful applications of gate RNN[9].

So in this paper, LSTM is applied to predict short ship motion[4]. The setting of network parameters greatly determines the ultimate effect of network. Therefore, optimizer is used to optimize parameters of network during training of neural network, including SGD, Adagrad, Momentum, Adam, RMSProp, etc. These optimizers are based on gradient descent of object function on parameters. However, there exists a problem in all above optimizers, that the iterative process may fall into a local optimum and cannot achieve the global optimum, which will affect the effect of model.

Therefore, it is proposed that using PSO to optimize LSTM training process in this paper[10]. Since the initial position of the particles in PSO algorithm are randomly generated, the probability of the parameters entering a local optimum is reduced. Meanwhile, the reduction of use of gradient makes training process easier, since the PSO does not need to use gradient information.

II. LONG SHORT-TERM MEMORY

The LSTM model are consist of a set of interconnected memory modules. Each network module contains one or more auto-correlated memory cells and three proliferation units: input gates, output gates, and forget gates. Forget gates determine how much of c_{t-1}, the cell state at the previous time node, is reserved to the current moment t. Input gates determine how much of x_t, the network's input is reserved to c_t, the cell state at current time node. The architecture of LSTM is as shown in Fig.1.

The whole basic equation of LSTM can be described as:

1. For input gate:

$$i_t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i) \tag{1}$$

2. For forget gate:

$$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_i) \tag{2}$$

3. For cell:

$$C_t^*=\tanh(W_c\cdot[h_{t-1},x_t]+b_c) \tag{3}$$

$$C_t=f_t*C_{t-1}+i_t*C_t^* \tag{4}$$

4. For output gate:

$$o_t=\sigma(W_o\cdot[h_{t-1},x_t]+b_o) \tag{5}$$

$$h_t=o_t*\tanh(C_t) \tag{6}$$

Where i_t represents the output of input gate, f_t represents the output of forget gate, C_t represents the output of the cell of LSTM, and h_t represents the real output of the hidden layer of LSTM. The fundamental reason why LSTM can avoid the gradient vanishing is the sum of the output of input gate and forget gate in the cell.

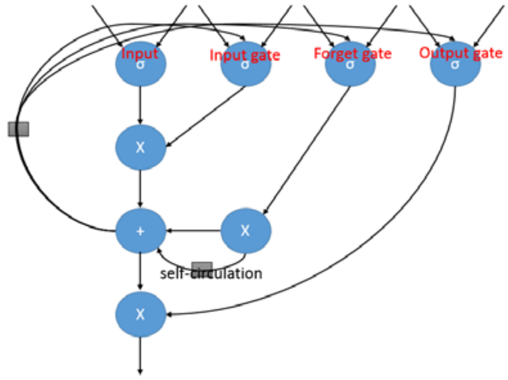


Fig.1. the Architecture of LSTM

III. PARTICLE SWARM OPTIMIZATION

PSO (Particle swarm optimization, PSO) was proposed by Kennedy and Eberhart in 1995, one algorithm of particles flying in the solution space and landing at the best solution, improved on basis of the Hepper's model of simulating birds swarm or fish swarm. PSO is simple and easy to implement, without any gradient information and less parameters. It can be said that PSO is not only suitable to scientific research, but also engineering application.

In PSO, every solution is taken as a particle, and judge the current position by its own current fitness function during every iteration. Besides, the velocity of particle decides its moving distance and direction every step. The architecture of PSO is as presented in Fig.2.

In every iteration, the velocity and position of particle P_i, $i=\{1,2,\ldots,N\}$, N representing the total number of particles, updates according to the following principles:

$$V_{id}^{k+1}=wV_{id}^{k}+c_1r_1\left(p_{id}^{k}-X_{id}^{k}\right)+c_2r_2(p_{gd}^{k}-X_{id}^{k}) \tag{7}$$

$$X_{id}^{k+1}=X_{id}^{k}+V_{id}^{k+1} \tag{8}$$

Where V_{id}^k and X_{id}^k represent respectively the d_{th} component of the k_{th} velocity and position vector of the i_{th} particle, c_1 , c_2 represents respectively the learning coefficient of particle P_i in local best position and global best position, r_1, r_2 are two random number between 0 and 1 to increase the randomness of search, w represents inertia weight in order to adjust the ability to search in solution space.

The reason why choose PSO as an optimization is at its fast convergence and global search optimality.

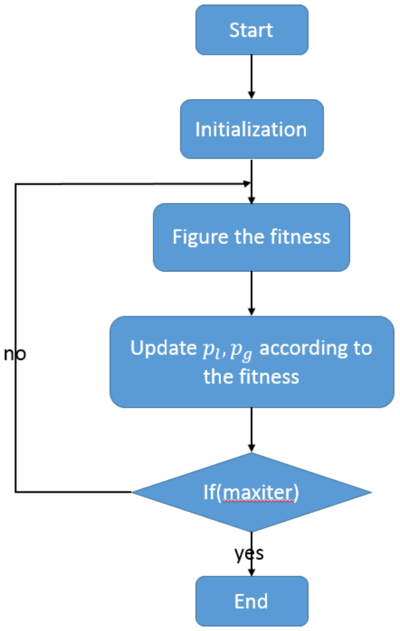


Fig.2. the Architecture of PSO

IV. HYBRID LSTM WITH PSO OPTIMIZATION

In artificial neural network, parameters of the network greatly influence final effect of model in real world task. Until now, the whole of common optimizers for parameter are based on the gradient of loss function on parameters, including Adam, Adagrad, SGD, etc. The main idea of these methods is to use the gradient to find the optimal direction of parameter, based on gradient descent algorithms. This method cannot guarantee the solution obtained in global solution space, although it realizes fast convergence of the model and adaptive adjustment of the learning rate. Therefore, using PSO to optimize LSTM is proposed in this paper, called "LSTM-PSO", which increases the probability of searching the optimal solution in global solution space[12, 13]. The architecture of LSTM-PSO is shown as Fig.3.

Where p_{best} is the best particle position optimized by PSO, and $maxiter$ is the PSO training iteration number. The whole training process is consist of two iterations, the outer training iteration of LSTM and the training iteration of PSO. In every iteration of neural training, the update of parameter depends on the gradient of the target function on the parameter no longer, but according to the best position gotten from PSO.

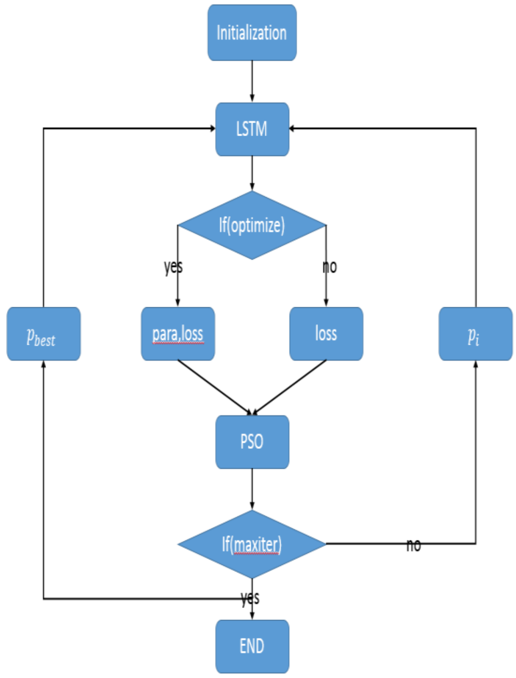


Fig.3. the Architecture of LSTM-PSO

The initialization of LSTM parameter is no longer decided by random, but optimized by PSO. During training, the loss function L_t^i, $i=1,2,\ldots,maxiter$, and parameter p_t^j, $j=1,2,\ldots,n$, where $maxiter$ is number of training, n is the number of parameter to be optimized every layer and t is current time node, are fed into PSO, and then the parameter group corresponding to the minimum loss, after finishing whole searching iteration of PSO, is returned to LSTM to continue the following training[14].

V. EXPERIMENTS

The ship's heaving and pitching rotation data is used as training, validation and test data in this paper, and the sampling period is given 0.016 second, which means the distance between each time node is 0.01s. What's more, the future 5s (300 time steps in the future) is set as the forecast time, which means the model will give the predict result of 5s in the future once.

In this paper, MSE is used as loss function to training model and test effect. MSE can be described as (9):

$$MSE=\frac{1}{p}\sum_{i=1}^{p}(y_t^i-y^i)^2 \quad (9)$$

Where p is the predict distance, t is the t_{th} time node, y_t^i is the output of model in t_{th}, y^i is the real value in t_{th}. In the experiments, to avoid the non-convergence of model, the model is trained with PSO based on the model optimized by Adam. In order to test the final forecast effect of LSTM-PSO model, other five optimizers have been used to train basic LSTM network and forecast the ship motion in the future 4s in this paper, including Adam, Adagrad, SGD, RMSprop, AdaDelta. As shown in Fig.4-9. In order to ensure the fairness of the experiment, all comparative experiments are based on the model trained by the Adam optimizer.

In the experiment, in order to guarantee the convergence of model, the combination of PSO and Adam is used as the optimizer to optimize model during training. The random parameters are optimized by PSO, then updated by Adam. Considering the train time and compute resource, the train and test are run at the small dataset, the ratio of train set, validation set, and test set is 2:2:1, and the train iterations are set at 1000.

Figure 4. Predict result of ship motion in the future 5s with Adam

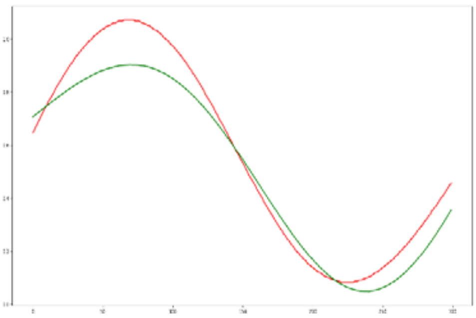


Figure 5. Predict result of ship motion in the future 5s with Adagrad

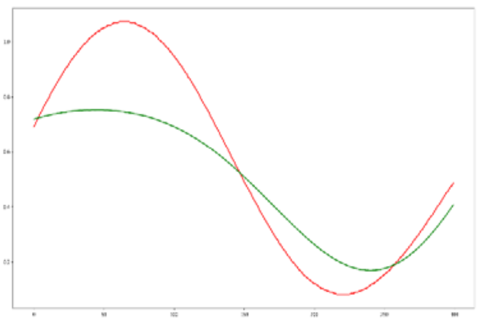


Figure 6. Predict result of ship motion in the future 5s with SGD

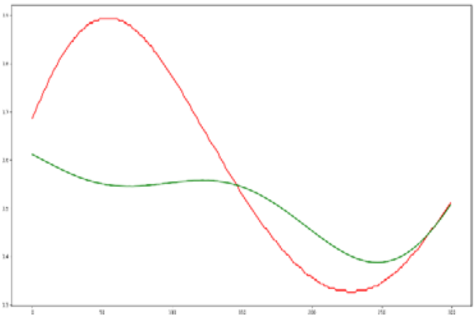


Figure 7. Predict result of ship motion in the future 5s with RMSprop

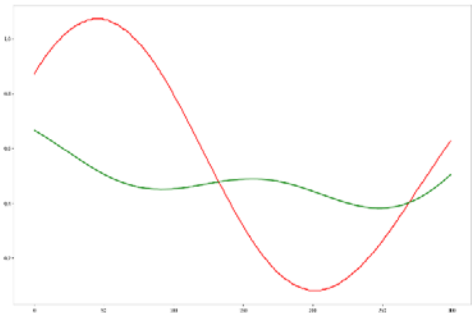


Figure 8. Predict result of ship motion in the future 5s with Adadetla

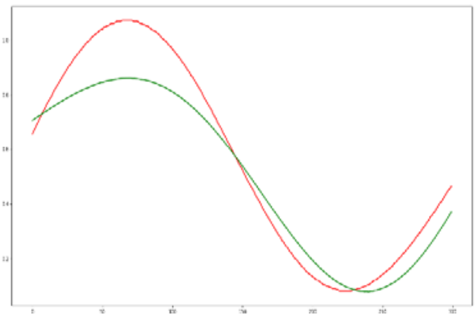
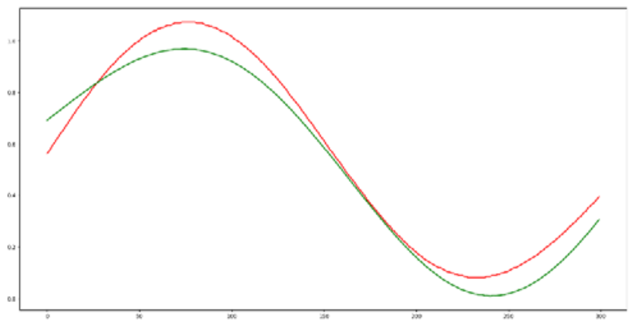


Figure 9. Predict result of ship motion in the future 5s with PSO-Adam



The accuracies and losses of the same model with five optimizers and PSO above on train set and validation set at 40th train step are as presented in Table 1.

TABLE 1. THE ACCURACIES AND LOSSES OF SIX OPTIMIZERS

Optimizer	Acc	loss
Adam	0.7847	7.4847e-05
Adagrad	0.4472	1.5334e-04
SGD	0.1058	0.0013
RMSprop	0.2518	1.7062e-04
Adadetla	0.6196	9.5228e-05
PSO-Adam	0.8448	8.0681e-04

VI. CONCLUSION

In this paper, LSTM model is combined with PSO-Adam optimization to improve the forecast accuracy in ship movement. Adam Added with PSO to optimize LSTM can increase the probability of searching the best solution in global solution space. However, a lot of noise exist in the original data, reducing the prediction ability of model. Therefore, this combined model will be improved, with joining Kalman filter, to reduce the noise in the original data. What's more, the initial particles of PSO are based on Adam to guarantee the convergence in this paper.

REFERENCES

[1] Jie X, Chaozhong W, Zhijun C, et al. A novel estimation algorithm for interpolating ship motion[C]//Transportation Information and Safety (ICTIS), 2017 4th International Conference on. IEEE, 2017: 557-562.

[2] Borkowski P. The ship movement trajectory prediction algorithm using navigational data fusion [J]. Sensors, 2017, 17(6): 1432.

[3] Zhao X, Xu R, Kwan C. Ship-motion prediction: algorithms and simulation results[C]//Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on. IEEE, 2004, 5: V-125.

[4] Cheng X, Chen S, Diao C, et al. Simplifying Neural Network Based Model for Ship Motion Prediction: A Comparative Study of Sensitivity Analysis[C]//ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering. American Society of Mechanical Engineers, 2017: V001T01A016-V001T01A016.

[5] Schmidhuber J. Deep learning in neural networks: An overview [J]. Neural networks, 2015, 61: 85-117.

[6] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 2014: 3104-3112.

[7] Tian Y, Zhu Y. Better computer go player with neural network and long-term prediction [J]. arXiv preprint arXiv:1511.06410, 2015.

[8] Peng X, Dong H, Zhang B. Echo State Network ship motion modeling prediction based on Kalman filter[C]//Mechatronics and Automation (ICMA), 2017 IEEE International Conference on. IEEE, 2017: 95-100.

[9] A. Graves. Supervised Sequence Labelling with Recurrent Neural Networks. Textbook, Studies in Computational Intelligence, Springer, 2012.

[10] Li X, Lv X, Yu J, et al. Neural Network Application on Ship Motion Prediction[C]//Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2017 9th International Conference on. IEEE, 2017, 1: 414-417.

[11] Chouikhi N, Ammar B, Rokbani N, et al. PSO-based analysis of Echo State Network parameters for time series forecasting [J]. Applied Soft Computing, 2017, 55: 211-225.

[12] Wang H, Yan X. Optimizing the echo state network with a binary particle swarm optimization algorithm [J]. Knowledge-Based Systems, 2015, 86: 182-193.

[13] Ye Y, Lu Z L. Neural network short-term traffic flow forecasting model based on particle swarm optimization [J]. Computer Engineering and Design, 2009, 30(18): 4296-4298.

[14] Chouikhi N, Ammar B, Rokbani N, et al. PSO-based analysis of Echo State Network parameters for time series forecasting [J]. Applied Soft Computing, 2017, 55: 211-225.