IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 24, 2020, accepted June 25, 2020, date of publication July 2, 2020, date of current version July 15, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.3006499

Analysis and Application of Grey Wolf Optimizer-Long Short-Term Memory

JINXIN PAN, BO JING, XIAOXUAN JIAO, AND SHENGLONG WANG
Prognostics and Health Management Laboratory, Air Force Engineering University, Xi'an 710038, China
Corresponding author: Xiaoxuan Jiao (1091313541@qq.com)

This work was supported in part by the Aeronautical Science Foundation of China under Grant 20142896022.

ABSTRACT Long short-term memory (LSTM) is widely applied in both academic and industrial fields. However, there is no reliable criterion on selecting hyperparameters of LSTM. Currently, although some widely used classic methods such as random search and grid search have obtained success to some extent, the problems in local optimum and convergence still exist. In this research, we propose to use grey wolf optimizer (GWO) to search for the hyperparameters of LSTM. Through the method, the superiority of metaheuristic in global optimization and the strength of LSTM in predicting are combined. In this model, number of hidden layer nodes and learning rate of LSTM are set as preys, and grey wolf pack has a simple but efficient mechanism to search for the optimal hyperparameters. The benchmark tests on several basic functions were utilized, and the results were verified by a comparative study with random search, support vector regression and several other regression methods. Specifically, we applied this algorithm in predicting the degradation trend of the airborne fuel pump. As a result, the ergodicity and convergence of the algorithm are proved mathematically based on Markov processes theory. The benchmark tests show that the GWO-LSTM model holds for predicting data with low overall slope and high partial fluctuation. The application in airborne fuel pump shows that, trained by dataset with 5700 points, the proposed model could predict sequence of 300 points with root mean square error 0.617 after 30 iterations of optimizing, which is 2.512 previously. The result further demonstrates that the proposed algorithm is applicable to make prediction with high accuracy. Overall, the effectiveness of GWO-LSTM model is verified from theoretical proof to benchmark tests and then to actual product application.

INDEX TERMS Evolutionary computation, LSTM, Markov processes, prediction algorithms, airborne fuel pump.

I. INTRODUCTION

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction [1]. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics [1]–[4]. Long short-term memory (LSTM) constitute a very powerful class of computational models, capable of instantiating almost arbitrary dynamics [5]. In recent years, LSTM has become a widely used model for a variety of problems [6].

Performance of many deep learning algorithms depend critically on hyperparameters [7], [8]. However, due to the lack of precise mathematical relationship between hyperparameters and results, there is no definite criterion for hyperparameters selection. In practice, the most popular methods used to search hyperparameters are Bayesian optimization [9], grid search [10] and random search [11]. Nevertheless, such methods have some limitations. For example, the computational expense of grid search grows dramatically with the number of hyperparameters in the model, and the random search method has a risk of falling into local optimum [8], [12]–[14]. The problem of determining hyperparameters has become a bottleneck restricting the accuracy of deep learning.

Metaheuristic optimization techniques have become very popular over the last two decades. Surprisingly, some of them such as Artificial Bee Colony (ABC) [15], Ant Colony Optimization (ACO) [16], and Particle Swarm Optimization (PSO) [17] are fairly well-known among not only

The associate editor coordinating the review of this manuscript and approving it for publication was Soon Xin Ng.

computer scientists but also scientists from different fields. Regardless of the differences between the metaheuristics, most of metaheuristics are nature-inspired (inspired from some principles in physics, biology, etc.), and metaheuristic optimization techniques are proved to have abilities to avoid local optima and outstanding superiority in ergodicity and convergence [18]–[21]. Metaheuristics has attracted more and more attention in hyperparameter setting.

Aiming at the global optimization capacity of metaheuristics, in this paper, we propose to adopts the grey wolf optimizer (GWO) to search for the hyperparameters of the LSTM. To verify the effectiveness of the algorithm mathematically, we prove the ergodicity and convergence of the algorithm based on Markov processes theory, the result shows that the sequence of grey wolf pack converges to the optimal hyperparameters of LSTM with probability 1. The algorithm is then benchmarked on several basic functions, the results show that the GWO-LSTM algorithm is able to provide very competitive results for predicting data with low overall slope and high partial fluctuation. Finally, we consider predicting the pressure degradation trend of the airborne fuel pump based on GWO-LSTM algorithm. The results show that the root mean square error is only 0.617 after 30 iterations of optimizing, which is 2.512 before optimizing.

The major contributions can be concluded as following. Firstly, we presented the combination of GWO and LSTM. The results of benchmark tests and application in airborne fuel pump show that, the proposed model could provide very competitive accuracy compared to other algorithms. Secondly, the route to test the proposed model is worth mentioning. We completed the test of GWO-LSTM from theory prove to benchmark tests and then to application in real products, which increasing the credibility of the model. Thirdly, the predicting of degradation data from airborne fuel pump is helpful to assess the status of the products. Through prediction, the lifespan of the pump could be estimated, so that we can change the pump before it breaks. Such idea is also in line with PHM (prognostics and health management).

II. GWO-LSTM ALGORITHM ANALYSIS

A. GREY WOLF OPTIMIZER

Wolf is a kind of fierce and wise animal. In the natural world, the grey wolf pack has created a set of efficient hunting skills in response to the harsh natural environment. Humans have imitated the wolf pack activity to solve practical problems. There are precedents in history. Genghis Khan formatted the army with the wolf warfare method, the German submarines used the wolf pack attack tactics during the World War II, and the US electronic warfare "wolf pack attack system" has all embodied the wisdom of the wolf canine [21].

There is a strict hierarchy in the grey wolf pack. The grey wolf pack can be divided into four levels as shown in the Figure 1. Among them, the α wolf is the head wolf, which is mainly responsible for the decision of the entire wolf pack's habitat, hunting, and moving behaviors. The α wolf is the core

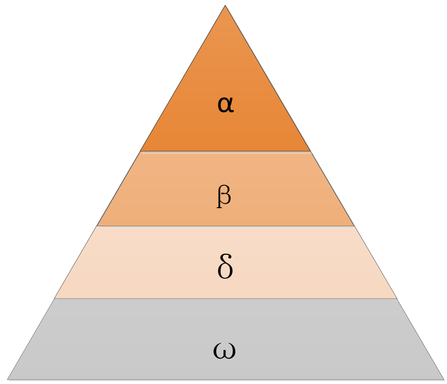


FIGURE 1. The hierarchy of grey wolf pack. Dominance decreases from top down.

of grey wolf pack. The β wolf is the second in command of the grey wolf pack, who subordinate to the α wolf and assists to make decisions. And β wolf is the first candidate for the head wolf after the α wolf dies. The δ wolf consists of sentry wolves, young wolves, care wolves, etc., that obeys the α and β wolves, and assists in managing the grey wolf pack. The ω wolf is composed of other members of the grey wolf pack, it is the lowest layer of the grey wolf pack, that obeys the management of the upper layer, and faces some risks in survival.

The grey wolf optimizer was proposed by Australian scholar Mirjalili in 2014, and the core of the algorithm is to simulate the division and cooperation in the hunting process of the grey wolf pack. The algorithm considers the optimal solution as a prey, and adopt the grey wolf pack to continuously approach the prey. It can be described as follows:

1. Structure division of the grey wolf pack. Set the hunting space of the grey wolf pack to an N × M Euclidean space, where N is the number of grey wolves, and M is the dimension of the prey. The position of each wolf could be expressed as $X_i = (x_1, x_2, \ldots, x_M)$ (i = 1,2, ..., N). Then, evaluate the distance between each grey wolf and the prey according to the hyperparameters of the position. The three wolves in the best positions are selected as α wolf, β wolf, and δ wolf, and the remaining wolves are ω wolf.

2. Search for the prey. The search activities of the grey wolf pack were completed under the guidance of α, β, and δ. Under the call of the three head wolves, they continuously search for prey. The mathematical model of searching is referred to equations (1)-(9).

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}(t)\right| \tag{1}$$

$$\vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}(t)\right| \tag{2}$$

$$\vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}(t)\right| \tag{3}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \tag{4}$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \tag{5}$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \tag{6}$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{7}$$

$$\vec{C} = 2\vec{r}_2 \tag{8}$$
$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{9}$$

where vector X(t) denotes the current position of the moving wolf, vector X(t+1) denotes the position after searching. Vector X_α, X_β and X_δ is the position of α wolf, β wolf, and δ wolf. Vector r_1, r_2 take random values between (0, 1) during each searching iteration. Components of vector a decrease linearly from 2 to 0 over the course of iterations. '·' denotes the Hadamard product [21].

After each search process, the structure of the grey wolf pack is redecided. The three wolves closest to the prey are automatically converted into α wolf, β wolf, and δ wolf, and the next round of search is organized.

When the times of search iterations is still small, the value of a is relatively large. Due to the precise position of the prey is not known yet, the grey wolf pack tends to expand the search range. A part of grey wolf pack is shrinking in the direction of α wolf, β wolf, and δ wolf, and other part of grey wolf pack move in the opposite direction. For optimization, this search mechanism could decrease the risk of falling into local optimal values.

As the times of search iterations continues to increase, the magnitude of vector A continues to decrease. At that time, the grey wolf pack has basically grasped the position of the prey, the grey wolf pack gradually tends to shrink, and starts to surround and storm.

The random change of vector r_1 and r_2 bring uncertain factors to the search of the grey wolf pack, which also conforms to the distortion in the information transmission of the grey wolf pack in the natural environment and the randomness of the grey wolf moving. As for the algorithm, it can also decrease the risk of falling into the local optimal value.

B. LSTM

In principle, recurrent networks can use their feedback connections to store representations of recent input events in the form of activations (short-term) memory. However, error signals flowing backward in time tend to blow up or vanish. Based on this, Jürgen Schmid Huber proposed LSTM. The key to LSTMs is the cell state, which works as a conveyor belt. The cell state runs straight down the entire chain, with only some minor linear interactions. Regulated by structures called gates, LSTM does have the ability to remove or add information to the cell state. Thus, LSTM was explicitly designed to avoid the long-term dependency problem. [22].

As shown in Figure 2 and 3, each cell of the LSTM contains three gates, that are the forget gate, the input gate, and the output gate. The x(t) and h(t-1) are the inputs, and the computation of each cell can be defined by a series of equations as (10)-(15):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{10}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{11}$$
$$g_t = \tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \tag{12}$$

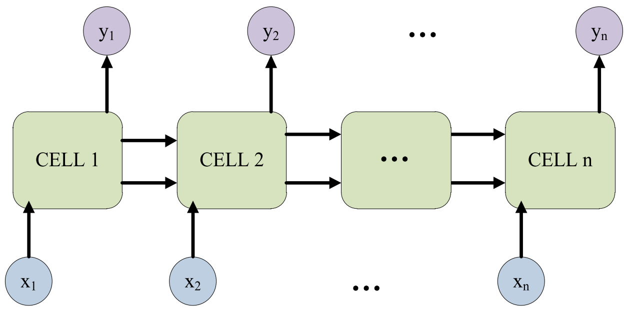


FIGURE 2. Schematic diagram of LSTM network.

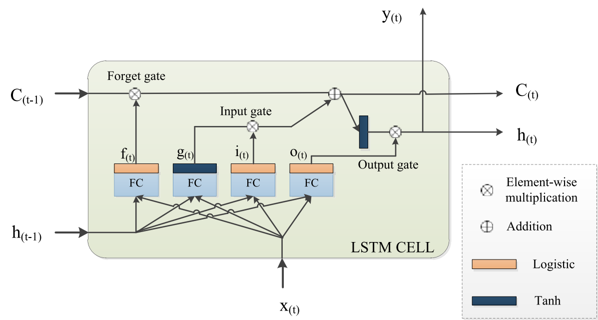


FIGURE 3. Schematic diagram of the internal structure of a cell.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{13}$$
$$c_t = f_t \cdot C_{t-1} + g_t \cdot i_t \tag{14}$$
$$y_t = \tanh(f_t \cdot C_{t-1} + g_t \cdot i_t) \cdot o_t \tag{15}$$

where the variable t is the timestamp. The cell state is C_t, and the output is y_t, which also serves as the input for the next timestamp. '·' denotes the Hadamard product, i_t, f_t, g_t and o_t are the output of the gates. W_i, W_f, W_o, W_g, b_i, b_f, b_o and b_g are coefficient matrixes. Via the function of the different gates, LSTM memory units can capture the complex correlation features within time series in both short and long term, which is a remarkable improvement compared with RNN [23], [24].

As the input and output of different cells interact with each other, the error also spreads with the data. Therefore, the loss function $L(t)$ of the LSTM can be divided into two blocks, one is the loss at the time t, and the other is the loss spread back from timestamp after the time t:

$$L(t) = \begin{cases} l(t) + L(t+1) & t < \tau \\ l(t) & t = \tau \end{cases} \tag{16}$$

where τ is the index of the last timestamp of LSTM. As calculating the back-propagation gradient error, the error of output at time t and the error flowing backward through C_t and h_t both need to be considered. Due to space reasons, the formula for backpropagation will not be described in detail.

C. GWO-LSTM MODEL ANALYSIS

Set m numbers of hyperparameters in LSTM as prey of the grey wolf pack, and take the actual effect of data prediction by LSTM as a criterion for evaluating the position of each

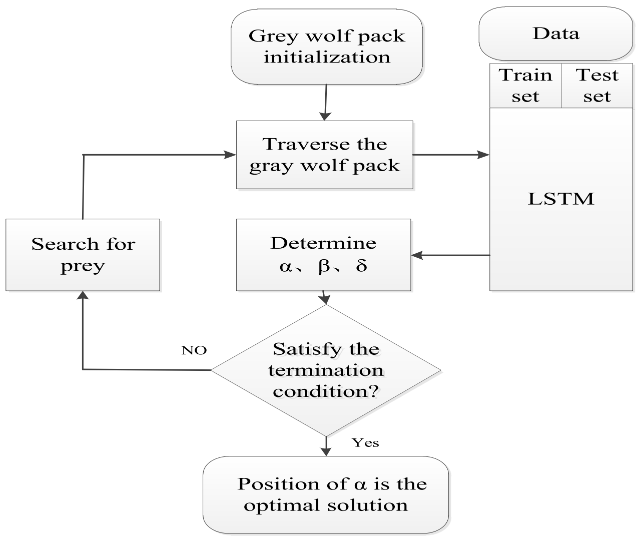


FIGURE 4. The flowchart of GWO-LSTM.

grey wolf. Then simulate to search the prey iteratively. The flowchart is shown in Figure 4.

Step1: Determine the number of grey wolf and the hyper-parameters of LSTM to be optimized. Determine the upper and lower limits of the optimization space. Then randomly generate the grey wolf pack in the space, and determine the number of search iterations.

Step2: Substitute the hyperparameters of LSTM corresponding to the position of each grey wolf. Divide data into two parts as train-set and test-set, then predict the next sequence of train-set by studying the trend of the data. Compare the predicted sequence with the true data in test-set, and calculate the error between predicted sequence and true data. The three wolves with the smallest error are set to be α wolf, β wolf, and δ wolf.

Step3: Under the call of α, β, and δ, the grey wolf pack searches for prey, and the position of each wolf changes according to formulas (1)-(4).

Step4: Repeat Step2-3 for specified times. The hyperparameters corresponding to the position of α wolf are the optimal hyperparameters of LSTM.

III. ERGODICITY AND CONVERGENCE ANALYSIS OF GWO-LSTM

In searching for the optimal hyperparameters of LSTM, the direction and distance of the grey wolf pack's movement are mainly determined by the random vectors r_1 and r_2 in [0,1], the positions of the grey wolf pack in the next timestamp is only related to the current time and position. Therefore, the motion process of the grey wolf pack has no aftereffect, and the sequence of hyperparameters in optimization is Markov process. The following is to prove the ergodicity and convergence of the GWO-LSTM through the relevant properties of Markov chain.

Theorem 1: Set a homogeneous finite Markov chain X with $\{1, \ldots, N\}$ as the state space, and $P = [p_{ij}]$ as the one-step transition probability matrix. Then the necessary and sufficient condition of X being ergodic is that there is a positive integer m such that at least one column of P^m has all elements greater than 0.

Lemma 1: The sequence of the track grey wolf pack go through constitutes a homogeneous finite Markov chain.

Proof: Set the hunting space of the grey wolf pack as a European space of $N \times M$, in which $Xi = (x_1, x_2, \ldots, x_M)$ is the position of the i-th wolf. It can be known from formulas (1) to (9) that the state transition of X_i from step k to step $(k + 1)$ is determined only by the position of step k and the random coefficients r_1 and r_2. Due to the number of grey wolf pack N and the parameter dimension M are positive integers, and there are upper and lower bounds on the search range, and there is a limit on the accuracy of the hyperparameter values, the sequence of the track grey wolf pack optimized is homogeneous finite Markov chain.

Lemma 2: The sequence of the track grey wolf pack go through is ergodic.

Proof: In the European space of hunting, ξ is a hyper-parameter to be optimized, the upper and lower bounds are assumed to be a and b. Set bound a is corresponding to state 1, and bound b is corresponding to state N. Assume that in formulas (1) to (9), ξ has a maximum Euclidean distance for transition from state 1 to state N is d, and the probability is p. Then there must exists $k < [(a - b)/d] + 1$, that if the steps of transition probability matrix is equal to k, the probability from state 1 to N will be greater than 0. Due to that the Euclidean distance from any other state to N is less than d, the transition probability from any other state to N in step k is greater than 0. Therefore, there exit a positive integer $m = k$, that all elements of the N-th column in P^m are greater than 0. According to Theorem 1, the Markov chain is ergodic.

Theorem 2: Literature [25] has proved that if an evolutionary algorithm meets a criterion that any state in the space is reachable and the track sequence is monotonic, the algorithm converges to the global optimal solution with probability 1.

Lemma 3: The sequence of the track α wolf go through converges to the global optimal solution of LSTM with probability 1.

Proof: According to Lemma 2, the sequence of the track grey wolf pack go through is ergodic. Therefore, any state in GWO-LSTM is reachable. In the process that grey wolf pack search for the LSTM hyperparameters, the most dominant figures are α wolf, β wolf, and δ wolf. As the value of the A continuously decrease, the search range of the grey wolf pack is shrinking toward the α wolf, β wolf, and δ wolf. If and only if the position of any other grey wolf is better than α wolf, β wolf, or δ wolf, the status of α wolf, β wolf, or δ wolf will be replaced. Therefore, the values of hyperparameters corresponding to the track of the α wolf are constantly optimized, and the sequence of the track α wolf go through is monotonic. According to Theorem 3, the sequence of the track α wolf go through converges to the globally optimal hyperparameters of LSTM with probability 1.

TABLE 1. Comparison of the prediction effect on different data. WN means white noise, which is added to simulate the noise in data collecting.

	GWO-LSTM	RS-LSTM	Bay	Elastic	GWO-SVR	GBR
sinx	0.0585	0.0715	0.0030	0.4538	0.0029	0.0037
sinx + WN	0.2091	0.2986	1.2956	1.5624	1.0545	1.0608
$1/\sqrt{x}$	0.0018	0.0053	0.0031	0.0049	0.0032	0.0036
$1/\sqrt{x}$ + WN	0.1118	0.3508	1.0408	1.0588	1.0486	1.0796
X^2	1.240 $\times 10^6$	1.853 $\times 10^6$	1.8809 $\times 10^3$	1.8642 $\times 10^8$	1.2273 $\times 10^{12}$	1.6054 $\times 10^{11}$
X^2 + WN	6.390$\times 10^6$	1.064$\times 10^7$	2.086$\times 10^3$	1.8642 $\times 10^8$	1.2273 $\times 10^{12}$	1.6153 $\times 10^{11}$
e^x	3.773$\times 10^9$	2.3914 $\times 10^{10}$	0.1231	0.0088	3.466$\times 10^{10}$	2.866$\times 10^{10}$
e^x + WN	1.022$\times 10^{10}$	1.3916 $\times 10^{10}$	1.0901	2.8006	3.466$\times 10^{10}$	2.866$\times 10^{10}$

(WN=White Noise)

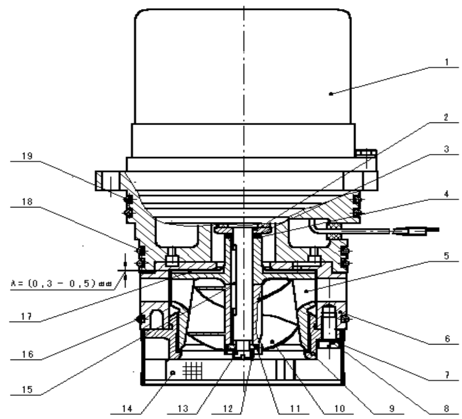
IV. ALGORITHM BENCHMARK

To discover the effectiveness of GWO-LSTM for the prediction on different types of time series, we selected primary functions such as sinx, $x^{1/2}$, $x^{-1/2}$, e^x to generate data, each kind of data has two states, added white noise or non-added white noise. Then we divided data into two set as train-set and test-set. We structured GWO-LSTM to predict the time series of train-set through studying the trend, and analyzed the error compared with the true data in test-set. Some other predicting algorithms as GWO-SVR, RS-LSTM, Bayesian ridge regression, elastic network regression, and gradient boosting regression are also applied as comparison. Hyperparameters of regression methods are all optimized fairly. The steps are as follows:

Step1: Use a function to generate data with 10300 numbers, then set the first 10,000 numbers as the train-set and the last 300 numbers as the test-set.

Step2: Structure GWO-LSTM network, set number of hidden layer nodes and the learning rate as hyperparameters to be optimized, with variable range of 1 to 100 and 0.0001 to 0.1. Timestep is also a hyperparameter. To simplify the model, we consider not to optimize timestep. So, we set the timestep of LSTM as 20 according to experience [26], [27]. Then we can predict the time series of train-set by GWO-LSTM, and we can get 300 numbers as prediction.

Step3: Compare the prediction data with the true sequence in test-set, and calculate the root mean square of the difference between the predicted value and the actual value, which will be taken as the criterion for judging the prediction performance.



1 Motor 2 Washer 3 Disc spring 4 Adjusting washer
5 Impeller 6 Diffusion tube 7 Washer 8 Cylinder head screw
9 Oil pump port 10 Axial flow impeller 11 Adjusting shim 12 Shim
13 Screws 14 Filters 15 Keys 16 Seals
17 Countersunk head screws 18 Seals 19 seals

FIGURE 5. The structure of airborne fuel pump.

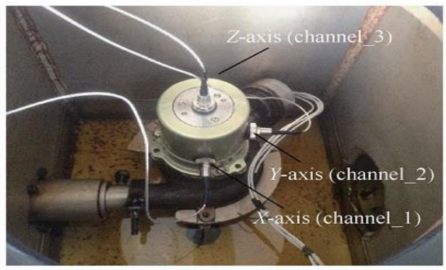


FIGURE 6. The installation of the airborne fuel pump test bench.

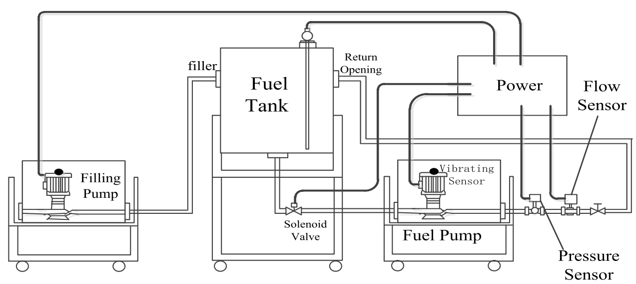


FIGURE 7. Schematic diagram of the airborne fuel pump test bench.

The prediction processes of Bayesian ridge regression, RS-LSTM, elastic network regression, GWO-SVR, and gradient boosting regression are consistent with GWO-LSTM. We adopt the root mean square of the difference between the predicted value and the actual value as the criterion. The results are shown in Table 1.

According to the analysis of the results, for data with trend of high slope, such as data generated by the x^2, e^x functions, the error of GWO-LSTM is larger than the elastic network regression and Bayesian ridge regression. Nevertheless, for data with trend of low slope, especially the data with white noise added, such as sinx, $x^{-1/2}$, GWO-LSTM model

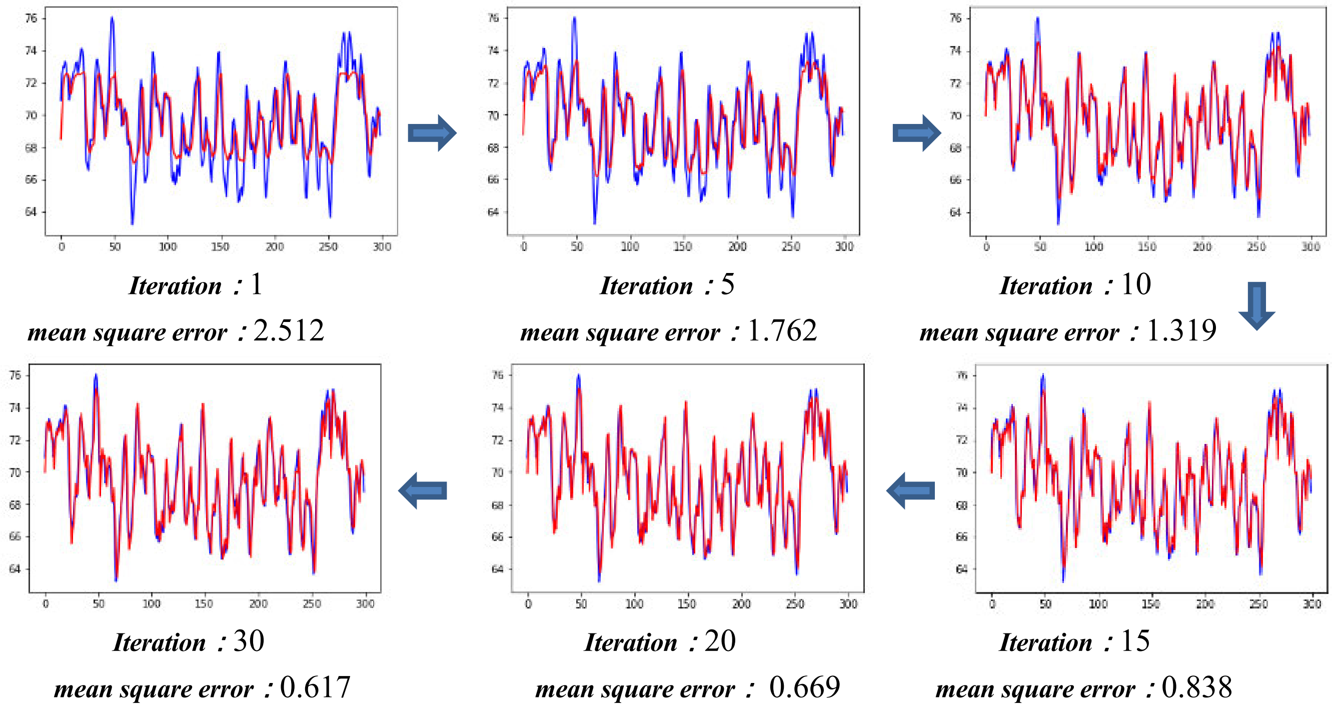


FIGURE 8. **Effect of GWO-LSTM model under different iteration times. The history data is given in the Fig.9. The blue lines denote the actual data after the history data, the red lines denote predicted data trained with history data. To highlight the effect of predicting, the ordinate is shrunk to 64-76 kPa.**

provide very competitive results. It can be concluded that GWO-LSTM model is suitable for predicting data with low overall slope and high partial fluctuation. The conclusion accords with the analytical capabilities for complex input-output relationship of deep learning.

The No Free Lunch (NFL) [28] theorem proved that there is no metaheuristic best suited for solving all optimization problems. Overview the GWO-LSTM model, the core is to increase time cost to get higher accuracy. Every iteration GWO search for hyperparameters, LSTM will be trained with the whole data, and the time cost will plus one times. So, the time complexity is linearly related with searching iteration. Therefore, the proposed algorithm is suitable to deal with cases that requires high precision and consider less about time complexity.

V. APPLICATION IN AIRBORNE FUEL PUMP

The airborne fuel pump is a core component of the fuel system, and the pump is responsible for the fuel supply and fuel transfer for the aircraft. Searching the degradation law of the airborne fuel pump is the basis for the life prediction of the airborne fuel pump. The time sequence of the degradation data of the airborne fuel pump is the key to predict the trend of breaking down, then estimate the life span of the airborne fuel pump [29].

A. STRUCTURE AND WORKING MODE OF AIRBORNE FUEL PUMP

The structure of the fuel pump is shown in Figure 5. When the fuel pump is energized, the flat key transmits the torque of the motor shaft to the impeller and the axial flow impeller. Fuel flows from the storage tank to the impeller, and the centrifugal force generated by the rotation of the impeller is thrown from the impeller blade to the impeller outer diameter. After the fuel flows out of the impeller outlet, it enters the fuel collection pipe of the fuel system and then be pressed into the aircraft fuel main and delivered to the fuel tank [30].

B. CONSTRUCTION OF AIRBORNE FUEL PUMP TEST BENCH

This test bench consists of fuel supply system, program-controlled power supply, and data acquisition system. The fuel supply system is shown in Figure 6 and 7. The fuel pump pumps the oil from the test pump box to the fuel tank. For cycling, the oil in the fuel tank returns to the test pump box by gravity through the solenoid valve. The program-controlled power supply can stabilize the output voltage, and the frequency and amplitude can be adjusted. The data acquisition system mainly includes sensors, data acquisition unit and its software. The sensors include vibration sensors, pressure sensors, and flow sensors. To detect vibration caused by diffuser damage, blade damage, leakage and bearing wear, the position of the vibration sensor was set close to the bearing of the pump. Three vibration sensors (Figure 6) are installed on this experimental bench, and the pressure sensor and flow sensor are installed in the transfer pipe at the pump outlet. The installation position is shown in Figure 7 [29].

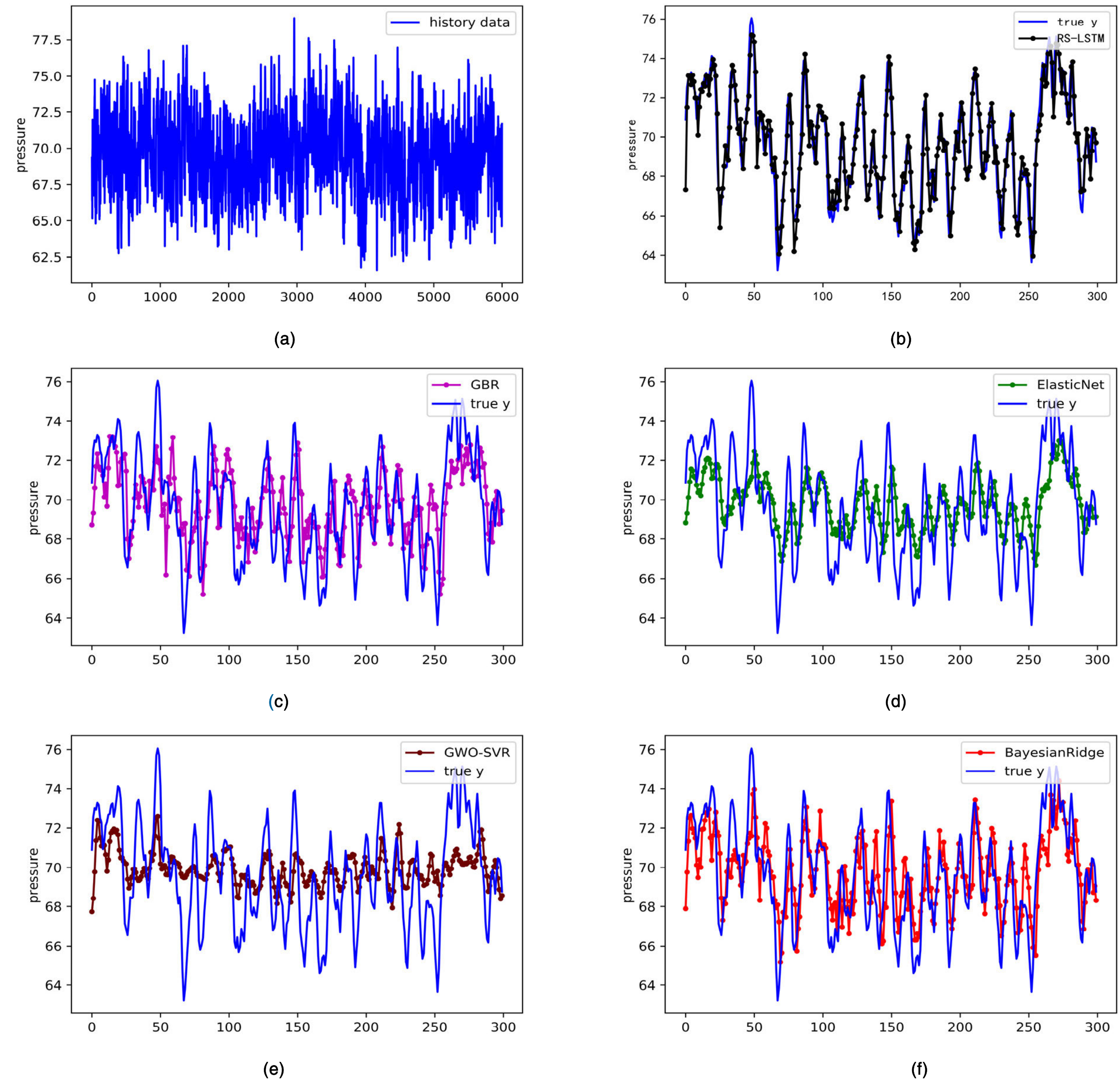


FIGURE 9. Prediction Effect: (a) history data; (b) RS-LSTM; (c) gradient boosting regression; (d) elastic network regression; (e) GWO-SVR; (f) Bayesian ridge regression. To highlight the effect of predicting, the ordinate is shrunk to 64-76 kPa.

C. PREDICTION OF AIRBORNE FUEL PUMP DEGRADATION TREND

The outlet pressure value is an important indicator of the fuel delivery capacity of the airborne fuel pump, which is a good characteristic parameter to measure the performance of the fuel pump. The test adopts pressure as the characteristic for the degradation of the airborne fuel pump performance. We set the rated voltage of the airborne fuel pump as supply power, 115V, 200Hz alternating current was used for a 100-hour degradation test. A sample is taken every minute to obtain a total of 6,000 points of data. Set the last 300 points as the test-set and the rest 5700 points as the train-set. The data processing method is the same as above mentioned.

Through the preliminary analysis of the experimental data, the degradation of the pressure of the airborne fuel pump is very slow, and the pressure data has a certain degree of fluctuation during the degradation process. Based on the study we searched, we consider to use the GWO-LSTM to predict the degradation data of pressure, which will help us to assess the working state of the pump.

We set number of hidden layer nodes and the learning rate of the LSTM algorithm as hyperparameters to be optimized in GWO-LSTM. Then we used GWO-LSTM to predict the next 300 numbers of data by studying the trend of the train-set data with 5700 numbers. We still adopted the root mean square of the difference between the predicted value and the actual value as the criterion for judging the prediction performance. The comparison between the predicted data and the actual data is shown in Figure 8. As the grey wolf pack continues to optimize the hyperparameters of the LSTM, the accuracy

TABLE 2. Mean square error of the prediction.

Methods	Mean square error
Bay	3.0528
RS-LSTM	0.9026
Elastic	4.2035
GWO-SVR	5.3172
GBR	4.3250
GWO-LSTM	0.6170

is also continuously improved. The blue line represents the actual data and the red line represents the predicted data. After 30 iterations, the root mean square is only 0.617, which is fairly lower than 2.512 before the optimization.

Other predicting algorithms as Bayesian ridge regression, RS-LSTM, elastic network regression, GWO-SVR, and gradient boosting regression are also applied as comparison. The prediction processes are consistent with GWO-LSTM. The results are given in Figure 9, and the mean square error of the methods are given in Table 2. The results show that the accuracy of GWO-LSTM in predicting degradation data is much higher than any other algorithm, which further proves that the proposed algorithm is applicable to predict degradation data with high accuracy.

VI. CONCLUSION

In this study, we propose to adopt the grey wolf optimizer (GWO) to search for the best hyperparameters of LSTM. The ergodicity and convergence of the algorithm are proved based on Markov processes theory. The results of prediction show that the GWO-LSTM algorithm is able to provide very competitive results for predicting data with low overall slope and high partial fluctuation. Finally, we apply the algorithm on predicting the pressure degradation trend of the airborne fuel pump, the results show that the root mean square error is only 0.617, which is fairly lower than 2.512 before the optimization.

The limitations of the proposal are as follows: First, the algorithm could provide good results only for predicting certain types of data. As for the degradation data of airborne fuel pump, the accuracy of prediction is higher than any other algorithms we selected. But for data generated by e^x, the accuracy of prediction is not satisfactory. Second, for the process grey wolf pack searching for preys, as the number of iterations increase, the time spend on training the network will relatively increases, which could bring additional cost on prediction.

For future work, we are going to search for more cases that suitable for prediction based on GWO-LSTM, and seek the relation between task complexity and hyperparameters, so as to make further optimization [8].

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," (in English), *Nature*, vol. 521, pp. 436–444, May 2015.
[2] J. Schmidhuber, "Deep learning in neural networks: An overview," (in English), *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[3] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," (in English), *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.
[4] M. Abdughani, J. Ren, L. Wu, J. M. Yang, and J. Zhao, "Supervised deep learning in high energy phenomenology: A mini review," (in English), *Commun. Theor. Phys., Rev.*, vol. 71, no. 8, pp. 955–990, Aug. 2019.
[5] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," (in English), *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
[6] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," (in English), *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[7] Z. Wang, T. Ogawa, and Y. Adachi, "Influence of algorithm parameters of Bayesian optimization, genetic algorithm, and particle swarm optimization on their optimization performance," (in English), *Adv. Theory Simul.*, vol. 2, no. 10, Oct. 2019, Art. no. 1900110.
[8] Z. Cai, Y. Long, and L. Shao, "Classification complexity assessment for hyper-parameter optimization," (in English), *Pattern Recognit. Lett.*, vol. 125, pp. 396–403, Jul. 2019.
[9] J. Moćkus, V. Tiesis, and A. Źilinskas, "The application of Bayesian methods for seeking the extremum," (in English), *Towards Global Optim.*, vol. 2, pp. 117–129, 1978.
[10] S. Yang, M. Li, X. Liu, and J. Zheng, "A grid-based evolutionary algorithm for many-objective optimization," (in English), *IEEE Trans. Evol. Comput.*, vol. 17, no. 5, pp. 721–736, Oct. 2013.
[11] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," (in English), *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.
[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," (in English), *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[13] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2546–2554.
[14] R. Biedrzycki, P. Zawistowski, and B. Twardowski, "Deep learning optimization tasks and Metaheuristic methods," (in English), *Fundamenta Informaticae*, vol. 168, nos. 2–4, pp. 185–218, Sep. 2019.
[15] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: Artificial bee colony (ABC) algorithm and applications," (in English), *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 21–57, Jun. 2014.
[16] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization—Artificial ants as a computational intelligence technique," (in English), *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, Nov. 2006.
[17] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," (in English), *IEEE Trans. Evol. Comput.*, vol. 10, no. 3, pp. 281–295, Jun. 2006.
[18] G. Rudolph, "Convergence analysis of canonical genetic algorithms," (in English), *IEEE Trans. Neural Netw.*, vol. 5, no. 1, pp. 96–101, Jan. 1994.
[19] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," (in English), *ACM Comput. Surv.*, vol. 35, no. 3, pp. 268–308, Sep. 2003.
[20] C. Huang, Y. Li, and X. Yao, "A survey of automatic parameter tuning methods for metaheuristics," (in English), *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 201–216, Apr. 2020.
[21] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf optimizer," (in English), *Adv. Eng. Softw., Article*, vol. 69, pp. 46–61, Mar. 2014.
[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[23] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," (in English), *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.

[24] L. Xu, Q. Li, J. Yu, L. Wang, J. Xie, and S. Shi, "Spatio-temporal predictions of SST time series in China's offshore waters using a regional convolution long short-term memory (RC-LSTM) network," (in English), *Int. J. Remote Sens.*, vol. 41, no. 9, pp. 3368–3389, May 2020.
[25] T. Back, *Evolutionary Algorithms in Theory and Practice*. London, U.K.: Oxford Univ. Press, 1996.
[26] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," (in English), *Neurocomputing*, vol. 275, pp. 167–179, Jan. 2018.
[27] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," (in English), *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018.
[28] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.
[29] X. X. Jiao, B. Jing, X. Q. Qiang, Q. D. Liu, J. Li, and W. Zhou, "Fault diagnosis of airborne fuel pump and experimental platform research," *J. Vib. Shock*, vol. 36, no. 1, pp. 120–128, 2017.
[30] J. X. Pan, B. Jing, Z. G. Li, S. L. Wang, and X. X. Jiao, "Modeling and experimental design of electrical stress accelerated degradation of airborne fuel pump," *J. Electron. Meas. Instrum.*, vol. 33, no. 11, pp. 50–56, 2019.

JINXIN PAN received the B.Sc. degree from Air Force Engineering University, in 2019, where he is currently pursuing the M.Sc. degree. His main research interest includes prognostics and health management.

BO JING received the M.Sc. degree from Air Force Engineering University, in 1996, and the Ph.D. degree from Northwestern Polytechnical University, in 2002. She is currently a Professor with Air Force Engineering University. Her main research interests include prognostics and health management, design for testability, sensor networks, and information fusion.

XIAOXUAN JIAO received the M.Sc. and Ph.D. degrees from Air Force Engineering University, in 2014 and 2019, respectively. He is currently a Lecturer with Air Force Engineering University. His main research interests include information fusion, fault diagnosis, and prognostics.

SHENGLONG WANG received the B.Sc. degree from Air Force Engineering University, in 2018, where he is currently pursuing the M.Sc. degree. His main research interest includes prognostics and health management.