APSO-LSTM: An Improved LSTM Neural Network Model Based on APSO Algorithm

Keqiao Chen

1Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, 16803, the United States

*Corresponding author’s e-mail: kqc5533@psu.edu

Abstract. In the LSTM neural network model, the updating of the weights and threshold parameters depends on the gradient descent algorithm. When the number of hidden layers increases, the convergence rate decreases, and the adjustment of the weights may fall into local extremum, which affects the generalization ability and prediction performance of the LSTM model. Based on this, this paper proposes an improved LSTM neural network model based on APSO algorithm (APSO-LSTM). In this model, the root mean square error is designed as the fitness function, and APSO algorithm is used to build the optimization system. In the verification stage, the weight parameters of each neuron are globally optimized to improve the prediction performance of the model. The experimental results on time series datasets of UCI show that the prediction performance of APSO-LSTM model is significantly better than that of the standard LSTM model, which verifies the rationality of the model.

1. Introduction

With the explosive growth of information, neural networks and artificial intelligence have become the front research hotspot [1]. At present, neural network models have been widely used in speech recognition [2], natural language processing [3,4], video recognition [5,6] time sequence prediction and other tasks. The long-short term memory (LSTM) model [7] was proposed in 1997 by Hochreiter et al. The LSTM model can improve the traditional recurrent neural network (RNN), and introduces a gated unit mechanism to inhibit the gradient disappearance to a certain extent. Due to the capacity of processing time series data, the LSTM model has been applied to the field of time series prediction and has made great achievements.

While the LSTM neural network model still has some shortcomings, such as low learning efficiency, and gradient disappearing when the number of the hidden layer increases. Many experts and scholars are looking various methods to improve the LSTM model. For example, in 2018, Li et al. utilized the CNN model to optimize the LSTM model for news text classification, and got remarkable results [8]; Song et al. proposed an improved deep LSTM neural network, which can predict the trend of time series to a certain extent [9]. But most of the above improved models are based on the network structure of LSTM, which can’t solve the problem that the weight parameters of LSTM are easy to fall into local extremum. At present, more and more researches focus on swarm intelligence algorithms, and utilize them to optimize the weights and thresholds of artificial neural networks. For example, Lin [10] put forward a LSTM neural network model based on genetic algorithm (GA-LSTM) for stock price analysis; ElSaid [11] proposed an optimized LSTM neural network model based on ant colony algorithm (ACO-LSTM), etc.
Based on this, this paper proposes an improved LSTM neural network model based on APSO algorithm (APSO-LSTM). The RMSE value of the prediction results in LSTM neural network is designed as the fitness function in this model, which utilizes the APSO algorithm to construct the optimization space. And then in the model verification stage, the weight values between the LSTM network nodes are globally optimized to improve the model performance. Among them, the APSO algorithm is an improved particle swarm optimization algorithm proposed by Yang et al. in 2010 [12]. Compared with the standard PSO algorithm, the APSO algorithm is easier to obtain the global optimal solution and has a faster convergence speed.

2. APSO algorithm and the LSTM neural network

2.1. PSO algorithm and APSO algorithm

Particle swarm optimization (Particle Swarm Optimization, PSO) is a heuristic optimization algorithm proposed by Eberhart [13]. By simulating the biological process of bird foraging, the particles are iterated continuously in a reasonable space to search the feasible solution. Among them, each particle represents a candidate solution for an optimization problem.

Formulas (1) and (2) are the particle velocity and position update equations in the PSO algorithm.

\[
V_{t+1}^{i,j} = wV_{t}^{i,j} + \alpha r_{t}^{i,j} \left( y_{t}^{i,j} - x_{t}^{i,j} \right) + \beta r_{t+1}^{i,j} \left( \mathbf{y}^{i,j} - x_{t}^{i,j} \right)
\]

\[
x_{t+1}^{i,j} = x_{t}^{i,j} + V_{t+1}^{i,j}
\]

Where, at time t, \( V_{t}^{i,j} \) is the particle’s speed; \( x_{t}^{i,j} \) is its position; \( y_{t}^{i,j} \) is the historical best position of the particle in search process; \( \mathbf{y}^{i,j} \) is the global best position of the whole population, and the subscript \( j \) is the number of spatial dimensions. \( w \) is the inertia weight, \( r_{t}^{i,j} \) and \( r_{t+1}^{i,j} \) are random numbers; \( \alpha \) and \( \beta \) present learning parameters. In formula (1), \( wV_{t}^{i,j} \) refers to the current state of particles, \( \beta r_{t+1}^{i,j} \left( \mathbf{y}^{i,j} - x_{t}^{i,j} \right) \) means that particles start from the state of the whole population to carry out global optimization; \( \alpha r_{t}^{i,j} \left( y_{t}^{i,j} - x_{t}^{i,j} \right) \) means that particles start from their own state to carry out local optimization.

It has been proved that [14], in most optimization scenarios, we needn’t to use the historical best position to update the particle state in PSO algorithm, which will increase the training cost and cause slow convergence, and random number technology can realize the diversity of particles. The APSO algorithm [12] needn’t to set the initial speed of particles, and can utilize random number technology to replace it. The process is simplified, the convergence speed is accelerated, and the global optimal is easier to obtain. Formulas (3) and (4) are renewal equations of particle velocity and position in APSO algorithm.

\[
V_{t+1}^{i,j} = wV_{t}^{i,j} + \alpha r_{t}^{i,j} \left( y_{t}^{i,j} - x_{t}^{i,j} \right)
\]

\[
x_{t+1}^{i,j} = (1-\beta)x_{t}^{i,j} + \beta y_{t}^{i,j} + \alpha r
\]

In formula (3), \( r \) is a random number whose value range is between [0,1]. Using the random number technique to replace \( \alpha r_{t}^{i,j} \left( y_{t}^{i,j} - x_{t}^{i,j} \right) \) in formula (1) can greatly improve the convergence speed and the meaning of other parameters is consistent with the standard PSO algorithm. In formula (4), the parameter \( r \) makes the particles more mobile. If the value of \( \alpha \) is reasonable, the APSO algorithm can completely avoid the phenomenon of local extremum. In general, \( \alpha \in [0.1,0.5], \beta \in [0.2,0.7] \).
Due to its excellent search performance, the APSO algorithm is used to optimize and adjust the weight parameters of the LSTM neural network in the model verification stage.

2.2. The LSTM neural network

In the LSTM neural network, the unique gated units are used to learn and memorize the sequence data to maintain long-distance time series information dependence and achieve high-precision prediction. The LSTM neuron structure is shown in Figure 1.

![Figure 1. The structure of the LSTM neuron.](image-url)

As shown in Figure 1, the LSTM neuron has the input gate, output gate, and forget gate. The input gate mainly processes input data. The forget gate determines the current neuron’s retention of historical information. The output gate represents the output result of the neuron. Let the input sequence is \((x_1, x_2, \ldots, x_t)\), then at time \(t\), the calculation formula of each parameter of LSTM neuron is as follows:

\[
\begin{align*}
    i_t &= S(W_i \ast [h_{t-1}, x_t] + b_i) \\
    f_t &= S(W_f \ast [h_{t-1}, x_t] + b_f) \\
    o_t &= S(W_o \ast [c_t, h_{t-1}, x_t] + b_o) \\
    c_t &= f_t \ast c_{t-1} + i_t \ast \tanh(W_c \ast [h_{t-1}, x_t]) \\
    h_t &= o_t \ast \tanh(c_t)
\end{align*}
\]

Among them, \(x_t\) is the input to the LSTM neuron at time \(t\); \(h_{t-1}\) indicates the output state of the hidden layer at time \(t-1\); \(i_t, f_t,\) and \(o_t\) are the input of the input gate, the forget gate, and the output gate at time \(t\) respectively; \(W_i, W_f,\) and \(W_o\) are respectively the weight matrix of the input and the input gate, the forget gate and the output gate of the neuron at time \(t\); \(b_i, b_f,\) and \(b_o\) are their corresponding offset vectors; \(W_c\) is the weight between the input and the cell unit; and \(h_t\) represents the output of the hidden layer at time \(t\); \(S\) is the Sigmoid function.

The deep LSTM neural network is composed of LSTM neural units, and thus constructs the network model with multiple hidden layers, and continuously removes redundant data by the forget gates to achieve higher accuracy. Therefore, the LSTM model has better time series predictive performance.
3. The APSO-LSTM model

3.1. Particle Encoding
The APSO-LSTM model proposed in this paper adopts the real number encoding. Figure 3 shows the LSTM neural network structure with 3 hidden layers. Based on this, the coding mode of APSO algorithm is discussed.

![LSTM neural network](image)

Figure 2. The structure of LSTM neural network with 2 hidden layers

A chromosome in the AGA-LSTM model corresponds to an individual in the population, and also represents a set of weight values between nodes in the LSTM model.

According to the LSTM network structure in Figure 2, Figure 3 is drawn, which represents the encoding mode of a particle in the APSO-LSTM model. For example, \( w_{2}^{HH} \) represents the weight between the second neuron in the input layer and the second neuron in the first hidden layer, and so on.

![Particle encoding](image)

Figure 3. The encoding of particles

3.2. Fitness function
In the APSO-LSTM model, the fitness function of particles is designed to evaluate these particles of the population. In the model validation phase, the RMSE of the output value of the LSTM model and the actual value is designed as the fitness function to measure the importance of particles. The smaller RMSE is, the more reasonable the weights setting of the corresponding LSTM model of the particle is, and the stronger the generalization ability of the model is.

The root mean square error, usually expressed as RMSE. It shows that the RMSE of the global optimal solution obtained by the particle \( i \) in the iterative process on the validation set. The calculation process is shown in (10).

\[
RMSE_i = \sqrt{\frac{1}{n_v} \sum_{j=1}^{n_v} (d_{ij} - y_{ij})^2}
\]

(10)

Among them, \( n \) is the number of population; \( n_v \) is the number of validation set; \( d_{ij} \) is the prediction result of LSTM model corresponding to particle \( i \); \( y_{ij} \) is the actual value of validation set.

3.3. The flow of APSO-LSTM model
In this paper, the APSO algorithm is used to optimize the weights between the nodes of the LSTM neural network, which can enhance the generalization ability.
The APSO-LSTM model, that is, after the initial training of the LSTM network, the initial weights are mapped through APSO algorithm in the model verification stage, and the weights between each node are mapped to a certain dimension attribute of the particle, so that each particle represents the solution set of the candidate weights of the whole neural network. Figure 4 shows the flow of LSTM neural network model based on APSO optimization, $n$ is the number of iterations, $N$ is the maximum number of iterations and the main steps are as follows:

1. Initialize the LSTM model and APSO algorithm, such as the structure of LSTM network, the number of nodes, the number of APSO population, the number of iterations and other parameters.

2. Train the LSTM neural network using test set to obtain the default initial weights.

3. Use Formula (10) to calculate the particles in the APSO population.

4. Search for the global optimal particles. Comparing the fitness value of each particle with that of the optimal particle, the new global optimal particle is selected as the one with the smaller one.

5. The velocity and position of all particles are updated according to formula (3) and (4).

6. Add 1 to the number of iterations $n$, and judge whether the current value of $n$ is greater than the maximum number of iterations. If so, skip to step (7); otherwise, skip to step (3).

7. Output the global optimal particle, which corresponds to the optimal weight distribution of the LSTM network.

4. Experimental results and analysis

In this paper, the time series prediction models based on GRU, LSTM, PSO-LSTM and APSO-LSTM neural networks are established. Among them, PSO-LSTM is an LSTM model that optimizes the network weights based on the standard PSO algorithm. And by comparing their prediction accuracy, the performance of the proposed APSO-LSTM model is verified. The experimental environment is as follows: deep learning framework tensorflow1.10.0, the algorithm programming language is Python3, and the computer operating system is Ubuntu 16.4.

4.1. Datasets

In this paper, the shampoo sales data set (SSD) and the daily lower temperature data set (MDTD) of DataMarket database and the ozone level detection data set (OLD) of the UCI database are selected. Among them, SSD dataset depicts the sales data of a shampoo in three years, and this experiment mainly forecasts its monthly sales volume; MDTD dataset describes the weather temperature data of Melbourne for 10 consecutive years, and this experiment mainly forecasts its daily minimum temperature; the OLD dataset records the ground ozone concentration data, and this experiment mainly forecasts the ozone concentration in the future period, which all belongs to the time series prediction problem.

4.2. Evaluation index of models

After preprocessing, the first 70% of the datasets is taken as training sets, 20% as verification sets and 10% as test sets. After model training and verification, the prediction models based on GRU, LSTM, PSO-LSTM and APSO-LSTM neural networks are established respectively, which are applied to the
time series analysis of the above three datasets. The evaluation index of the model adopts the sum of error squares SSE, as shown in formula (11):

$$\text{SSE} = \sum_{j=1}^{n_{\text{test}}} (y^*_j - y^*_j)^2$$  \hspace{1cm} (11)

Among them, \(n_{\text{test}}\) is the number of test sets, \(y^*_j\) is the prediction result of the model, \(y^*_j\) is the actual value. The smaller the SSE value, the higher the prediction accuracy of the model.

4.3. Analysis of experimental results

Table 1 shows the SSE values of 4 models of GRU, LSTM, PSO-LSTM and APSO-LSTM on SSD, MDTD and OLD data sets. Figure 5 is a histogram drawn according to Table 1.

| Data set  | The prediction model |
|----------|----------------------|
| SSD      | GRU  | 322     | LSTM  | 298.4 | PSO-LSTM | 277.8 | APSO-LSTM | 245.6 |
| MDTD     | 281.8 | 309.6 | 267.1 | 243.2 |
| OLD      | 389.7 | 361.3 | 310.3 | 301.8 |
| Average SSE value | 331.2 | 323.1 | 285.1 | 263.5 |

As can be seen from Table 1, the SSE values of the LSTM model under the three different data sets are almost the largest except on the SSD dataset, followed by the GRU model and PSO-LSTM, and the SSE values of the APSO-LSTM model are the lowest on each data set, indicating that the model has the highest time series prediction accuracy, and its generalization ability are best.

Figure 5 shows the trend of the average SSE value of the three models under each dataset. The average SSE value of the GRU model is 2.4% lower than that of the LSTM model, while the average SSE value of the PSO-LSTM model is 13.9% lower than that of the LSTM model and 11.8% lower than that of the LSTM model. And the proposed APOS-LSTM model is 7.6% lower than he PSO-LSTM model. Obviously, the AGA-LSTM model has the smallest error and the best prediction accuracy. This proves that PSO algorithm and ASPO algorithm can promote the weight optimization of LSTM neural network, and APSO algorithm has the best effect and is easier to obtain the global optimal solution.
Table 2. The prediction results of 4 prediction models on Melbourne weather. (Unit: °C)

| Day | Actual temperature | GA-LSTM | ACO-LSTM | PSO-LSTM | APSO-LSTM |
|-----|--------------------|---------|----------|----------|-----------|
| 1   | 20.7               | 18.4    | 18.5     | 19.6     | 20.5      |
| 2   | 17.9               | 18.9    | 17.2     | 18.6     | 18.3      |
| 4   | 14.6               | 12.1    | 12.6     | 12.7     | 13.7      |
| 5   | 18.3               | 17.2    | 17.5     | 19       | 18.4      |
| 6   | 17.1               | 16      | 16.3     | 16.5     | 16.7      |
| 7   | 20.3               | 21.5    | 18.9     | 19.6     | 20.1      |
| 8   | 19.6               | 21.2    | 19.5     | 20.5     | 19.9      |
| 9   | 15.5               | 14.3    | 14.9     | 16.1     | 15.1      |
| 10  | 16.8               | 16.2    | 16.4     | 17.3     | 16.9      |

The MSTD dataset records the weather temperature data of Melbourne for 10 consecutive years. In order to further enhance the persuasiveness of the experiment, during the validation process, the actual value of daily minimum temperature in a month of Melbourne for 10 consecutive days and the predicted values of four models are recorded, and the results are shown in Table 2. In order to compare the prediction performance of APSO-LSTM model with other three models, figure 6 and Figure 7 are drawn according to Table 2.

Figure 6 shows the prediction results of APSO-LSTM model for Melbourne weather for 10 consecutive days. It can be seen that the maximum error between the predicted curve and the actual temperature curve is no more than 1 °C, and the prediction effect is precise.

As shown in Figure 7, the predicted results of four models (LSTM, GRU, PSO-LSTM and APSO-LSTM) on Melbourne weather for 10 consecutive days are shown. It can be seen that the error value of LSTM model is the largest, and the maximum error exceeds 3 °C. Only the curve of APSO-LSTM model is closest to the actual value curve, which further shows the effectiveness and innovation of the proposed APSO-LSTM model.
5. Conclusion

In this paper, aiming at the problems of slow convergence speed and easy to fall into local extremum when modifying weights in standard LSTM neural network, an improved LSTM neural network model based on APSO optimization is proposed. The experimental results show that, compared with GRU, LSTM and PSO-LSTM models, the average SSE value of APSO-LSTM model in 3 datasets is the minimum, and the prediction performance is significantly improved. Therefore, APSO algorithm provides a novel idea for optimization of the LSTM neural network.

References

[1] Wu J, Long J, Liu M. Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm[J]. Neurocomputing, 2015, 14(8):136-142P

[2] Zhang Y, Lu X. A Speech Recognition Acoustic Model Based on LSTM-CTC[C]// 2018 IEEE 18th International Conference on Communication Technology (ICCT). IEEE, 2019

[3] Yao L, Guan Y. An Improved LSTM Structure for Natural Language Processing[C]// 2018 IEEE International Conference of Safety Produce Informatization (IICSI). IEEE, 2018.

[4] Meng F, Lu Z, Tu Z, et al. Neural Transformation Machine: A New Architecture for Sequence-to-Sequence Learning[J]. 2015.

[5] Donahue J, Hendricks L A, Rohrbach M, et al. Long-term Recurrent Convolutional Networks for Visual Recognition and Description.[M]// AB initio calculation of the structures and properties of molecules /. 2015.

[6] Liu D, Bellotto N, Yue S. Deep Spiking Neural Network for Video-Based Disguise Face Recognition Based on Dynamic Facial Movements[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(6):1843-1855.

[7] HOCHREITERS, SCHMIDHUBERJ. Long shortterm memory[J]. Neural Computation, 1997, 9(8): 1735-1780

[8] Li C, Zhan G, Li Z. News Text Classification Based on Improved Bi-LSTM-CNN[C]// 2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE Computer Society, 2018.

[9] Song B, Fan C, Wu Y, et al. Data Prediction for Public Events in Professional Domains Based on Improved RNN- LSTM[J]. Journal of Physics Conference Series, 2018, 976(1):012007.

[10] Lin M, Chen C. Short-term prediction of stock market price based on GA optimization LSTM neurons[C]//Proceedings of the 2018 2nd International Conference on Deep Learning Technologies. 2018: 66-70.
[11] ElSaid A E R, El Jamiy F, Higgins J, et al. Optimizing long short-term memory recurrent neural networks using ant colony optimization to predict turbine engine vibration[J]. Applied Soft Computing, 2018, 73: 969-991.

[12] Yang X S. Nature-inspired metaheuristic algorithms[M]. Luniver press, 2010.

[13] Eberhart R, Kennedy J. Particle swarm optimization[C]//Proceedings of the IEEE international conference on neural networks. 1995, 4: 1942-1948.

[14] Gandomi A H, Alavi A H. Krill herd: a new bio-inspired optimization algorithm[J]. Communications in nonlinear science and numerical simulation, 2012, 17(12): 4831-4845.