Short-term Load Forecasting Model Based on IBFO-BILSTM

Zhengnan Zhou, Guoqing Wu* and Xudong Zhang
School of Electrical Engineering, Nantong University, Nantong, China

*Corresponding author e-mail: wgq@ntu.edu.cn

Abstract. Short-term power load forecasting has always been a very important position in power grid operation. Improving prediction accuracy has great significance for improving the power company's dispatch efficiency and economic benefits. In view of the feed-forward neural network cannot learn the time series relationship between load data, a hybrid depth model based on improved bacterial foraging algorithm (IBFO) and long short-term memory (LSTM) network is proposed for short-term power load forecasting. Using historical load data and its associated temperature data and date information as input, the initial weights and bias are optimized by improved bacterial foraging algorithm, and the multi-layer bidirectional LSTM neural network is used to mine the hidden time series relationship between input data, and finally output prediction. The proposed prediction model was verified by using the historical load data of Nantong. The data shows that the proposed model has higher prediction accuracy than other models.

1. Introduction

Electricity has long been one of the indispensable energy sources for modern human beings. As the main body of electric energy production and consumption, dispatching and distribution, power system control regional power load interval is conducive to the rational development and utilization of regional energy resources, and can meet the growing demand of the regional national economy for electricity. As an important part of power system scheduling, power load prediction has been a hot topic of research by scholars concerned.

The methods of power load forecasting are broadly divided into two categories [1], one is mathematical statistics method, which mainly includes regression analysis method, time series method, gray prediction method, Auto Regressive Integrated Moving Average (ARIMA) model, etc. These methods are simple and easy to implement, but they are subject to many factors and rely too much on the model to express the nonlinear characteristics of the power load data. The other type is artificial intelligence technology prediction methods, including expert system method [2], artificial neural network [3], Support Vector Machine (SVM) [4] and so on. These methods can well exploit the nonlinear relationship between load and climate, policy and other factors, but lack of discussion on their own timing factors. Long Short-Term Memory (LSTM) neural network [5] is ideal for load prediction because it adds time-sharing prior knowledge to the feed-in neural network, enabling it to combine nonlinearity with timing.

However, LSTM network has the disadvantages of being prone to local optimal, relying heavily on the selection of the initial weight bias, and the difficulty of selecting the hyper-parameters of deep neural network. In order to overcome the above shortcomings and improve the accuracy of power load
forecasting, some scholars have been influenced by heuristic algorithms, and they have improved their neural networks and achieved good results. For example, the literature [6] uses the genetic algorithm (GA) to find the optimal hidden layer number of the LSTM network, the number of hidden units in each hidden layer, and the optimal time series of the input layer. The literature [7] used particle swarm optimization (PSO) algorithm to screen the length of the output layer, the type of neurons in the hidden layer, and the hyper-parameters of the activation functions of each layer, both of which have good performance, but they do not consider the influence of the initial weight bias.

In view of the above problems, a hybrid model of short-term load forecasting, namely IBFO-LSTM model, combined with Improved Bacterial Foraging Optimization (IBFO) and deep LSTM network is proposed. The model uses a multi-layer LSTM neural network with a large amount of load data, date information, meteorological data, etc. as inputs, and deeply constructs the characterization semantics of load fluctuations. Meanwhile, the standard bacterial foraging algorithm is improved to have better optimization ability. The purpose is to find the optimal initial weight and bias of the LSTM network to improve the convergence ability of the LSTM network. Experimental results show that the IBFO-LSTM model has high prediction accuracy compared with other algorithms.

The paper is organized as follows. In section 2, we introduced an improved bacterial foraging optimization algorithm. In section 3, we introduced the IBFO-BILSTM prediction model. The results of prediction were shown in section 4. Finally, in section 5, we gave summaries and conclusions.

2. Improved bacterial foraging optimization algorithm

In 2002, Passino simulated the foraging behavior of Escherichia coli and proposed a bacterial foraging optimization algorithm, which is favored by scholars at home and abroad because of its calculation method of obtaining optimal solution through multiple iterations of global search. Its mathematical model is mainly divided into three parts: chemotaxis, reproduction and migration. The standard BFO algorithm has some problems in finding the optimal weight bias of the neural network, such as insufficient search precision, large search workload and repeated work. The three operations of the standard bacterial foraging algorithm are described and improved below.

2.1. Chemotaxis

The change produced by moving the bacteria i in one direction is defined as

$$\theta(j+1,k,l) = \theta(j,k,l) + C(i) \frac{\Delta(i)}{\sqrt{\Delta^T(i) \Delta(i)}}$$  

(1)

Where $\theta(j,k,l)$ is the position of the bacterium i after the l-th migration, the k-th reproduction and the j-th chemotaxis; $C(i)$ is the chemotaxis step size; $\Delta(i)$ is the random direction vector.

The step size of the standard BFO algorithm is fixed, considering that the step length should be increased in the early stage of the algorithm to carry out the global search, and in the later stage should reduce the step to improve the search accuracy, the bacterial individual adaptive coefficient $\alpha$ is introduced here, taking into account the effect of optimal bacterial location and iteration times on the bacterial individual. The improved step length formula can be expressed as

$$C = C_{\min} + (C_{\max} - C_{\min}) \beta$$

(2)

$$\beta = \alpha(1 - e^{-\frac{j+k+l}{j_{\max} + k_{\max} + l_{\max}}}) + (1 - \alpha) \frac{j+k+l}{j_{\max} + k_{\max} + l_{\max}}$$

(3)
\[ \alpha = \sin\left(\frac{\pi}{2} \frac{N - i}{N_c}\right) \]  

(4)

Where \( C_{\text{max}} \) and \( C_{\text{min}} \) are the minimum and maximum step values; \( \beta \) is the step adjustment factor; \( \alpha \) is the individual adaptive coefficient of bacteria; \( J_i \) is the fitness value of the current bacteria; \( J_{\text{max}} \) is the maximum fitness value recorded in this algorithm; \( j_{\text{max}}, \ k_{\text{max}}, \ l_{\text{max}} \) are the maximum number of times of chemotaxis, reproduction and migration.

2.2. Reproduction

For the i-th bacterium, define the fitness function as

\[ J = \sum_{j=1}^{N_c+1} J(i, j, k, l) \]  

(5)

The standard BFO algorithm arranges the fitness and removes the bacteria with less adaptability in the first half, and divides the other half of the more adaptable bacterial individual into two identical sub-bacteria. This step follows the rules of nature's survival and is conducive to increasing the convergence speed of the algorithm, but the diversity of the direction of population search has decreased.

In view of above shortcomings, the copying operation was improved as follows: after retaining the first half of the \( S_{r1} \) bacteria who has higher fitness value, the latter half of the bacteria \( S_{r2} \) directly replicated the optimal bacteria to perform the chemotaxis operation. Thus, the \( S_{r1} \) bacterial group performs the chemotaxis operation of the formula (1), retaining the diversity of the search and the search globality of the algorithm. The \( S_{r2} \) bacterial group performs the chemotaxis operation of formula (6), which further improves the search accuracy of the algorithm.

\[ \theta(j+1, k, l) = \theta_{\text{best}}(j, k, l) + C(i) \frac{\Delta(i)}{\sqrt{\Delta^t(i)\Delta(i)}} \]  

(6)

2.3. Migration

Bacterial individuals of the standard BFO algorithm will migrate according to the following mechanisms:

\[ x' = \begin{cases} x', & \text{if } q < P_{\text{ed}} \\ x, & \text{otherwise} \end{cases} \]  

(7)

Where \( x' \) is the new region after migration, \( q \) is a random number in the interval \((0, 1)\), and \( P_{\text{ed}} \) is the set migration probability.

Based on the idea of roulette, the fixed migration probability is calculated as

\[ P = P_{\text{ed}} \ast \frac{J_{\text{max}} - J_i}{J_{\text{max}} - J_{\text{min}}} \]  

(8)

Where \( J_{\text{max}} \) and \( J_{\text{min}} \) are the maximum and minimum fitness values.
In this way, bacteria with higher fitness will migrate with less probability, whereas bacteria with less fitness will have greater migration probability.

3. IBFO-BILSTM prediction model

3.1. BILSTM neural network model

LSTM is a type of Recurrent Neural Network (RNN). In order to solve the problem of gradient explosion and gradient disappearance in the traditional RNN network, LSTM adds a memory unit that can store long-term information based on RNN. This special structure makes the LSTM network have unique advantages in learning the hidden knowledge of time series prediction. The unit structure is shown in Figure 1.

![Figure 1. Unit structure of LSTM.](image)

The basic unit of the LSTM network is composed of the forget gate $f_t$, the input gate $i_t$ and the output gate $o_t$. The forget gate will discard some information from the transient cell state, and the input gate will identify what new information remains in the cell state. LSTM update the cell state through following functions.

\[
f_t = \sigma(W_{sf}x_t + W_{hf}h_{t-1} + b_f) \tag{9}
\]

\[
i_t = \sigma(W_{si}x_t + W_{hi}h_{t-1} + b_i) \tag{10}
\]

\[
o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{11}
\]

\[
c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{12}
\]

\[
h_t = o_t \odot \tanh(c_t) \tag{13}
\]

Where $W_{sf}, W_{hf}, W_{si}, W_{hi}, W_{xo}, W_{ho}, W_{xc}, W_{hc}$ are the corresponding weight matrix; $b_f, b_i, b_o, b_c$ are the corresponding bias matrices; sigmoid is activation function; $\odot$ represents point multiplication.
The unidirectional LSTM calculates from the beginning to the end of the sequence, so it can only predict the future from current and past inputs. Sometimes, the prediction may need to be determined by several inputs before and after the current input, which is more accurate. Therefore, a bidirectional LSTM has been proposed, using two recurrent neural networks, one calculates the hidden output $h^f_t$ from the beginning, and the other calculates the hidden output $h^b_t$ from the back. Figure 2 shows an unfolding bidirectional LSTM. This paper forms a bidirectional LSTM according to the following parallel processing:

$$h_t = h^f_t + h^b_t$$  \hspace{1cm} (14)

### 3.2. IBFO-BILSTM neural network model

#### 3.2.1. Model structure

The prediction model architecture diagram constructed in this paper is shown in Figure 3. The model structure method is described as follows.

1) Input layer: The input is an array of cells composed of a matrix whose dimensions are $M \times N$ (365 \times 48), where $M$ is the total days of training data, and $N$ is the number of sequential steps. Each sample is a row of cell arrays. Each circle represents a timing input and the input matrix is described as

$$X = [x_1, x_2, x_3, x_4, x_5, x_6]$$  \hspace{1cm} (15)

Where $x_1$ is the single-sequence power load; $x_2$ is month number (1-12); $x_3$ is day of month (1-31); $x_4$ is day of week (1-7); $x_5$ is the time node of day (1-48); $x_6$ is the average temperature of the day.

2) Bidirectional LSTM layer: Two bidirectional LSTM layers both contain 20 hidden neurons. In order to avoid over-fitting, both layer have a dropout layer, whose rate is 0.2.

3) Fully connected layer: Two simple hidden layers contain 30 and 6 hidden neurons.

4) Output layer: Only output one-dimensional load data.
3.2.2. Optimization. Before training, IBFO algorithm is used to optimize the initial weights and bias of the LSTM network. Specific steps are as follows:

1) Structurally coding the weight and bias parameter of the BILSTM network correspond to the individual bacteria in the IBFO algorithm, and the value of the weight and bias parameter represents the position of the bacteria in the search space;

2) In order to expand the search range and ensure the high precision of the weight and bias, the coding rules use floating point number coding. Set the bias and weight in the range of [-1, 1], assign it to the population of the IBFO algorithm, and randomly initialize the population;

3) Decoding the dimensional information of the bacterial population into the weight and bias parameters, simulating the output of the network, and calculating the square of the error as the fitness function of the algorithm. The sum of the squared errors is:

$$
SSE = \sum_{i=1}^{T} (y_i - \hat{y}_i)^2
$$  \hspace{1cm} (16)

where: \( T \) is the total number of prediction results, that is, the total number of time steps; \( y_i \) and \( \hat{y}_i \) are the actual load value and the predicted load value of the i-th sampling point on the forecast day.

4) Iterate according to the IBFO algorithm until the maximum iteration step termination algorithm is reached or the optimal bacteria are searched;

5) Obtain optimized initial weights and bias for prediction.

In the training process, Adam (adaptive moment estimation) optimization algorithm is adopted. Adam is a first-order optimization algorithm that can replace the traditional stochastic gradient descent process. It can iteratively update the neural network weights based on the training data. The algorithm is computationally efficient and requires less memory. It is suitable for solving optimization problems with large-scale data and parameters, and has a good advantage for solving the unsteady problem of large noise and sparse gradient of power load data.
4. Experiment and result analysis

To verify the situation of the model in the actual forecast, this paper uses historical load values and temperature data collected in Nantong, whose sample interval is 30 minutes. The software platform of this paper is based on Matlab2018a’s Deep Learning Toolbox framework.

4.1. Data pre-processing

Data such as load and temperature for the first 11 months were selected as the training set, and the load data for the 12th month was used as the test set.

In order to improve the convergence speed of the model, the data needs to be Z-score standardized so that the data conforms to the standard normal distribution. Z-score can be expressed as

$$x^* = \frac{x - \mu}{\sigma}$$

(17)

Where $\mu$ is the mean of the sample data; $\sigma$ is the standard deviation of the sample data; $x^*$ is the normalized value; $x$ is the value before normalization.

4.2. Model evaluation

To evaluate the performance of the proposed IBFO-BILSTM model, three common metrics were used.

$$y_{MAPE} = \frac{1}{T} \sum_{i=1}^{T} \left| \frac{y_i - \bar{y}}{y_i} \right|$$

(18)

$$y_{RMSE} = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (y_i - \bar{y}_i)^2}$$

(19)

$$y_{MFA} = \frac{1}{T} \sum_{i=1}^{T} \left( 1 - \frac{|y_i - \bar{y}_i|}{y_i} \right) \times 100\%$$

(20)

4.3. Results and discussions

In this experiment, the control variable method is used to perform stepwise model parameter tuning. The influence of learning rate on model prediction performance is shown in Table 1. It can be seen that when learning rate is at 0.03, model performance is the best.

| Learning rate | Epochs | Batch Size | MAPE/% |
|---------------|--------|------------|--------|
| 0.01          | 2000   | 32         | 2.95   |
| 0.02          | 2000   | 32         | 2.74   |
| 0.03          | 2000   | 32         | 2.63   |
| 0.04          | 2000   | 32         | 2.81   |

Table 1. Performance of IBFO-BILSTM model on different learning rate.

The parameters of the IBFO algorithm are set as follows: number of bacteria $S=10$, times of chemotaxis $N_c=4$, times of bacterial swimming $N_s=4$, maximum swimming step size $C_{max}=0.05$, ...
minimum swimming step length $C_{min}=0.01$, number of times of reproduction $N_{re}=4$, times of migrations $N_{ed}=4$, and the migration probability $P_{ed}=0.125$.

In order to verify the prediction results of the proposed model better, it is compared with the SVM model, the MLP model and the ARIMA model.

| Model      | MAPE/% | RMSE   | MFA/% |
|------------|--------|--------|-------|
| SVM        | 3.00   | 24.32  | 97.00 |
| MLP        | 3.08   | 26.08  | 96.92 |
| ARIMA      | 2.63   | 21.19  | 97.37 |
| IBFO-BILSTM| 2.05   | 17.18  | 97.95 |

Table 2 is an overall comparison of the load forecast errors of each model for a given day in December. It can be seen that the proposed model has less error compared with other models, with an average prediction accuracy of 97.95% and a MAPE of 2.05%, which is 0.95%, 1.03%, and 0.58% higher than other models. It reached 17.18, which was 7.14, 8.9, and 3.82 lower than other models.

Figure 4 shows the curve of the actual load change trend of each model. It can be seen that when predicting the peak-to-valley value, the three can't learn the load fluctuation law very well, especially when there are obvious gaps at 15 nodes. The IBFO-BILSTM model proposed in this paper has the best fitting effect on the actual load, among which there are 10 points with absolute relative error less than 1% and 20 points at 1%-2%.

5. Conclusion
In this paper, the improved bacterial foraging algorithm is applied to the initial weight bias selection of bidirectional LSTM neural network. A short-term electric load forecasting method based on IBFO-BILSTM model is proposed. By constructing a multi-layer hybrid neural network, time-series hidden
information of different characteristics such as load temperature and date are extracted to predict the power load.

The data shows that compared with other algorithms, the proposed prediction model can better improve the accuracy of short-term power load forecasting, and has better applicability to load data with timing characteristics. The limitation of this paper is that the data used still lacks more factors that affect the power load, such as electricity price and rainfall. Future research work can expand the application range of the bacterial foraging algorithm and further integrate with the bidirectional LSTM neural network to explore the optimal selection of hidden parameters such as the number of hidden layers, activation functions, and optimization functions.

Acknowledgments
This work is supported by National Natural Science Foundation of China (No. 61273151); Jiangsu Provincial Science and Technology Department Policy Guidance Program (Production, Research and Research Cooperation) - Forward-looking Joint Research Project (No. BY2015047-04); Science and Technology Plan Project of Nantong (No. GY12016042).

References
[1] Srivastava A. K., Pandey A.S., Singh D., Short-term load forecasting methods: A review, (2016) International Conference on Emerging Trends in Electrical, Electronics and Sustainable Energy Systems, ICETEESES 2016, art. no. 7581373 , pp. 130 - 138.
[2] Sun Heru, Wang Wei, Electric short-term load forecasting using artificial neural networks and fuzzy expert system, Advances in Intelligent and Soft Computing, v 112, p 699 - 707, 2011, Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE2011) November 19-20, 2011, Melbourne, Australia: Volume 3: Computer Networks and Electronic.
[3] Chen L.-G., Chiang H.-D., Dong N., Liu R.-P, Group-based chaos genetic algorithm and non-linear ensemble of neural networks for short-term load forecasting, (2016) IET Generation, Transmission and Distribution, 10 (6), pp. 1440 - 1447.
[4] Zhang Anan, Zhang Pengxiang, Feng Yating, Short-term load forecasting for microgrids based on DA-SVM, COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, v 38, n 1, p 68-80, January 7, 2019.
[5] Kong W., Dong Z.Y., Jia Y., Hill D.J., Xu Y., Zhang Y, Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network, (2019) IEEE Transactions on Smart Grid, 10 (1), art. no. 8039509, pp. 841 - 851.
[6] Bouktif S., Fiaz A., Ouni A., Serhani M.A., Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches, (2018) Energies, 11 (7), art. no. 1636.
[7] Wen L., Zhou K., Yang S., Lu X., Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting, (2019) Energy, 171, pp. 1053 - 1065.