Power Load Prediction Model Based on Long Short Term Memory and Sparrow Search Algorithm

Weiyue Huo\textsuperscript{1,}\textsuperscript{*}, and Jing Zhou\textsuperscript{1}
\textsuperscript{1}Zhejiang University, Hangzhou, China

\textsuperscript{*} Corresponding author: 1102155477@qq.com

Abstract. Power load forecasting is an important reference for power demand evaluation. In order to promote the accuracy and stability of load short-term prediction, this paper brings up an innovative power load prediction method based on improved Long Short Term Memory (LSTM) recurrent neural network. Sparrow Search Algorithm (SSA) is employed to optimize the initial weights and bias of LSTM layers, which can boost the accuracy and convergence rate of the model. Finally, the processed dataset is used to train and test the SSA-LSTM model and the simulation demonstrates superiority of the proposed method.

1. Introduction

With the growth of the economy and population, energy consumption and the peak load of power network has dramatically increased over the past decades. This reveals the importance of power system planning, operation, and control. Accurate prediction of the power load is vital for energy management, equipment update and power grid construction. Overestimation of power load leads to unnecessary reservation and extra cost. On the other hand, underestimation may increase overload risk and result in a massive power outage [1]. In the deregulated electricity market, improvement of the load prediction accuracy will produce considerable economic and social benefits [2]. According to the interval of prediction period, power load prediction can be divided into following three categories: (1) Short-term load forecasting (STLF) which is usually from one hour to one week [3]. (2) Medium term load forecasting (MTLF), ranging from one week to one year, and (3) Long-term load forecasting (LTLF) which is longer than one year [4]. Regardless of different types of load prediction, it is a considerable challenge to establish a precise, robust and flexible energy-prediction model. Currently, different methods have been proposed to achieve an accurate prediction, which can be classified into physical models artificial intelligence-based models [5][6] and hybrid models [7][8]. To incorporate the nonlinearity of power load series, AI based models have become one of the most popularly-used methods with outstanding generalization capacity and self-adaptive ability [9][10]. The advent of smart meters has made the energy consumption data available, making AI data-driven forecasting models possible. The classic AI models applied in power load prediction mainly contain the artificial neural network (ANN)[11], the decision tree (DT), and the support vector machine (SVM)[12]. ANN and SVM are the most widely used models AI methods. However, ANN and SVM also has some obvious shortcomings. ANN is difficult to choose input parameters networks structure. What’s more, it is easy to fall into the local minimum point. SVM can better solve the problems of local minimum point, but SVM is hard to be implemented for large-scale training samples. Among AI based models, long short term memory networks strategies (LSTM) have gradually attracted widespread attention in energy forecasting, it has intrinsic advantages in learning nonlinear and time-
series data[13]. As an effective nonlinear recurrent neural network, LSTM network links the input data over time through the design of the network structure, making it suitable for dealing with energy forecasting problems.

With the development of machine learning techniques, the state-of-the-art approaches are usually able to fit the training sample pretty well. However, single model sometimes does not guarantee a good performance for out-of-sample forecasting, incurred by the limitation of training sample and overfitting problem[9]. As a remedy to this problem, ensemble learning technique has been employed for training and forecasting, where several AI models are generated and integrated based on bagging or boosting. In this paper, an improved LSTM model combined with Sparrow Search Algorithm is proposed. SSA has many advantages such as fast convergence and good global positioning ability. Because the selection of initial value of weights and bias in LSTM model is difficult and has significant effect on the performance of power load prediction, SSA optimization algorithm is utilized to optimize the initial model state replacing the traditional traversal method.

The rest of the paper is organized as follows. Section II of this paper demonstrates the algorithm principles of LSTM and SSA and elaborates the proposed load forecasting methodology. In Section III, daily hourly power load data from September 2013 to August 2015 in a district of Nanjing is used for analysis and verification. Finally, we draw the conclusion of the paper and discuss the future work in Section IV.

2. The SSA-LSTM Model

This part briefly introduces LSTM algorithm and Sparrow Search algorithm (SSA).

2.1 Long Short-Term Memory Architecture

Long Short-Term Memory (LSTM) have emerged as an effective and scalable model for sequential data recently. Unlike Earlier methods such as Recurrent Neural Network, LSTM model is both adaptive and effective at capturing long-term temporal features. They do not suffer from the optimization obstacles that plague simple recurrent networks (SRNs) and have been widely used to deal with many difficult sequence problems. This includes handwriting recognition, natural language processing, Power load data prediction and so on [14].

![Figure 1. Schematic of LSTM block](image)

The core feature of LSTM lies in cell state and gate structure. A schematic of LSTM basic cell can be seen in Fig. 1. The hidden layer cell of LSTM has 3 inputs at time $t$: input $x_t$ of the current input layer, the output $h_{t-1}$ of the previous hidden layer, and the state value $c_{t-1}$ of the previous cell. Each cell generates two outputs, which are the current output $h_t$ and current cell state value $c_t$. The cell also features three gates (input, forget, and output). These gates can be defined by the following set of equations:

**Forget gate:**

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

**Input gate:**

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

Where $\sigma$ is the sigmoid activation function, $W_f$, $W_i$, $b_f$, and $b_i$ are the weight matrices and bias vectors, respectively.
Output gate:

\[ o_t = \sigma(W_t[h_{t-1}, x_t] + b_t) \]  

(3)

Where \( W \) and \( b \) present the weight and bias of each gate, \( \sigma \) is the sigmoid function.

Memory cell:

\[ c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh (W_c[h_{t-1}, x_t] + b_c) \]  

(4)

Cell output:

\[ h_t = o_t \otimes \tanh (c_t) \]  

(5)

Where \( \tanh \) is the activation function of the LSTM cell and the \( \otimes \) presents matrix multiplication.

LSTM can record past information for a long time and has outstanding learning ability for both large and small datasets.

2.2 Sparrow Search Algorithm

Sparrow Search Algorithm (SSA) is a new swarm intelligence optimization algorithm proposed by K. X. Jia and B. Shen in January 2020[14]. This algorithm is inspired by the behavior of sparrow and uses sparrow position to present optimization variables. It has been proved that SSA has better accuracy, convergence speed, stability and robustness compared with other swarm intelligence algorithms.

1) Population behavior characteristics: In the sparrow population, sparrow population is divided into producer and scrounger. Producers are responsible of searching for food. The scrounger follows the producer in order to get food. According to observations, sparrows can flexibly switch between producer and scrounger.

2) Mathematical Model: In the mathematical model of SSA, the position of sparrow population is represented by the following \( n \times d \) matrix:

\[ X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,d} \end{pmatrix} \]  

(6)

In this equation, \( n \) represents the amount of the sparrow population; \( d \) is optimization dimension. Each individual sparrow’s fitness is shown in the following equation:

\[ \text{Fitness} = [f(X_1), f(X_2), \ldots, f(X_n)]^T \]  

(7)

Each element in \( X \) Fitness is the fitness value of this sparrow.

The producers search the foraging area on a large scale and update the location according to (8)

\[ X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp \left( \frac{-i}{\lambda x T} \right) & \text{if } R_2 < ST \\ X_{i,j}^t + Q \cdot L & \text{if } R_2 > ST \end{cases} \]  

(8)

where \( t \) represents iteration. \( X_{i,j}^t \) is the coordinate of the \( j \)th dimension of sparrow \( i \). \( R_2 \in [0,1] \) indicates a warning value and \( ST \) is the safety threshold. \( ST \) usually takes a value of 0.8. \( Q \) is a normally distributed random number. \( L \) is a matrix where all the elements are 1, and the size is \( 1 \times d \). \( T \) is total number of iteration.

When \( R_2 < ST \), it shows that the sparrow population is safe and will search on a large scale. Conversely, the sparrows quickly move to other areas. For scroungers, they update the position according to (9)

\[ X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp \left( \frac{X_{i,j}^t - X_{i,j}^*}{\lambda} \right) & \text{if } i > n/2 \\ X_{p}^{t+1} + [X_{i,j}^t - X_{p}^{t+1}] \cdot A^+ \cdot L & \text{otherwise} \end{cases} \]  

(9)
Among them, $X_p$ is the best position among producers, $X_{worst}$ is the worst position of the population in this iteration, $A$ is a $1 \times d$ matrix and the elements are randomly assigned as 1 or -1. It indicates search direction.

When $i > n/2$, it means that these sparrows are very hungry, and they will fly to other places by themselves. The remaining scroungers move around the producers because of the rich food and compete with the them to become the producers.

The algorithm assumes that 10% to 20% of the sparrows can be aware of the danger. The sparrows were randomly selected and their position is updated according the equation (10):

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \kappa \cdot |X_{i,j}^t - X_{best}^t| & \text{if } f_i > f_g \\ X_{i,j}^t + K \cdot \left( \frac{|X_{i,j}^t - X_{worst}^t|}{(f_i-f_w) + \epsilon} \right) & \text{if } f_i = f_g \end{cases}$$

(10)

Among them, $X_{best}$ is the best position of the population in this iteration; $\kappa$ is the parameter which controls step; $K$ is a random number of 1 or -1; The fitness value of sparrow is $f_i$, $f_g$ and $f_w$ are the best and worst fitness values of the population. $\epsilon$ is a constant.

**Figure 2.** Program flow chart of SSA

In (10), when $f_i > f_g$, sparrows are at the edge of the population, so they fly to the best position $X_{best}$ of the population. $f_i = f_g$ indicates that these sparrows are already in the best position $X_{best}$. Due to their alertness, they need to move closer and stay away from the worst position $X_{worst}$ in the meanwhile. Through the above description and modeling of sparrow population behavior, the flow chart of SSA is shown in Fig. 2.
2.3 The Proposed Power Prediction Algorithm

To enhance both prediction accuracy and stability, this paper proposes a hybrid model SSA-LSTM that employs the SSA to optimize the initial weights and bias of LSTM. The major steps of the SSA-LSTM model are elaborated below:

1) Model initialization: Initialize the parameters of the SSA including sparrow population, position, composition and the maximum number of iteration. Initialize the structure of LSTM and the weights and bias of LSTM model are taken as optimization targets.

2) Objective function: The objective function of SSA is the mean absolute error (MAE) of predicted value of untrained LSTM model compared with raw data.

3) Optimization: Update the positions of the sparrows based on the results of the objective functions achieve the optimal initial value of LSTM when reach the maximum iteration number.

4) LSTM training: The particle value of SSA is used as the LSTM model weights and bias. Then the training dataset is input into optimized LSTM model to get the final forecasting model.

3. Experiment and Analysis

3.1 Data Description

In order to evaluate the proposed model, the experiment was carried out using the actual load data. This is a dataset reports on daily hourly power load and relevant variables from September 2013 to August 2015 in a district of Nanjing, China.

In a power system, the weather information is usually the dominant variable in driving the electricity demand. This dataset includes the date-time, power load, temperature, pressure, wind direction, wind speed and the cumulative number of snow and rain collected from Supervisory Control and Data Acquisition (SCADA) system.

The Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE) are used as metrics to evaluate the performance of the model. The three error measures are displayed as follows:

\[
RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (R_t - F_t)^2}
\]

(11)

\[
MAE = \frac{1}{T} \sum_{t=1}^{T} |R_t - F_t|
\]

(12)

\[
MAPE = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{R_t - F_t}{R_t} \right|
\]

(13)

where \(R_t\) is the real data, \(F_t\) is the predicted value, and \(T\) means the total number of testing data. The entire training data is divided into two groups: training and validation sets. Training sets is used to train different models while the load prediction using validation data set is employed to test the model effectiveness. Comparisons among different models are given below.

3.2 Experiment Results and Analysis

The prediction errors for each reference model are given in TABLE (I), and the comparison of prediction results of each reference model for datasets are given in Fig. 3.

Table 1. Comparison of different models.

| Model     | RMSE(kW) | MAE(kW)   | MAPE(%) |
|-----------|----------|-----------|---------|
| ANN       | 205.99   | 161.146   | 0.0304  |
| LSTM      | 118.507  | 87.8497   | 0.0167  |
| SSA-LSTM  | 110.224  | 64.887    | 0.0104  |
As can be seen from TABLE(I) and Fig. 3, the following conclusion can be achieved.

1) The proposed SSA-LSTM model have lower forecasting error, indicating an improved accuracy of power load prediction.

2) Compared with conventional LSTM model, the SSA-LSTM model reduces the prediction error by about 40%, suggesting its potential for practical application.

3) The SSA-LSTM prediction method proposed in this paper follows the curve of real values closely, which shows outstanding performance of out of sample.

4. Conclusion

In this paper, an innovative show-term load forecasting model combining SSA with LSTM has been proposed. The proposed methodology has been used to predict the power load in a district of Nanjing. The Experiment result shows the proposed method is superior to the traditional LSTM model and ANN. The proposed algorithm results about 90% of accuracy with the validation test. Improving LSTM model with sparrow search algorithm can overcome the defect of falling into local optimal point easily and further optimize the network performance. In the future research, we plan to employ the proposed algorithm to deal with other time series data to check its effectiveness and accuracy.

5. References

[1] S. Fan, L. Chen, and W. Lee, IEEE Transactions on Industry Applications, 45, no. 4, pp. 1460–1466, (2009).

[2] D.W.Bunn and E.D.Farmer, Comparative Models for Electrical Load Forecasting. Wiley, (1985).

[3] W. Chandramitasari, B. Kurniawan, and S. Fujimura, SAIN, (2018), pp. 43–48.

[4] S. Mishra and S. K. Patra, in TENCON 2008 IEEE Region 10 Conference, (2008), pp. 1–6.

[5] H. Hippert, C. Pedreira, and R. Souza, IEEE Transactions on Power Systems, 16, no. 1, pp. 44–55, (2001).

[6] J. W. Taylor and P. E. McSharry, IEEE Transactions on Power Systems, 22, no. 4, pp. 2213–2219, (2007).

[7] Szilagyi and P. Wira, in 2018 ICPS, (2018), pp. 20–25.

[8] X. Liao, X. Kang, M. Li, and N. Cao, in 2019 ICITBS, (2019), pp. 305–308.

[9] J. W. Taylor and R. Buizza, IEEE Transactions on Power Systems, 17, no. 3, pp. 626–632, (2002).

[10] W. Le, in 2018 (CICED), (2018), pp. 1462–1466.

[11] S. Singh, S. Hussain, and M. A. Bazaz, in 2017 (ICIIP), 2017, pp. 1–5.

[12] R. Gao, L. Zhang, and X. Liu, in Proceedings of the 10th World Congress on Intelligent Control and Automation, 1048–1051, (2012).

[13] X. Hui, W. Qun, L. Yao, Z. Yingbin, S. Lei, and Z. Zhisheng, in 2017 ICPRE, (2017), pp. 589–591.

[14] J. Xue and B. Shen, 8, 22–34, (2020).