Short-term load forecasting based on MA-LSSVM

Liangkun Wei, Jianguo Li *
School of Shanghai DianJi University, Shanghai, China

*Corresponding author e-mail: lijiang@sdju.edu.cn

Abstract. Aiming at the problems of low accuracy and poor accuracy of short-term load forecasting (STLF), monkey algorithm (MA) and least square support vector machine (LSSVM) are combined for STLF. The input factors of the model are load data and meteorological information. MA is used to optimize the kernel function parameter $\sigma$ and the regularization parameter $\lambda$ of LSSVM to obtain the optimal solution of the least square support vector machine prediction model. Then a STLF model was established.

1. Introduction
Electricity has become one of the most important energy sources in today’s society, and it plays an indispensable role in many fields [1]. Accurately mastering the past power load data and mining the potential value of the data are of great significance to the grid companies and people’s lives [2]. STLF is one of the important methods of the power load forecasting system. Through forecasting, we can better plan the power generation capacity of power plants and dispatch power plants, and improve grid safety management. Therefore, short-term load is an important research topic in smart grid projects.

At present, STLF has received extensive attention at home and abroad. With the deepening of research topics, more and more scholars use traditional methods and intelligent algorithms to improve prediction accuracy. In the literature [3], the historical data of power load is clustered according to the distance between the two data, and the combination method is used to divide the categories into three categories. The same type of data is used as the input of the neural network for load forecasting, which improves the accuracy of load forecasting. In the literature [4], a large user power load forecasting method based on load clustering, RBF neural network conjugate gradient learning and similarity weighting synthesis is proposed, and good forecasting results have been achieved. In the literature [5], CSO is used to optimize the BP neural network, and the prediction model of the CSO-BP neural network is established. The results show that the prediction effect is better.

In order to solve the problems in the above-mentioned literature, this paper takes advantage of MA high accuracy, rapid convergence process and strong ability to deviate from the local optimum, and proposes a load forecasting model based on MA optimization LSSVM to solve the blind selection of LSSVM to improve the accuracy of load forecasting.

2. Algorithm introduction

2.1. Principle of Least Squares Support Vector Machine
The improvement of LSSVM to SVM is that it uses the sum of square errors to replace the original loss function, and changes the inequality constraints to equality constraints [6], reducing the complexity of
calculations, and making the prediction accuracy and operating speed better than SVM. Suppose \( n \) training samples are \((x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), 1 \leq i \leq n\), where \( x_i \) is the correlation vector that affects the predicted power, and \( y_i \) is the predicted power. Output vector, use nonlinear mapping \( \phi(x) \) to construct optimal decision function formula (1) in high-dimensional feature space:

\[
f(x) = \omega^T \phi(x) + b
\]  

In the formula, \( \omega \) is the weight, and \( b \) is a constant.

LSSVM transforms the regression problem into a constrained optimization problem by using the principle of structural minimization. The optimization goal formula (2) is:

\[
\min \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \lambda \sum_{i=1}^{n} \xi_i^2
\]

The constraint formula (3) is:

\[
y_i [\omega^T \phi(x_i) + b] = 1 - \xi_i, i = 1, 2, \ldots, n
\]

In the above formula, \( \xi_i \) is the relaxation factor, and \( \lambda \) is the regularization parameter.

List the Lagrange function formula (4) of the above optimization problem:

\[
L(\omega, b, \xi, \alpha) = \frac{1}{2} \|\omega\|^2 + \gamma \sum_{i=1}^{n} \xi_i^2 - \sum_{i=1}^{N} \alpha_i [\omega \phi(x_i) + b + \xi - y_i]
\]

According to the conditions, the L function takes the partial derivative of each variable and makes the result zero to obtain the formula (5):

\[
\begin{align*}
\frac{\partial L}{\partial \omega} &= 0 \Rightarrow \sum_{i=1}^{n} \alpha_i \phi(x_i) = \omega \\
\frac{\partial L}{\partial b} &= 0 \Rightarrow \sum_{i=1}^{n} \alpha_i = 0 \\
\frac{\partial L}{\partial \xi} &= 0 \Rightarrow \alpha_i = \lambda \xi_i \\
\frac{\partial L}{\partial \alpha} &= 0 \Rightarrow \phi(x_i) \omega + b - y_i + \xi_i = 0
\end{align*}
\]

A linear equation (6) can be obtained through the above conditions:

\[
\begin{bmatrix}
0 & 1 & \cdots & 1 \\
1 & k(x_1, x_1) & \cdots & k(x_n, x_1) \\
\vdots & \vdots & \ddots & \vdots \\
1 & k(x_n, x_1) & \cdots & k(x_n, x_n)
\end{bmatrix}
\begin{bmatrix}
\frac{1}{\lambda} \\
\frac{1}{\lambda} \\
\vdots \\
\frac{1}{\lambda}
\end{bmatrix}
\begin{bmatrix}
\alpha \\\
a_1 \\\n\vdots \\
a_n
\end{bmatrix} =
\begin{bmatrix}
b \\
y_1 \\
\vdots \\
y_n
\end{bmatrix}
\]

The formula (7) of the LSSVM prediction model is obtained by solving the above formula as:

\[
y_i = \sum_{i=1}^{n} a_i k(x_i, x_i) + b
\]

Among them, \( k(x_i, x_i) \) is the kernel function.

The four commonly used kernel functions of LSSVM are polynomial kernel function, radial basis function, sigmoid function, and linear function. The radial basis kernel function (8) is:

\[
k(x, x) = \exp \left( \frac{||x-x||^2}{2\sigma^2} \right)
\]

Because of its relatively simple structure, strong adaptability, and fast convergence speed, this paper chooses it as the kernel function of the model.

2.2. Principle of Monkey Algorithm

MA is a swarm intelligence algorithm that simulates mountain climbing in monkey swarms. It is mainly used to solve global optimization problems with continuous variables. The basic idea is to simulate the three processes of climbing, looking, and jumping of monkey climbing a mountain, so as to perform new and iterative intelligent calculations [7].

In order to find the most value point in the entire domain, suppose there are \( m \) groups of monkeys, each group of \( n \) monkeys, a total of \( m \times n \) monkeys climbing the mountain. By initializing each monkey to start from a random location where it is, and climb up with a specific step length \( a \). When climbing to the highest point of the field, look around at the sight distance \( b \). If you can find a mountain
that is higher than its own position, jump to this mountain, and then continue the climbing process, and so on.

2.2.1. **Representation of the solution.** Define $M$ as the population size of the monkey group. If a monkey $x_i(i = 1,2,3 \cdots M)$ is randomly selected in the population, the position of the monkey is $x_i(x_{i1},x_{i2},x_{i3}, \cdots x_{in})$, each component $x_{ij}$ in the formula represents the actual position of each monkey in each dimension. The actual position of each monkey represents a feasible decision vector for the optimization problem.

2.2.2. **Initialization.** The initialization of the population has an important impact on the global convergence and optimization effect of the algorithm. If the population is initialized in the optimal solution neighborhood, the algorithm will converge faster and the accuracy of the solution will be higher. If the population is initialized in a larger range during initialization, it is far from the optimal solution, the algorithm's convergence speed is slow, and the global optimal solution accuracy obtained is not high. If the population is initialized in a small range, the algorithm can easily fall into a local optimal solution. Therefore, the basic MA adopts random initialization in the solution space, and the process is as follows (9):

$$x_{i,j} = x_{\text{min},j} + (x_{\text{max},j} - x_{\text{min},j}) \cdot \text{rand}$$  \hspace{1cm} (9)

Among them, $x_{\text{min},j}$ and $x_{\text{max},j}$ indicate that there is an upper bound $x_{\text{max},j}$ and a lower bound $x_{\text{min},j}$ in the j-th dimension space. Rand means to generate a random real number in the interval $[0,1]$.

2.2.3. **Climbing process.** The climbing process of MA is a process of gradually improving the value of the objective function of the optimization problem through iteration [8]. Different from other algorithms, the post-group algorithm uses the idea of pseudo-gradient to design the climbing process of monkey $i$:

A randomly generated vector $\Delta X_i = (\Delta x_{i1}, \Delta x_{i2}, \cdots \Delta x_{in})$, $\Delta x_{ij}$ satisfies the formula (10):

$$\Delta x_{ij} = \begin{cases} a, & \text{with probability } \frac{1}{2} \\ -a, & \text{with probability } \frac{1}{2} \end{cases}$$  \hspace{1cm} (10)

Among them, the parameter $a(a > 0)$ is the step length of the climbing process, and its size is mainly determined by the optimization problem itself. The larger the solution, the larger the climbing step, and the smaller the solution, the smaller the climbing step.

The pseudo gradient of the objective function $x_i$ is (11):

$$f'_i (x_i) = \frac{f(x_i + \Delta x_i) + f(x_i - \Delta x_i)}{2\Delta x_i}$$  \hspace{1cm} (11)

Among them, $j = 1,2,3 \cdots n$.

Calculate $y_i = x_{ij} + a \cdot \text{sign}(f'_i (x_i))$ by $Y = \langle y_1,y_2, \cdots y_n \rangle$.

Among them, $j = 1,2,3 \cdots n$.

If $Y$ is feasible, let $Y = X_i$; otherwise, keep the value of $X_i$ unchanged.

Repeat the above steps until the maximum number of times is reached or there is almost no change in the objective function value during the two iterations.

2.2.4. **Look-jump process.** After the monkey group has climbed the set number of times, each monkey reaches the highest mountain in the current position, that is, reaches the local optimal value. At this time, the monkey looks for a point in the field of vision that is better than the current position by looking at it, and then jumps away from the current position. This process is mainly to speed up the monkey swarm algorithm to find the optimal solution to the problem. Proceed as follows:

Randomly generate a real number $y_j$ in interval $(x_{ij} - b,x_{ij} + b), j = 1,2, \cdots n$, and $Y = \langle y_1,y_2, \cdots, y_n \rangle$. 
When $Y$ meets the condition $f(y_i) \geq f(x_i)$, then let $x_i = y_i$. Otherwise, repeat the previous step until a $y$ point that satisfies the condition is generated.

Use $y$ as the initial position and repeat the climbing process.

### 2.2.5. Somersault process.

The main purpose of the somersault process is to force the monkey group to move from the current search area to a new area, so as to avoid the monkey group falling into the local optimum. Choose the center of all monkeys’ positions as the pivot point, and each monkey will turn to a new area in the direction pointing to the pivot point or the opposite direction. For the $i$-th monkey, the translation process is as follows:

Randomly generate a real number $\alpha$ in the interval $[c, d]$. The interval is called the flip interval, and it usually depends on the specific situation.

Let $y_i = x_{ij} + \alpha(p_j - x_{ij})$, and $p_j = \frac{\sum_{i=1}^{M} x_{ij}}{M}, j = 1, 2, \ldots, n$ be the midpoint of all monkey positions. Point $P = (p_1, p_2, \ldots, p_n)$ is called the somersault pivot. If $\alpha \geq 0$, the monkey will flip in the direction that the position points to the fulcrum of the somersault, and vice versa.

When $Y = (y_1, y_2, \ldots, y_n)$ meets the condition, then $x_i = y$. Otherwise, repeat the above steps until a feasible $y$ is produced.

### 2.3. Improved LSSVM prediction model

In this article, MA is used to optimize the kernel function parameter $\sigma$ and the regularization parameter $\lambda$ of LSSVM to keep them at the optimal level. The specific steps are as follows:

1. Set the relevant parameters of the monkey group algorithm, use the parameter $\sigma$ and the parameter $\lambda$ as the initial position of the monkey, and initialize the population;

2. The input contains the input vector and the expected output value. Use the forecast date history data as the training sample, including temperature, week type, and load. Training samples include temperature series on a certain day, weekly series, and load series on the previous day or the past two days. The output vector is the load sequence at a certain time. 24 hours a day, 24 points, 24 training sessions;

3. The formula for calculating the fitness value is as follows (12). Use the prediction value obtained by learning the training sample with the LSSVM model, and calculate the absolute percentage error of each individual monkey as the fitness value of the MA according to the formula;

$$F = \sum_{i=1}^{M} \left| \frac{A_i - P_i}{A_i} \right|$$

Among them, $A_i$ is the actual measured value, and $P_i$ is the predicted value;

4. Use the climbing process to search for local optimal solutions;

5. Reuse the look-jump process to search for a better position and climb to a better position;

6. Use somersaults to move to a new position and search again;

7. When the number of iterations reaches $T_{max}$, the optimization ends and the optimized parameters are output, otherwise the number of iterations $t = t + 1$, go to step (3);

8. Input the optimized parameters into LSSVM to obtain the prediction model and make predictions; the step flow chart is shown in Figure 1.
3. Case analysis

3.1. Basic data
In this experiment, 768 data from February 1 to March 4, 2015 in a certain area of Suqian, Jiangsu Province were used as training samples to predict the 24-hour coincidence of the next day. Algorithm parameter settings: monkey swarm algorithm population size $M = 50$, climbing step length $a = 0.01$, climbing times $N_c = 100$, observation field $b = 0.5$, somersault interval $[c, d] = [-1, 1]$ and maximum The number of iterations $N = 60$.

3.2. Experimental results and analysis
Compare the predicted results of the MA-LSSVM model with the actual load value, the PSO-LSSVM model and the predicted value of the LSSVM model, as shown in Figure 2.

It can be seen from Figure 2 that the overall change trend of the forecast results of the three STLF models is consistent with the actual power change trend, but the load forecast curve of the MA-LSSVM forecast model has a higher degree of fit to the actual load.

It can be seen from Figure 3 that the minimum prediction error of the prediction model proposed in this paper is 1.69MW, and the maximum value is 3.51MW, both of which are smaller than the maximum and minimum errors of the PSO-LSSVM prediction model and the LSSVM prediction model, and it can
be seen that the MA=LSSVM model is significantly smaller than the other two models in the fluctuation of the prediction error range, and the error fluctuation range is smaller and more concentrated. It can be seen that the forecasting effect of the forecasting model proposed in this article is very good.

![Figure 3. Absolute error curve.](image)

The calculation results of the three error indicators are shown in Table 1.

| Predictive model | $e_{MSE}$ | $e_{MAPE}$ | $e_{MAD}$ |
|------------------|-----------|------------|-----------|
| MA-LSSVM         | 6.852     | 1.074      | 2.531     |
| PSO-LSSVM        | 13.902    | 1.539      | 3.620     |
| LSSVM            | 24.149    | 2.004      | 4.754     |

According to Table 1, the mean square error of the prediction model proposed in this paper is 50.7% lower than the PSO-LSSVM prediction model and 71.6% lower than the LSSVM prediction model. The other two error evaluation indicators are also lower than the PSO-LSSVM prediction model LSSVM prediction model. This verifies the validity of the MA-LSSVM model.

4. Conclusion
This article proposes a method of optimizing the kernel function parameter $\sigma$ and the regularization parameter $\lambda$ of LSSVM through MA to establish a STLF model. The accuracy and effectiveness of this model are verified through experiments, and the results show that the MA-LSSVM prediction model has higher accuracy and better results.

References
[1] jiangdong. Study on Two-layer Real-time Optimization Game in Active distribution network [J]. Electrical Measurement & Instrumentation, 2018, 55(18): 50-56.
[2] Wei Ruidong, Yu Peng, Gao Song, et al. Power communication network routing configuration mechanism based on multiple ant colony algorithm [J]. Journal of Beijing University of Posts and Telecommunications, 2017(S1): 89-92.
[3] Shi Yuheng, Zhao Na, Wang Ling, et al. Preliminary study on the forecast model of daily maximum power load in Beijing area [J]. China Electric Power, 52(8).
[4] Li Tingshun, Wang Wei, Liu Zesan. Electric load GELM-WNN forecasting method considering uncertain intervals [J]. Computer Engineering, 2019(7):315-320.
[5] Wang Kejie, Zhang Rui. Research on short-term power load forecasting method based on improved BP neural network [J]. Electrical Measurement & Instrumentation, 2019, 56(24).

[6] Li Yuancheng, Fang Tingjian, Yu Erkeng. Research on support vector machine method for short-term load forecasting [J]. Proceedings of the Chinese Society of Electrical Engineering, 2003, 23(006): 55-59.

[7] Zhang Yajie. Monkey group algorithm and its application research [D]. Xidian University.

[8] Zhao R, Tang W. Monkey algorithm for global numerical optimization [J]. Journal of Uncertain Systems, 2008, 2(3):165-176.