Power load prediction method based on kernel extreme learning machine with t-distribution variation bat algorithm

Caijuan Qi¹, Qian Zhang², Xing Tian¹, Kun Zhang¹, Wei Tang³*

¹ State Grid Ningxia Electric Power Eco-Tech Research Institute, Yinchuan, Ningxia Province, 750004, China
² Beijing Kedong Electric Power Control System CO., LTD, Beijing, 100192, China
³ China Agricultural University, Beijing, 100083, China
*Corresponding author’s e-mail: wei_tang@cau.edu.cn

Abstract. To improve the accuracy of power system load forecasting, according to nonlinearity and uncertainty of load sequence, a power load forecasting method combined with kernel extreme learning machine (KELM) and adaptive variation bat optimization algorithm (AMBA) is proposed in this paper. The model determines mutation probability of the current optimal individual according to the population fitness variance and the value of the current optimal solution and the t-distribution variation is performed on the global optimal individual, and bat individual after mutation is subjected to secondary optimization. Then AMBA is used to optimize the network parameters of kernel function extreme learning machine, so that make full use of advantages of fast learning speed and generalization ability of KELM to realize rapid prediction of load. AMBA-KLEM model is used to predict load of Ningxia Power Grid, and the results are compared with ELM model and KELM model. The results show that the model has higher load prediction accuracy.

1. Introduction
With the development of the power market, the importance of load forecasting is becoming more and more obvious, and the requirements for load forecasting accuracy are getting higher and higher. The traditional prediction method is relatively mature, and the prediction result has certain reference value. However, to further improve the prediction accuracy, it is necessary to improve the traditional method [1]. At the same time, with the continuous advancement of modern science and technology, the theoretical research is gradually deepened, the emergence of emerging interdisciplinary theory, represented by grey theory, expert system theory, fuzzy mathematics, etc., also provides a solid theoretical basis and mathematical basis for the rapid development of load forecasting [2].

The extreme learning machine (ELM) is a new algorithm for the single hidden layer feedforward neural network (SLFN) [3]. This algorithm randomly assigns the connection weight of the input layer and the hidden layer node and the threshold of the hidden layer neurons of the network, and the whole training process does not need to iterate once, and only needs to adjust the number of neurons in the hidden layer. The only optimal solution can be obtained. However, since the weights and thresholds between the input layer and the hidden layer of the ELM are randomly generated, the stability of the network structure is poor to some extent [4]. The KELM algorithm uses the kernel function to map the
input samples to high-dimensional kernel space, which overcomes the random fluctuation of the model output caused by the random matrix $H$ in the ELM, simplifies the setting process of the number of hidden layer nodes, and makes the model have better generalization ability and rapidity [5]. In addition, the global optimization algorithm is easy to fall into the local optimal problem. At present, the main method to solve this problem is to increase the population size, but it can't solve the premature convergence problem fundamentally[6].

This paper presents a new prediction model (AMBA-KELM), kernel extreme learning machine based on adaptive mutation bat algorithm. Firstly, in order to improve generalization performance of extreme learning machine, solve potential problem caused by the least squares method and the randomness initialization of the hidden layer node threshold and the input layer weight, the kernel extreme learning machine model is adopted; Then, in view of the deficiency of bat algorithm that is prone to fall into local optimal, it can judge whether the bat algorithm falls into local optimal according to the population fitness variance and the value of the current optimal solution, then t-distribution variation on the optimal individual is proposed, which can effectively avoid premature convergence of bat algorithm. A load prediction model combined with the kernel extreme learning machine and adaptive mutation bat algorithm based on the variance of the population's fitness is established.

2. Kernel extreme learning machine

The kernel extreme learning machine is an algorithm derived from extreme learning machine theory [7]. KELM is a single-layer feed forward neural network algorithm. Compared with the basic ELM algorithm, its ability to solve regression prediction problems is stronger. Compared with the support vector machine algorithm, it has similar or better prediction accuracy and calculation speed is faster.

The basic extreme learning machine model can be expressed as:

$$f(x) = h(x)\beta$$

(1)

Where $h(x)$ is the hidden layer output matrix and $\beta$ is the output layer weight matrix.

ELM guarantees regression prediction accuracy by minimizing output error, shown in equation (2):

$$H\beta = T$$

(2)

Where $H$ is the hidden layer output matrix of extreme learning machine; $T$ is expected output vector.

The ELM algorithm guarantees the generalization ability of the neural network by minimizing the output weight $\beta$. Usually $\beta$ takes its least squares solution:

$$\hat{\beta} = H^+T = H^T(HH^T)^{-1}T = H^T\left(\frac{1}{C} + HH^T\right)^{-1}T$$

(3)

Where $H^+$ is the Moore-Penrose generalized inverse of the hidden layer output matrix $H$. Add the parameter $1/C$ to the main diagonal in the unit diagonal matrix $HH^T$, so that its eigenvalue is not zero, and then the weight vector $\hat{\beta}$ is obtained. This makes ELM more stable and its generalization is better.

For KELM algorithm, the neural network characteristic equation is the same as basic ELM, but kernel function is introduced to obtain better regression prediction accuracy, as shown in equation (4):

$$f(x) = h(x)H^T\left(\frac{1}{C} + HH^T\right)^{-1}T = \begin{bmatrix} K(x,x_1) \\ \vdots \\ K(x,x_N) \end{bmatrix} \left(\frac{1}{C} + \Omega_{ELM}\right)^{-1}T$$

$$\Omega_{ELM} = \exp(-\gamma \|x_i - x_j\|)$$

(4)

(5)

Where $\Omega_{ELM}$ is the selected kernel function, usually taking a Gaussian kernel function; the output weight of KELM model is:

$$\beta = \left(\frac{1}{C} + \Omega_{ELM}\right)^{-1}T$$

(6)

In the KELM model, the specific form of feature mapping function $h(x)$ of the hidden layer node is.
not specifically given, but only the specific form of the kernel function $K(x, x)$ needs to be known to find the value of output function. At the same time, because the kernel function directly adopts the form of inner product, it is not necessary to set the number of hidden layer nodes when solving the output function value, so that the initial weight and offset of the hidden layer need not be set, however, because parameter $\gamma$ of kernel function and parameter $C$ of generalized inverse matrix affect regression prediction performance, optimization solution is still needed.

3. Bat algorithm

3.1. Basic Bat Algorithm Principle
Bat algorithm (BA), as a newly proposed global optimization algorithm, imitates bats to search for prey[8]. The specific steps are as follows:

   Step 1: Initialization. Calculate fitness value and select the current optimal individual $x^*$ according to fitness value.

   Step 2: Update parameters of bat search according to equations (6), (7) and (8).

$$Q_i = Q_{\min} + (Q_{\max} - Q_{\min}) \rho$$

$$v_i^{t+1} = v_i^t + (x_i^t - x^*)Q_i$$

$$x_i^{t+1} = x_i^t + v_i^{t+1}$$

Where $\rho$ is a random number obeys [0,1] uniform distribution, $v_i^t$ and $x_i^t$ respectively represent the speed and position, $x^*$ represents the current global optimal position, $Q_i$ is bat individual's pulse frequency.

   Step 3: Generate a random number $\text{rand}1$, if $\text{rand}1 > r_0$, select the current optimal solution, then each bat individual is updated to the global optimal individual neighborhood.

   Step 4: Generate a random number $\text{rand}2$. If $\text{rand}2 < A$, and $f(x_i) < f(x^*)$, this new solution is accepted.

   Step 5: The fitness of each individual is evaluated, position and fitness of the optimal individual were recorded.

   Step 6: Check if the end conditions are met. If end condition is met, iteration is stopped and the global optimal individual is output; otherwise, the procedure returns to step 2.

3.2. Basic Principle of Adaptive Mutation Bat Optimization Algorithm (AMBA)
The bat algorithm is a group intelligent optimization algorithm that relies on the interaction and influence between bat individuals. Like other intelligent optimization algorithms, it has the problem of easy to fall into premature convergence. However, whether it is precocity or global convergence, the bats in the population will appear "aggregate". A large number of experiments have shown that the uniformity of individual positions is equivalent to the same fitness of individuals. Therefore, according to this characteristic of bat individuals, the problem of premature convergence of bat algorithm is solved. This paper proposes a bat optimization algorithm based on adaptive mutation of group fitness.

Let the number of bats in the population be $n$, $f_i$ is the fitness of the $i$-th bat individual, $f_{avg}$ is the current average fitness of the population, and $\sigma^2$ is the population fitness variance of the population, then $\sigma^2$ can be defined as

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (f_i - f_{avg})^2$$

Where

$$f_{avg} = \frac{1}{n} \sum_{i=1}^{n} f_i$$

The $\sigma^2$ represents the degree of "convergence" of bats in the population. The smaller the value, the more the population tends to converge; otherwise it is in the random search phase. Obviously, $\sigma^2$ is not possible to assume that the algorithm achieves global convergence only by the fact that $\sigma^2$ is equal to zero, and further judgment is needed. It can be seen that if $\sigma^2$ is zero, and this time is not the
In the theoretical optimal solution or the expected optimal solution \( f^* \), then the population falls into local optimum. If you want to solve this problem, you must provide a mechanism to jump out of the local optimal re-search when the algorithm occurs premature convergence.

According to equations (6), (7), and (8), during the evolution process, each bat can track the current optimal individual to update its position. Therefore, if the algorithm occurs premature convergence, the optimal individual at this time must be the local optimal solution, and the population loses its evolutionary ability. Combining equation (7), if a mutation operation is applied at this time to change the optimal individual at this time, the bat can be entered into other areas for searching. This is the basic idea of the adaptive mutation mechanism proposed in this paper.

Considering that the rest of the bats may find a better position when tracking the current optimal individual, this paper sets the mutation operation as a random operator, that is, the probability of \( p_m \) is satisfied by the probability of satisfying the variation condition by \( x^* \), and the calculation formula of \( p_m \) is as follows:

\[
    p_m = \begin{cases} 
    k \sigma^2 < \sigma_d^2, f(x^*) > f_d \\
          0 \quad \text{others}
    \end{cases}
\]

Where \( k \) can take any value between \([0.1, 0.3]\); the value of \( \sigma_d^2 \) is related to the actual problem, generally takes a maximum value smaller than \( \sigma^2_d \); \( f_d \) can be set to the theoretical optimal value.

For the global optimal individual \( x^* \) mutation operation, t-distribution variation is used in this paper.

\[
    x^* = x^* \times (1 + t(\eta))
\]

Where \( t(\eta) \) is t distribution with iteration number of algorithm as freedom degree parameter.

The t distribution, also known as the student distribution, contains the parameter degree of freedom \( \eta \), and its curve shape is related to size of \( \eta \). The smaller degree of freedom \( \eta \), the flatter the t distribution curve, the higher the tail of the curve is. When the degree of freedom \( \eta = 1 \), t distribution curve is a Cauchy distribution curve. That is \( t(\eta = 1) \rightarrow C(0,1) \), where \( C(0,1) \) is Cauchy distribution; the greater degree of freedom \( \eta \), the closer t distribution curve is to normal distribution curve, when degree of freedom \( \eta \rightarrow \infty \). The t distribution curve is approximated as a Gaussian distribution curve. That is, \( t(\eta \rightarrow \infty) \rightarrow N(0,1) \), where \( N(0,1) \) is a Gaussian distribution. That is, Gaussian variation and Cauchy variation are two special cases of t-distribution variation.

In initial stage of algorithm, Iteration value is small, t-distribution variation is similar to Cauchy mutation, and has good global search ability. Iteration value is large in the later stage of the algorithm operation, the t-distribution variation is similar to Gaussian variation, and has good local search ability; In the middle of run, the t-distribution variation is between Cauchy distribution variation and Gaussian variation. Therefore, t-distribution mutation strategy combines the advantages of both Gaussian distribution and Cauchy distribution to improve the global search and local search ability of algorithm.

4. Establishment of AMBA-KELM model

When ELM model solves the Moore-Penrose generalized inverse \( H^* = H^T (HH^T)^{-1} \), the existence of complex collinearity may lead to \( HH^T \) singularity. Each time the model is modeled by ELM, the matrix \( H^* \) obtained is different, so that solution is obtained. The hidden layer output weights \( \beta \) are also inconsistent. final effect of prediction results in the output of ELM model is prone to random fluctuations, and the stability and generalization ability are not ideal. In order to further enhance the generalization ability and stability of ELM, the kernel function is introduced into ELM by comparing the principle of ELM and support vector machine (SVM). The KELM model is adopted, and the newly proposed AMBA algorithm is used to optimize the model parameters of KELM. The specific steps are:

1) Determine the topology of the ELM. That is, the numbers of neurons in the input layer, the number of neurons in the hidden layer, and the number of neurons in the output layer;

2) Initialize bat algorithm parameters, including setting the population number 50, the initial position is \( C \) and \( \gamma \);

3) The sum of error squares of expected value and predicted value of training samples was used as
5. Example analysis

In order to verify validity and prediction accuracy of AMBA-KELM model. The experiment selected Ningxia daily data from 1 March to August 31, 2017 for 12 hours (a total of 184). The obtained 184-day data was processed into 177 sets of 8-dimensional data, and the first 7-dimensional data of each group was input as a model for the first 7 days, and the 8th-day data of the next day was output as a model. The first 115 groups were selected as model training data, and the last 62 groups were used as test data. The number of hidden layer nodes is the number of training samples, population size is 50, and the maximum number of training times in network is set to 100, using the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE) as a basis for evaluation.

ELM, KELM and AMBA-KELM models are used to predict and compare based on the same sample data in this paper, the predicted results and curve of relative error are shown in Figure 1. From Fig.1, it can be seen that the load forecasting curve of AMBA-KELM is closer to the actual load curve and the relative error always fluctuates at a lower level, the prediction curve of AMBA-KELM model may show a more similar trend to the actual load. Figure 2 compares the convergence rate of AMBA algorithm and BA algorithm, as we can see from it, AMBA has a faster convergence rate. This is because the introduction of mutation strategy in AMBA can not only avoid premature convergence of algorithm, but also significantly improve the global convergence performance of algorithm, so that the algorithm can quickly converge. According to Table 1, every error of the AMBA-KELM model mentioned in this paper are the smallest among the 3 models mentioned above, and the maximum and minimum relative errors are significantly lower than other two. In conclusion, the prediction model proposed in this paper has smaller error fluctuation, more accurate predicted results and faster convergence speed.
### Table 1. Forecasting errors comparison.

| Model      | ELM | KELM | AMBA-KELM |
|------------|-----|------|-----------|
| MAE        | 328.07 | 163.43 | 119.68 |
| MAPE/%     | 3.1968 | 2.5780 | 1.1619 |
| RMSE       | 486.00 | 245.89 | 150.08 |
| $M_{\text{max}}/%$ | 1.0649 | 2.0891 | 0.8623 |
| $M_{\text{min}}/%$ | 14.8661 | 9.4139 | 4.5665 |

Note: $M_{\text{min}}$ and $M_{\text{max}}$ are the minimum and maximum relative errors respectively.

---

**6. Conclusions**

This article has the following three main tasks:

- a. In view of bat algorithm is easy to fall into local optimum, a bat optimization algorithm based on population fitness variance adaptive t-distribution mutation is proposed, which only needs to mutate the global optimal individual.

- b. In order to improve the generalization performance of the extreme learning machine and solve the problem caused by the random initialization of the input weight and the hidden layer node offset, the kernel extreme learning machine model is introduced. In order to improve the prediction performance of the KELM model, this paper chooses AMBA. The algorithm optimizes parameter $\gamma$ of kernel function and parameter $C$ of generalized inverse matrix, and establishes AMBA-KELM model.

- c. The AMBA-KELM prediction model established is applied to power system load forecasting, and the model is verified by the example to have higher prediction accuracy.

---

**References**

[1] Dou, Z.H., Yang, R.G., Zou, L., et al. (2014) Prediction method of wind speed on wind farm using optimal sub-set regression model. Renewable Energy Resources, 8: 1161-1167.

[2] Wang, W.G., Shen, J., et al. (2017) Short-term load forecasting based on fuzzy mean generating function improved by back-stepping theory. Water Resources and Power, 35: 208-211.

[3] Wang, W.G., et al. (2018) Prediction method of wind speed on wind farm using optimal sub-set regression model. Electrical Measurement & Instrumentation, 55: 19-24.

[4] Zhang, B., Liu, W., Li, S.T., et al. (2019) Short-term load forecasting based on wavelet neural network with adaptive mutation bat optimization algorithm. IEEJ Transactions on Electrical and Electronic Engineering, 14: 376-382.

[5] Dou, Z.H., Yang, R.G., Jiao, J. (2013) Method of short-term load forecasting based on mean generating function-optimal subset regression. Transactions of the Chinese Society of Agricultural Engineering, 29: 178-184.

[6] Liu, W., Dou, Z.H., Zhang, B. (2018) Short-term load forecasting based on elastic net improved GMH and difference degree weighting optimization. Applied Sciences, 8: 1603-1624.

[7] Wang, Y.J., Li, D.W., Gao, C., Zhang, H.H. (2015) Short-term power load forecasting based on improved PSO-SVM. Electrical Measurement & Instrumentation, 52: 22-25.

[8] Yang, W., Mao, J.L., Xiong, Y. (2018) A dynamic weighted adaptive particle swarm optimization for typical functional problems. Electronic Science and Technology, 31: 9-12.