Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Short-term power load forecasting of GWO-KELM based on Kalman filter

Xiaoyu Chen. Yulin Wang. Jianyong Tuo*.

Beijing University of Chemical Technology, Beijing 100029
**corresponding author: tuojy@mail.buct.edu.cn*

Abstract: Short-term power load forecasting plays a significant role in power system security management. The prediction model in this paper is the grey wolf optimization algorithm to optimize kernel extreme learning machine (GWO-KELM). First, the Kalman filter is used to reduce the noise for the random noise interference existing in the power load data. Then determine the input and output of the prediction model. In this paper, the ELM model of three different kernel functions is used for comparative experiments, and the mean absolute percentage error is used as the evaluation model index. It is concluded from the experimental results that the GWO-KELM model used in this article has the advantages of high prediction accuracy and strong generalization ability, so it is practicable to apply the model to short-term electric load forecasting.

Keywords: short-term power load forecasting, Kalman filter, kernel extreme learning machine, grey wolf optimization, mean absolute percentage error.

1. INTRODUCTION

Short-term power load forecasting plays a key role in power system operation (Saadat et al., 2014). Accurately predicting short-term power load can effectively improve the utilization rate of power generation equipment, and improve the profitability of power system operation. However, the change of electric load is restricted by many factors, such as economy, temperature. There is a complex nonlinear relationship between these factors and the electric load, which makes the electric load forecasting complicated.

Currently, there are a variety of power load forecasting methods, such as neural networks (Chen et al., 2009), support vector machines (Ye et al., 2012), wavelet packet analysis (Bi et al., 2004) and so on. There are some disadvantages to predict a model using a single machine learning algorithm. The traditional neural network algorithm has slow learning speed and the possibility of network train failure. Support vector machine is difficult to process large-scale training samples. Wavelet analysis needs to select wavelet base in advance, and the quality of wavelet base affects the experimental results. After the extreme learning machine was proposed (Huang et al., 2004), it has been effectively applied to short-term power load forecasting due to its fast learning speed, good convergence effect and strong generalization ability (Zhang et al., 2017). Using the extreme learning machine to predict the short-term power load of the Indian power market (Dash et al., 2016), experiments have obtained good prediction results. ELM is widely used in the fields of classification regression, clustering (Huang et al., 2006), and this paper uses the ELM prediction model.

Based on the above research results, the prediction model of this paper is grey wolf-optimized kernel extreme learning machine based on Kalman filter (KF-GWO-KELM). Kalman filtering removes noise and interference in power data and obtains smoother and more accurate data values. Then the experimental comparison of ELM models with different activation functions. The parameter optimization of KELM determines the prediction performance, so parameter optimization is also an important content. In the application of bankruptcy prediction, the grey wolf optimization algorithm has been used to optimize the parameters of KELM (Wang et al., 2017). In this paper, the grey wolf optimization algorithm performs parameter optimization and applied to short-term power load forecasting.

The main work and results are as follows: a) Use the combined forecasting model KF-GWO-KELM for short-term power load forecasting. b) The model can predict the power load values for the next 24 hours, 72 hours and 168 hours. It is used to schedule daily and weekly schedules, so this experiment is meaningful. c) Experimental results show that the model has high prediction accuracy and strong generalization ability.

The rest of the paper is as follows: The second part is the basic theory, which introduces the extreme learning machine, the grey wolf optimization algorithm and the Kalman filter. The third part is the predictive model framework. The fourth part is the experiment and analysis of the prediction model. The fifth part is a summary of the article.

2. BASIC THEORY

2.1 Extreme learning machine with different activation functions

The extreme learning machine (ELM) is a single-layer feedforward neural network with a three-layer network structure, including input layer, hidden layer and output layer. The number of nodes in input layer is determined according to the input variables, n input variables, the number of nodes is set to n. The ELM algorithm uses randomly generated the input weight ω and the bias b of the hidden layer node, so the learning efficiency is high. The experiment shows that it runs fast and has a generalization ability. The formula for ELM is as follows:

$$H\beta = T \tag{1}$$

where H is the output matrix of the hidden layer, β is the output weight. The specific representation of H is as follows:

$$H = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) \dots g(\omega_L \cdot x_1 + b_L) \\ \vdots \\ g(\omega_1 \cdot x_N + b_1) \dots g(\omega_L \cdot x_N + b_L) \end{bmatrix} \tag{2}$$

where N is dimensional space, L is hidden layer nodes, g is activation function, b is bias of randomly assigned, ω is the input weight vector.

And T represents the output matrix. The formula is as follows:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} \dots t_{1m} \\ \vdots \\ t_{N1} \dots t_{Nm} \end{bmatrix} \tag{3}$$

When training the model, first randomly assign the values of ω and b of ELM to complete the initialization. The hidden layer output matrix H can be uniquely determined. T is a known matrix. The output weight β is obtained by the formula:

$$\beta = H^{+}T \tag{4}$$

where T is the training data target matrix, H^{+} is the Moore-Penrose generalized inverse of the matrix H. H matrix depends on the choice of the activation function. Choosing different activation functions will affect the experimental results, and Table 1 shows three activation functions. The experiments compared the ELM models with different activation functions. ELM uses the default sigmoid activation function, ReLU-ELM uses the rectified linear unit activation function, and KELM uses the radial basis function.

Table1: The number of hidden layer nodes is different

Activation Function	Formula
Sigmoid Function	$y = \frac{1}{(1+e^{-x})}$
Rectified Linear Unit	$y = max(0, x)$
Radial Basis Function	$K(x, x_i) = exp(-\gamma\|\|x - x_i\|\|)$

In the neural network, the activation function of the hidden layer needs to be selected according to the characteristics of the data itself, and the optimal function is selected through experiments. Among them, KELM replaces random mapping. Then the output function of the KELM can be succinctly written as:

$$F(x) = h\beta = h(x)H^T\left(\frac{I}{C} + HH^T\right)^{-1} T$$

$$= \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{I}{C} + \Omega_{ELM}\right)^{-1} T \tag{5}$$

KELM performance is determined by the kernel function γ and the penalty parameter C. The parameters play an important role in the stability and generalization performance of the model. In this paper, GWO is used to optimize the parameters.

2.2 Grey Wolf Optimization Algorithm

The grey wolf optimization algorithm is a metaheuristic algorithm that simulates the wolf hunting mode, which was proposed in 2014 (Mirjalili et al., 2014). It has two social modes: one is hierarchical. The grey wolf has four ranks, and the top-down is α, β, δ and ω. The highest level α is the leader wolf. The second is the hunting. The algorithm has been widely used in the fields of parameter optimization and image classification since its introduction (Nurul et al., 2017).

In this paper, the grey wolf algorithm is used to optimize the parameters C and γ of the kernel extreme learning machine. The optimized parameters represent the prey and the wolf group hunting represents the process of dynamic optimization in the algorithm (Chen et al. 2019).

First step is that the wolves change their position depends on the prey position. These formulas are expressed as:

$$\vec{D} = \left|\vec{C} \cdot \vec{X}_p(n) - \vec{X}(n)\right| \tag{6}$$

$$\vec{X}(n+1) = \vec{X}_p(n) - \vec{A} \cdot \vec{D} \tag{7}$$

where n is the number of iterations, $\overrightarrow{X_p}$ is the position vector of the prey, $\vec{X}$ is the position vector of the grey wolf, $\vec{A}$ and $\vec{C}$ are coefficient vectors. Vector $\vec{A}$ and vector $\vec{C}$ are as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{8}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{9}$$

where component of $\vec{a}$ is linearly reduced from 2 to 0 during the iteration, and $\overrightarrow{r_1}$, $\overrightarrow{r_2}$ are random vectors in [0,1].

Second step is in the process of dynamically searching for prey, the grey wolf with smallest distance from the prey is always α, which indicates the best position for hunting, followed by β and δ, and the lowest ω changes its hunting position according to the first three levels of wolves. Expressed as:

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|, \vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|,$$

$$\vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right| \tag{10}$$

$$\vec{X}_1 = \vec{X}_\alpha - \overrightarrow{A_1} \cdot (\vec{D}_\alpha), \vec{X}_2 = \vec{X}_\beta - \overrightarrow{A_2} \cdot (\vec{D}_\beta),$$

$$\vec{X}_3 = \vec{X}_\delta - \overrightarrow{A_3} \cdot (\vec{D}_\delta) \tag{11}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{12}$$

Finally, the GWO algorithm terminates by completing the maximum number of iterations. The return value of α is the optimization result of parameters C and γ.

2.3 Kalman Filter

Kalman filtering is an algorithm that uses the linear system state equation to estimate the state of the system by the input and output observation data. It can remove the noises and disturbances in the system. It is currently a filtering method widely used in the field of communication and control.

The Kalman filter algorithm can be divided into two steps: prediction and update. The prediction equation estimates the prior value of the state variable and the prior estimate of the error covariance according to the state estimation value and input at the previous moment. The formula is:

$$\hat{x}_{\bar{k}} = A\hat{x}_{k-1} + Bu_{k-1} \tag{13}$$

where $\hat{x}_{k-1}$ is posterior state estimates at k-1, $\hat{x}_{\bar{k}}$ is the a priori state estimate at time k, u_{k-1} is the amount of input of the system at time k-1, A and B are the state transition matrix.

$$P_{\bar{k}} = AP_{k-1}A^T + Q \tag{14}$$

whereP_{k-1} is posteriori estimation covariance at k-1, $P_{\bar{k}}$ is the a priori estimate covariance at time k, Q is the covariance of the system process.

The update equation is responsible for achieving an improved posteriori estimate utilizing the a priori estimate and the new measured variable. The state update formula is:

$$K_k = \frac{P_{\bar{k}}H^T}{HP_{\bar{k}}H^T + R} \tag{15}$$

where K_k is the Kalman gain, H is the state variable to the observed transformation matrix, R is the measurement noise covariance.

$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k(z_k - H\hat{x}_{\bar{k}}) \tag{16}$$

$$P_k = (I - K_kH)P_{\bar{k}} \tag{17}$$

where z_k is the filter input, $\hat{x}_k$ and P_k are posteriori estimates of the a priori value corrected by the Kalman gain at time k.

The data collection will inevitably be disturbed by external noise, and the Kalman filter method is used for the power load data set. According to the prediction and update of the algorithm, the original data is processed to approximate the real data.

3. PREDICTION MODEL

3.1 Data source

The experimental data set used in this paper is the power load data set provided by the 9th Electrician Mathematical Contest in Modeling. The data set includes power load data for two regions from January 1, 2009 to January 10, 2015, and meteorological factor data for January 1, 2012 to January 17, 2015. The sampling interval of power load data is 15 minutes, 96 points per day, and the dimension is MW. Meteorological data includes daily maximum temperature, daily minimum temperature, daily average temperature, daily relative humidity, and daily rainfall information.

3.2 Model input and output

The model input variables include power load data and meteorological data, as follows:

(1) Power load data at the same time of the previous week, expressed as $L_1(t)$, $L_2(t)$, $L_3(t)$, $L_4(t)$, $L_5(t)$, $L_6(t)$, $L_7(t)$.

(2) Power load data at the first two moments of the previous day, expressed as $L_1(t-1)$, $L_1(t-2)$.

(3) Input the previous day's maximum temperature, minimum temperature and average temperature

(4) Input the relative humidity of the previous day.

(5) Precipitation on the previous day.

Model output variable is a column of electrical load data.14 input variables and 1 output power load data variable. The experimental samples in this paper are the electrical load data and meteorological data from January 1, 2013 to March 31, 2013. Including 3 months of weather and electricity data. Based on the model input, the model data set is determined to be 15 * 7968. According to predicting the future 24h, 72h, and 168h power data, that is, predicting future 96, 288, and 672 power load data, different numbers of input sets and test sets are determined.

3.3 GWO-KELM model

The steps of the GWO-KELM algorithm are as follows:

Firstly, select proper power load data and pre-process. The Kalman filter is used to filter the noise of the historical load value, and logarithm of filtered data. The weather data is processed by normalization method. The data are then divided into a training set and a test set. With the training set, ELM, ReLU-ELM, KELM, GWO-KELM models are used to determine the optimal parameters for modeling.

Secondly, in the model using the grey wolf optimization algorithm, the optimal C and γ are set as target prey. The initial grey wolf population and parameters were initialized, and the next generation grey wolf adjusted the position according to the target prey. Keeping the three best wolves and updating each wolf's position until the maximum number of iterations.

Thirdly, test the experimental model and return the MAPE to compare the performance of the model.

Figure 1 is the flow chart of power load forecasting in this paper. Simply speaking, it is to determine the power load data and pre-process it. The data are divided into training set and test set. The training set determines the optimal parameters for model. and the test set is used to evaluate the model with the MAPE index.

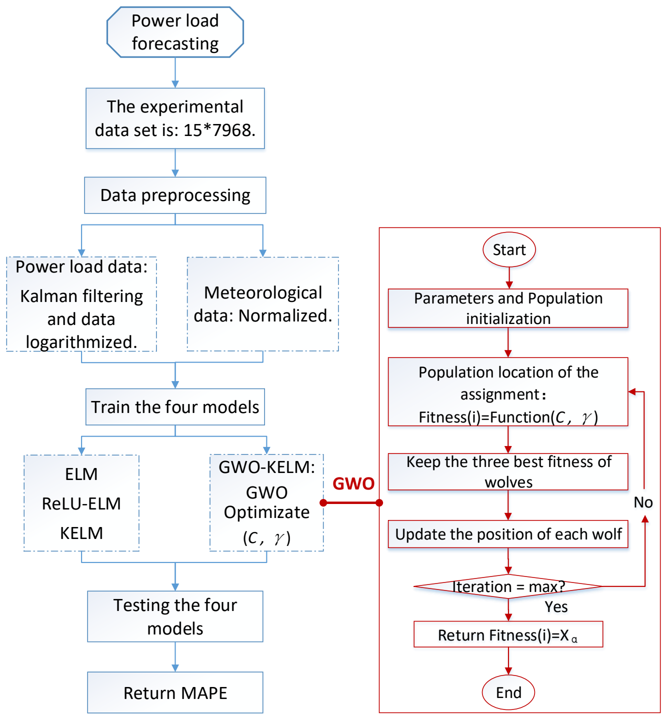

Fig.1 Power load forecasting algorithm flow chart

4. EXPERIMENTS AND ANALYSIS

This part describes the details of electric load forecasting experiment and the analysis based on the results.

4.1 Experimental analysis index

The experimental platform of the short-term power load forecasting model is Python 3.5.

In the power load forecasting, the average absolute percentage error is usually used as an evaluation index. The smaller the value, the higher the prediction accuracy of the model. The formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \qquad (18)$$

where n is the sample size, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

4.2 Experimental results and analysis

The following experimental results are evaluated using MAPE. It takes about 2 ~ 5 seconds to run the model once, and the model runs fast.

4.2.1 Determining the number of hidden layer neurons

For the number of hidden neuron nodes, the number of nodes is determined empirically or experimentally. The number of hidden layer nodes in this experiment is 20~35, and the optimal value is selected by experiment.

From Table2, the pros and cons of the model are closely related to the number of nodes in the hidden layer, and a suitable experimental model is established by selecting the appropriate number of hidden layer nodes. The optimal number of nodes for ELM is 20, and the rest of the model is 30.

Table2: The number of hidden layer nodes is different

Model / Nodes	ELM	ReLU-ELM	KELM	GWO-KELM
	MAPE	MAPE	MAPE	MAPE
20	11.593	7.629	12.927	4.826
30	15.099	4.627	9.925	3.979
35	14.031	4.771	26.558	4.687

4.2.2 Compare normalized and unnormalized

The data set is the electrical load data and the meteorological data, which belong to different eigenvalues and affect the accuracy of the prediction. Therefore, the dimension between the various features needs to be removed, and the common method is normalization. According to the characteristics of filtered power load data, logarithmic method is adopted for normalization processing, and the data range after normalization is [3, 4]. The formula is as follows:

$$L' = \log_{10} L \qquad (19)$$

The normalization of temperature, humidity and rainfall is to scale the value of the feature between 0 and 1, as shown below:

$$x_{norm}^{i} = \frac{x^i - x_{min}}{x_{max} - x_{min}} \qquad (20)$$

Table 3: Prediction results with Unnormalized data.

Model / Time/h	ELM	ReLU-ELM	KELM	GWO-KELM
	MAPE	MAPE	MAPE	MAPE
24	6.372	9.356	9.277	7.490
72	6.177	14.022	9.343	3.469
168	6.345	6.903	8.819	2.592
Ave	6.298	10.093	9.146	4.517

Table 4: Prediction results with normalized data.

Model / Time/h	ELM	ReLU-ELM	KELM	GWO-KELM
	MAPE	MAPE	MAPE	MAPE
24	5.594	4.478	3.023	2.890
72	4.338	3.866	5.302	3.485
168	3.531	5.953	2.352	1.123
Ave	4.487	4.765	3.559	2.499

4.2.3 ELM Models comparison with different activation functions

The following is a comparison of the predicted and actual power load values of the four models. Figures 2, 3, and 4 are predicted power load data for the next 24 hours, 72 hours, and

168 hours. The better the fitting effect of the curve in the experimental graph with the actual value, the better the performance of the prediction model. As can be seen from the figure, the predicted value of the GWO-KELM model is close to the actual value, and GWO-KELM model is more accurate than the ELM, ReLU-ELM and KELM models.

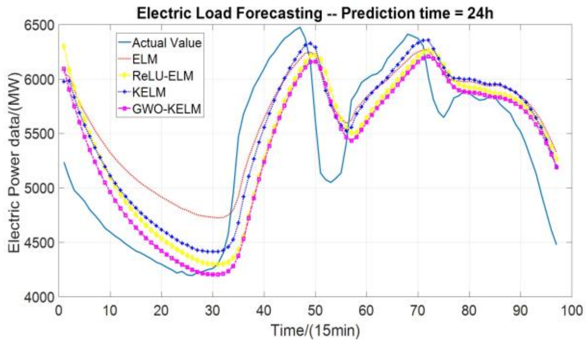

Fig.2. Predicted power load data for the next 24 hours.

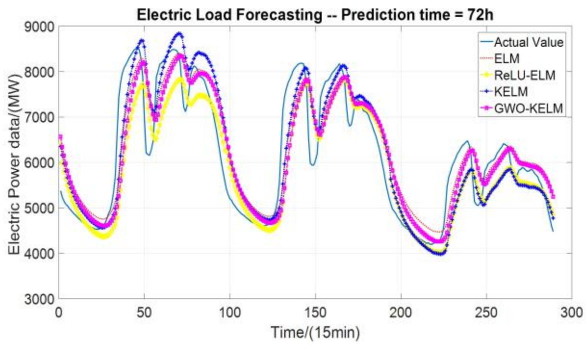

Fig.3. Predicted power load data for the next 72 hours.

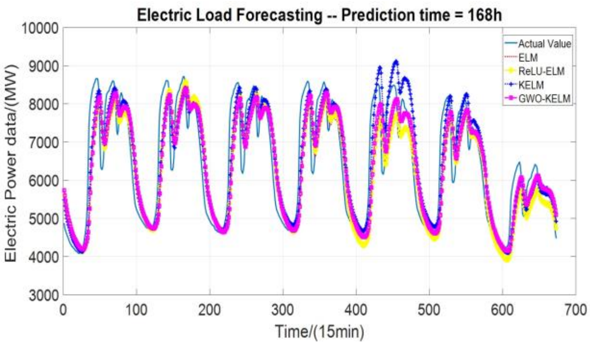

Fig.4. Predicted power load data for the next 168 hours.

5. CONCLUSIONS

In this paper, a power load prediction model based on the Kalman filter GWO-KELM algorithm is used, and the weather factors are considered. The algorithm uses Kalman filter to remove the noise, and uses the grey wolf optimization algorithm to optimize the kernel function parameter γ and penalty function C of the kernel extreme learning machine. The performance of the model was verified by comparison with other conventional methods. The experimental results show that the model can achieve better accuracy in short-term power load forecasting in 24 hours, 72 hours, and 168 hours.

In this work, we mainly consider the impact of weather on the electrical load. In fact, the input of the model can also select some known knowledges, such as working days, holidays, these factors can make the forecast more accurate, which will be done in the following study.

REFERENCES

Saadat B., Hooshmand R. A. and Parastegari M. (2014). Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. Energy 72, page 434-442.

Chen H., Lin Y., Chen Y., Chang C. and Hwang R. (2009). Short Term Power Load Forecasting by Using Neural Models. 2009 4th International Conference on Innovative Computing, Information and Control, ICICIC, page 1212-1215.

Ye N., Liu Y. and Wang Y. (2012). Short-term power load forecasting based on SVM. World Automation Congress (WAC), IEEE.

Bi Y., Zhao J. and Zhang D. (2004). Power load forecasting algorithm based on wavelet packet analysis. Power System Technology, 2004. PowerCon 2004. 2004 International Conference on, 1, page 987-990.

Huang G. B., Zhu Q. Y. and Siew C. K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE International Conference on Neural Networks - Conference Proceedings, 2(2), page 985-990.

Zhang W., Hua H. and Cao J. (2017). Short Term Load Forecasting Based on IGSA-ELM Algorithm. 2017 IEEE International Conference on Energy Internet (ICEI). IEEE, page 296-301.

Dash S K, Patel D. (2016). Short-term electric load forecasting using Extreme Learning Machine - a case study of Indian power market. IEEE Power, Communication & Information Technology Conference, IEEE.

Wang M., Chen H., Li H., Cai Z., Zhao X., Tong C., Li J. and Xu X. (2017). Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction. Engineering Applications of Artificial Intelligence, 63, page 54-68.

Huang G. B., Zhu Q. Y. and Siew C. K. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70(1-3), page 489-501.

Mirjalili S., Mirjalili S. M. and Lewis A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69, page 46-61.

Nurul Asyikin Z. and Zuriani M. (2017). Developing a Gold Price Predictive Analysis using Grey Wolf Optimizer. Research & Development. IEEE.

Chen X., Tuo J. and Wang Y. (2019). A Prediction Method for Blood Glucose Based on Grey Wolf optimization Evolving Kernel Extreme Learning Machine. 2019 Chinese Control Conference (CCC), Guangzhou, China, page 3000-3005.