Energy Consumption Forecast Based on Coupling PSO-GPR

Xinli Wang[1], Shijian Liu[1], Liping Yan[1] and Ning Wang[2]

[1]School of Economy and Management, North China Electric Power University, 071000, Baoding

[2]HeBei Construction & Investment Group CO. Ltd, 050051, shijiazhuang

Abstract. Consumption for energy has a great influence on sustainable development of economy and society. The scientific and rational energy development strategy can effectively guarantee social and economic stability and orderly development and energy development strategy is inseparable from the right to accurately predict energy demand and analysis. In order to improve the accuracy of prediction of energy consumption, hybrid forecasting model for energy consumption based on Gaussian process (GPR) and Particle Swarm (PSO) is proposed. Firstly, PSO algorithm is employed to optimize two parameters in covariance function, and then the initial value of parameters are obtained; next time series are trained in GPR model for energy consumption. Under the Bayesian framework, parameters in covariance function can be optimizing again. Finally, we can forecast energy consumption in well-trained GPR model, and the results can be compared with the auto-regressive integrated moving average model and exponential smoothing models. The results show that the proposed model has good stability and high prediction accuracy. It is suitable to be applied in forecasting consumption for energy.

Introduction

Energy consumption forecasting is a complex nonlinear system, which is affected by many factors. In the prediction of energy consumption, domestic and foreign scholars have established different econometric models, such as the grey prediction [1], DGEM. Chen Hongtao, Zhou Dequn built China's energy consumption GM(1,1) prediction model [2]. Liang Na and Zhang Jigang first use grey prediction to establish the prediction model, and then use the RBF neural network to estimate the parameters of the gray prediction model [3]. Li Hongmei, He Changzheng, Xiao Jin use Log and GMDH model to predict medium and long-term energy consumption of China [4]. However, due to the impact of many factors, there is a complex relationship between these factors. Therefore, it is difficult to predict the energy consumption by using the causal model or the structure proportion relationship, and it may also affect the accuracy of the prediction results. If the standard of the prediction model is to pursuit the maximizing prediction accuracy, the time series model will be the best choice [5]. Because the single prediction model is limited by the assumptions and the scope of application, the results are not ideal. Therefore, some scholars have combined the model, the combination forecasting model can effectively make up for the shortcomings of single model and improve the accuracy of prediction [6]. Based on the artificial neural network model, grey model and time series model, Yi Xue established the combination model to forecast the energy consumption of China in the next 6 years [7]. Liu Aiqin used the combination model of exponential smoothing and ARIMA to forecast energy consumption. Gauss process regression is a new machine learning technique based on Bayesian theory and statistical learning theory [8].

Gauss Process Regression(GPR)

GPR is a powerful probabilistic modeling tool. It can be used to define the prior distribution of the hidden function in a hierarchical Bayesian model, then inference can be performed directly on the function space by estimating the posterior distribution. Compared with the neural network and support vector machine, Gauss process regression has the advantages of easy realization, super parameter adaptive acquisition and output probability meaning. It is easy to combine with predictive

control, adaptive control, Bias filtering, etc. this method has developed rapidly in foreign countries, and has made a lot of research results. Firstly, GPR establishes the prior function of the model in the form of probability distribution, and then completes the transformation of the transcendental function to the posterior function based on the Bayesian framework, at the same time, the "super parameter" of the kernel function can be calculated, however, the choice of kernel function of SVM is usually based on empirical value or cross validation. The training process based on the training data is the process of selecting the parameters, which is one of the advantages of the Gauss model. The Bias principle provides a convincing and unified framework for computing, makes the it easy to calculate the Gauss regression model with the Gauss noise, and the model can be adjusted flexibly. Among them, $m(x)$ and $k(x, x'|\theta)$ represent the mean and covariance functions, ϕ and θ are the parameters of observation model and covariance function respectively. The reason why the covariance function plays an important role in the Gauss process regression is that it can reflect the priori assumption of the hidden function. GP is a set of arbitrary random variables with joint Gauss distribution, whose properties are determined by mean function and covariance function. GPR can choose different covariance functions, such as squared exponential(SE), rational quadeatic (RQ) and Matern covariance function. In this paper, we use the (SE):

$$k(x_i, x_j|\theta) = \sigma_{se}^2 \exp(\sum_{k=1}^{d} \frac{(x_{ik} - x_{jk})^2}{2l_k^2}) \tag{1}$$

$\theta = \{\sigma_{se}^2, l_1 \cdots l_k\}$ is a vector that contains all the parameters, σ_{se}^2 is the measurement parameter of total variance of the hidden function, l is scale parameter, which is used to control the degree of reduction of the correlation with the increase of the input dimension.

Taking the noise into account, we can establish a general model of Gauss process:

$$y_i = f(x_i) + \varepsilon_i, i = 1, \cdots, n \tag{2}$$

$\varepsilon \sim N(0, \sigma^2)$ and $f(x)$ obeys the Gauss process prior distribution. Assume that the hidden function is the mapping of the new input variable $\overline{x}$ to the predicted value $\overline{f}$, then the joint Gauss distribution between hidden variables f and $\overline{f}$ is:

$$\begin{bmatrix} y \\ \overline{f} \end{bmatrix} \sim N(0, \begin{bmatrix} K_{f,f} + \sigma^2 I & K_{f,\overline{f}} \\ K_{\overline{f},f} & K_{\overline{f},\overline{f}} \end{bmatrix}) \tag{3}$$

$K_{f,\overline{f}} = k(x, \overline{x}|\theta)$ and $K_{\overline{f},\overline{f}} = k(\overline{x}, \overline{x}|\theta)$

The resulting conditional distribution:

$$\overline{f}|y, x, \overline{x} \sim N(m(\overline{x}|\theta), \text{cov}(\overline{f})) \tag{4}$$

Among them, the mean function is

$$m(\overline{x}|\theta) = k(x, x|\theta)(K_{f,f} + \sigma_n^2 I)^{-1} y \tag{5}$$

The covariance function is

$$\text{cov}(\overline{f}) = k(\overline{x}, \overline{x}'|\theta) - k(\overline{x}, x|\theta)(K_{f,f} + \sigma_n^2 I)^{-1} k(x, \overline{x}'|\theta) \tag{6}$$

Particle Swarm Optimization(PSO)

In PSO, the entity is abstracted as a particle, and the initial value of the system is a group of random particles (a set of random solutions), the optimal solution is sought by iteration. The position of the particle is the solution to the problem, and the particle searches the optimal particle in the solution space, therefore, the best solution is to find the location of the updated particle, that is, how to update the particle position, in order to make the algorithm faster and better convergence to the optimal solution . If the current position of the particle i is better than its historical optimal position, it is updated to p_i. In addition, the historical optimal position for the entire group is $p_g = (p_{g1}, p_{g2}, \cdots, p_{gd})$. Then the j dimension value $(1 \le j \le d)$ of the i particle is updated by the

following two formulas:

$$v_{ij} = v_{ij} + c_1 \cdot rand() \cdot (p_{ij} - x_{ij}) + c_2 \cdot rand() \cdot (p_{gj} - x_{gj}) \tag{7}$$

$$x_{ij} = x_{ij} + v_{ij} \tag{8}$$

Among them, c_1, c_2 are nonnegative constants, which make particles have the ability to self summarize and learn from the best individuals in the group, so that the particles can move closer to their own historical optimal location and the global optimal position. $rand()$ is a random number between [0,1]. In addition, the speed of particle movement is limited between $[v_{min}, v_{max}]$. The PSO process is as follows:

Step 1, Initialization: generate the position and velocity of particles randomly in D dimensional space; Step 2, Position evaluation: use a constructed position objective function to evaluate each particle; Step 3: Update the historical optimal position of the particle and the global optimal position: compared with the optimal value of the historical position, if the position evaluation value is better than the historical optimal value of the particle, the current position is used to replace the historical optimal position of the particle; compared with the global optimal value, if the current evaluation value is better than the global optimal value of the group, then the global optimal value is replaced by the current value. Step 4, Update the position and velocity of the particles by the above formula; Step 5, Loop termination condition: perform second to fourth steps for each particle loop, until the loop termination condition is satisfied. The loop termination condition is similar to that of the genetic algorithm, which is an iterative number or a good fitness value.

Construction of PSO-GPR Combination Forecasting Model

The prediction process of this model is divided into three stages: in the first stage, according to the Gauss process regression theory, a general model for the regression of the process of the Gauss process is established, and uses the PSO to optimize two parameters of the covariance function σ_{se}^2 and l in order to get the initial value of the parameter which is close to the optimal value; In the second stage, we use the GPR model of the first stage and the initial value of the parameters to train the data, at the same time, the hyper parameters are optimized by using Bayesian theory. After the first stage particle swarm optimization, the initial value of the parameter is close to the optimal value, so the optimization process can be completed quickly and is not easy to fall into local optimal value; in the third stage, the new data are input into the GPR model to get the final prediction results.

Empirical Analysis

Forecast data and sample design. According to the time sequence of energy consumption in Hebei province from 1980 to 2012 as the sample data, the 2009-2013 energy consumption data as test data of the model of this paper to verify the accuracy of the prediction results for inspection. the fork represents the original time series, that is, the annual energy consumption, the green curve represents the predicted energy consumption.

Data preprocessing. In order to eliminate the influence of the data dimension on the prediction results, the input variables (years) and output variables (energy consumption) are normalized, each value in the sample data is in the range of [0,1]. First of all, we must determine the maximum and minimum values of the sample data , then use the following formula to normalize the data so that the data can be distributed uniformly in the effective data space.

$$z_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{9}$$

y_{max} and y_{min} represent the maximum and minimum values of the sample data respectively. At the end of the prediction process, the predicted value is normalized, so it is necessary to carry out the anti normalization operation to get the true value of the prediction.

Evaluating indicator. It is very important to select the reasonable evaluation index to evaluate the prediction effect. In this paper, the following two indicators are used to test the prediction results:

$$\sigma_{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\overline{y}(i) - y(i)\right|}{y(i)} \qquad (10)$$

$$\sigma_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\overline{y}(i) - y(i))^2} \qquad (11)$$

σ_{MAPE} is mean absolute percentage error; σ_{RMSE} is root-mean-square error; N is the sample number; $\overline{y}(i)$ is predicted value; $y(i)$ is true value. Energy consumption forecast based on PSO-GPR model. In this paper, the energy consumption of Hebei province from 1980 to 2008 was used as the data of PSO-GPR model and GPR model, and the corresponding GPR prediction model was obtained by the regression analysis of the process of the Gauss process. In the prediction model of Gauss process, the most important parameters are the two parameters of the covariance function σ_{se}^2 and l. In the GPR prediction model, these two parameters are set to be 5 and 10 respectively, and then use the Bayesian theory to optimize the parameters in the training process. However, the optimal value of the two parameters of the covariance function is far away from the initial value (5, 10). So in the combination forecasting model, the PSO algorithm is used to optimize the two parameters of the covariance function σ_{se}^2 and l, so that obtains the optimal initial value (46, 52) with respect to the initial value of the GPR prediction model, then, using Bayesian theory to optimize the parameters again, which will improve the accuracy of prediction. The kernel function of Gauss process regression is SE, which is the square exponential covariance function, In the PSO algorithm, the particle range is [1,100], population size is set to 50, the maximum number of iterations is 100.Because the fitness function of the PSO optimization algorithm used in this paper is to minimize the fitness function, so the root mean square error σ_{RMSE} is used as fitness function (see formula 12) . The results are shown in table 1.

Table 1 Predicted values of different models

Year	Energy consumption (10000 tons of standard coal) True value	PSO-GPR Predicted value	GPR Predicted value	ARIMA Predicted value	ES Predicted value
2009	25419	26247	25377	25067	25390
2010	27531	27404	27196	25821	26433
2011	29498	28208	29099	26583	27475
2012	30250	28631	31086	27354	28518
2013	29664	28668	33159	28133	29561

Using the evaluation index to analyze the error of energy consumption forecast in table 1, the results are shown in Table 2. It can be seen that the error of the combined forecasting model based on PSO-GPR is minimum, the average absolute error is 1093.49, the root mean square error is 3.29%, and the prediction accuracy of the model is improved significantly.

Table 2 Forecasting error analysis

Error	PSO-GPR	GPR	ARIMA	ES
RMSE	1093.49	1817.94	2110.87	1289.24
MAPE	3.29%	5.87%	6.44%	3.41%

Conclusion

Firstly, PSO algorithm has the characteristics of few parameters and easy to implement, and it can improve the generalization ability of Gauss process by using particle swarm optimization of the parameters of the Gauss process. The application of Coupling PSO-GPR in energy consumption time series prediction is of probability significance, which can meet the need of prediction.

Secondly, different kernel functions have different effects on machine learning and prediction ability, but the difference is small. In this paper, the SE kernel function is selected. The empirical results show that PSO has better prediction accuracy after optimizing GPR kernel function parameters. After the particle swarm optimization of Gauss process, the problem of the initial and local optimization and the determination of the parameters of the model are solved, therefore, the combination forecasting model based on PSO-GPR has a good application value in energy consumption forecast.

Acknowledgment

References

[1]Diyar Akay, Mehmet Atak. Grey Prediction with Rolling Mechanism for Electricity Demand Forecasting of Turkey [J]. Energy, 2007(32): 1670-1675.

[2]Yi-Shian Lee, Lee-Ing Tong. Forecasting Energy Consumption Using a Grey Model Improved by Incorporating Genetic Programming[J].Energy Conversion and Management,2011(52): 147-152.

[3]Xu Chong, Liu Baoguo, Liu Kaiyun, Guo Wei. An intelligent prediction model for landslide displacement time series analysis based on particle swarm coupled with regression method [J]. Rock and Soil Mechanics, 2011(6):1672-1674.

[4]He Yue, He Zhenglin, Zhou Xinyin. Combination warning systemof solid customer churn based on Data Mining[J].Soft Science,2012(1):130-131.

[5]Chen Hongtao, Zhou Ming. Study on energy consumption prediction based on GM(1,1) model[J].Mining Research and Development,2007,(7):77- 79.

[6]Liang Na, Zhang J. Prediction of energy consumption in China based on grey RBF network[J].Journal of Jiamusi University: Natural Science Edition, 2008, (2): 224- 226.

[7]Li Hongmei, He Changzheng, Xiao Jin. Long term prediction of energy consumption in China based on Log GMDH model [J]. Soft Science, 2012 (5): 51.

[8]Pandifu, Liu Hui, Li Yanfei. An optimal model for wind speed forecasting based on time series analysis and filtering algorithm [J]. Power System Technology, 2008 (32): 82-86.