Short-term power load forecasting based on combined kernel Gaussian process hybrid model

Liang Lingyu[1*], Wenqi Huang[2], Zhaojie Dong[3], Jiguang Zhao[4], Peng Li[5], Bingfang Lu[6]and Xinde Zhu[7]

[1234567]China Southern Power Grid Digital Grid Research Institute Co., Ltd., Guangzhou 510663

Abstract. As one of the countries with the most energy consumption in the world, electricity accounts for a large proportion of the energy supply in our country. According to the national basic policy of energy conservation and emission reduction, it is urgent to realize the intelligent distribution and management of electricity by prediction. Due to the complex nature of electricity load sequences, the traditional model predicts poor results. As a kernel-based machine learning model, Gaussian Process Mixing (GPM) has high predictive accuracy, can multi-modal prediction and output confidence intervals. However, the traditional GPM often uses a single kernel function, and the prediction effect is not optimal. Therefore, this paper will combine a variety of existing kernel to build a new kernel, and use it for load sequence prediction. In the electricity load prediction experiments, the prediction characteristics of the load sequences are first analyzed, and then the prediction is made based on the optimal hybrid kernel function constructed by GPM and compared with the traditional prediction model. The results show that the GPM based on the hybrid kernel is not only superior to the single kernel GPM but also superior to some traditional prediction models such as ridge regression, kernel regression and GP.

1Introduction

Power load forecasting is conducive to the intelligentization of fuel procurement, equipment maintenance and load distribution. For example, in the urban power system, power load forecasting is very important for power system management and energy trading [1-2]. Since the electric power process is a complex dynamic system, the prediction difficulty is relatively high, and the prediction effect of using manual observation and trend analysis methods is poor [3]. In recent years, after continuous exploration, domestic and foreign scholars have proposed a series of effective load intelligent forecasting methods, such as time series method, artificial neural network (ANN) and support vector machine (SVM) forecasting.

In 2000, Tresp first proposed the Gaussian Process Mixture (GPM) model [4]. Using the "divide and conquer" strategy, samples were divided into several groups, and each sample group was assigned a Gaussian Process (GP) model for learning prediction. It not only has better predictive ability, but also can output confidence interval [5]. Therefore, this paper combines single cores to construct a new kernel function on this basis, and selects the optimal combined kernel function for the load sequence through experiments, and then achieves the optimal prediction effect.

2Principle of Gaussian Process Mixture Model

Learning algorithm of Gaussian process mixture model

This paper adopts the iterative learning algorithm of hidden variable posterior hard partition proposed by Chen [6]. Compared with the traditional MCMC, VB or EM learning algorithm of the GPM model, the hard-partition iterative learning algorithm uses a sampling approximation strategy. In step E, the learning samples are allocated according to the maximum posterior probability criterion, and each step is estimated by the maximum likelihood method in step M. The undetermined parameters of the GP component greatly reduce the computational complexity of the algorithm. The specific implementation steps of the algorithm are as follows:

The first step: For a given learning sample, divide it into several groups by k-means clustering algorithm;

Step 2: Independent learning of each GP component participating in the mixing based on maximum likelihood estimation;

The third step: According to the maximum posterior probability criterion, re-designate the group of the learning sample. If the re-designated result is consistent with the previous round, the iterative algorithm stops and outputs the final result; otherwise, it returns to the second step.

*Corresponding author's e-mail: liangly1@csg.cn

Step 4: After the learning process is over, for a given test sample, if the corresponding target output is predicted, the group can also be specified according to the maximum posterior probability criterion. Then the test samples are assigned to the first group, and the prediction distribution can be obtained from the prediction formula of the single GP component. The required learning sample in this predictive formula is the learning sample assigned to the group in the last iteration.

3Prediction Algorithm of Combined Kernel Gaussian Process Mixture Model

GPM is a mixture of multiple relatively independent GPs, and each GP processes its corresponding sample component. There is a single Gaussian process with noise, and its expression is:

$$g(x) \sim N\left(m(x), k(x,x') + \sigma_v^2 \delta_{pq}\right) \tag{1}$$

Where is the mean function $m(x)$, $k(x,x') + \sigma_v^2 \delta_{pq}$ It is a kernel function. And for the Kronecker function δ_{pq} Say, when $p = q$, $\delta_{pq} = 1$, No $\delta_{pq} = 0$. In many cases, the Gaussian process $g(x)$ The mean function of can be normalized to $m(x) = 0$, and $k(x,x')$ You can choose different functions instead. In the kernel function selection, the predecessors have done some basic research, and the commonly used kernel functions are square exponential functions. (SE), Rational quadratic function (RQ) and Matern function (Ma) [7-10]. among them SEIs the most commonly used kernel function, And it has the best effect on infinitely differentiable time series forecasting, and it has high requirements for time series smoothness. RQIt is another commonly used core. Its advantage is that after the sequence phase space is reconstructed, as the delay increases, RQThe forecast effect is relatively stable. MaIt is a highly versatile core, and there are three common forms. Adjust the parameters to adapt to the sequence of different degrees of smoothness, but the parameters are not properly selected, based on MaKernel function GPM Even loss of predictive ability. The three kinds of function expressions are shown in (2) to (6).

1） Square exponential function (SE):

$$k_{SE}(x,x') = \sigma_f^2 \exp\left[-\frac{1}{2l^2}\|x-x'\|^2\right] \tag{2}$$

2） Rational quadratic function (RQ)

$$k_{RQ}(x,x') = \sigma_f^2\left[\left(1+\frac{1}{2\alpha l^2}\|x-x'\|^2\right)\right]^{-\alpha} \tag{3}$$

3）Matern function (Ma):

$D = 1$:

$$k_{Ma}(x,x') = \sigma_f^2 \exp\left[-\frac{1}{l}\sqrt{D\|x-x'\|^2}\right] \tag{4}$$

$D = 3$:

$$k_{Ma}(x,x') = \sigma_f^2\left[1+\frac{1}{l}\sqrt{D\|x-x'\|^2}\right]\exp\left[-\frac{1}{l}\sqrt{D\|x-x'\|^2}\right] \tag{5}$$

$D = 5$:

$$k_{Ma}(x,x') = \sigma_f^2\left[1+\frac{1}{l}\sqrt{D\|x-x'\|^2}+\frac{1}{l}\sqrt{D\|x-x'\|^2}\right]\exp\left[-\frac{1}{l}\sqrt{D\|x-x'\|^2}\right] \tag{6}$$

After the above three functions are transformed into matrix form, all their eigenvalues are not less than zero, that is, the above three functions are all positive semi-definite functions. According to Mercer's theorem, any positive semi-definite function can be used as the kernel function, so SE, RQ, and Ma can all be used as the kernel function of the GPM model. In addition, according to its combination and addition, all satisfy the conditions of Mercer's theorem, and it can also be used as the kernel function of GPM. The combined kernel functions used in this paper are formulas(7) to(10).

$$k_{SE+RQ}(x,x') = k_{SE}(x,x') + k_{RQ}(x,x') \tag{7}$$

$$k_{RQ+Ma}(x,x') = k_{RQ}(x,x') + k_{Ma}(x,x') \tag{8}$$

$$k_{Ma+SE}(x,x') = k_{Ma}(x,x') + k_{SE}(x,x') \tag{9}$$

$$k_{Ma+SE+RQ}(x,x') = k_{Ma}(x,x') + k_{SE}(x,x') + k_{RQ}(x,x') \tag{10}$$

Equations (1) to (6) are the variance of the kernel function, which controls the local correlation of input variables; is the feature width, which controls the smoothness of the model. Can make $\boldsymbol{\theta} = \{l, \sigma_f^2, \sigma_v^2\}$ It is a vector composed of undetermined hyperparameters contained in the GPM model, and its value needs to be determined during model learning.

4 Power load time series analysis

Analysis of power load sequence forecast characteristics

Let's start with the four aspects of autocorrelation function, partial autocorrelation function, maximum Lyapunov exponent and saturated correlation dimension, and analyze the characteristics of load series forecasting in depth.

(1) Autocorrelation function

The autocorrelation function describes in detail the dependence of a certain moment in the sequence on another moment. By setting the fixed time delay parameter, the correlation degree between the initial time and any time within the time delay range can be obtained. Now the autocorrelation function is calculated for the electric load sequence, and the maximum time delay is 200. The power load sequence in Figure 1 starts from the 20th time delay, and the value has fallen below the confidence interval, which proves that it is a set of nonlinear sequences.

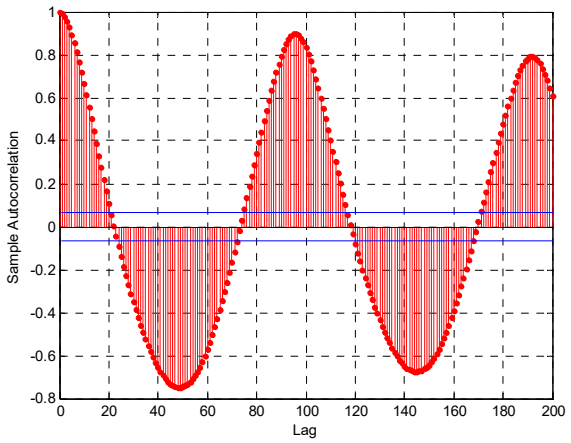


Figure 1 Autocorrelation function of power load sequence

(2) Partial autocorrelation function

The partial autocorrelation function is a good indicator of the stationarity of the time series. Now the partial autocorrelation function is calculated for the electric load sequence, and the maximum time delay is 100. The result is shown in Figure 2. It can be seen from the figure that there is a single large peak at time t+1, and the values after time t+5 mostly converge within the confidence interval, which proves that the power load sequence is non-stationary.

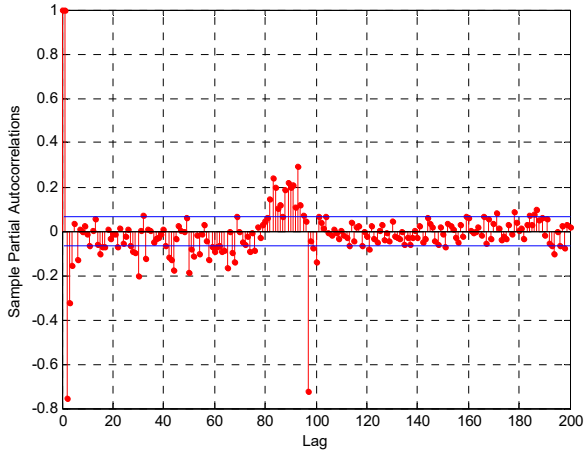


Figure 2 Partial autocorrelation function of power load sequence

(3) The largest Lyapunov exponent

Lyapunov exponent can well reflect the chaotic characteristics of time series. In this paper, the Wolf method loop is used to obtain the maximum Lyapunov exponent of the load sequence under 20 cycles. Since the minimum unit of the electric load sequence is 15 minutes, each cycle of the electric load is set to 3 hours. The result is shown in Figure 3. It can be seen from the figure that the maximum Lyapunov exponents of the series are all greater than zero, verifying that the load series have certain chaotic characteristics.

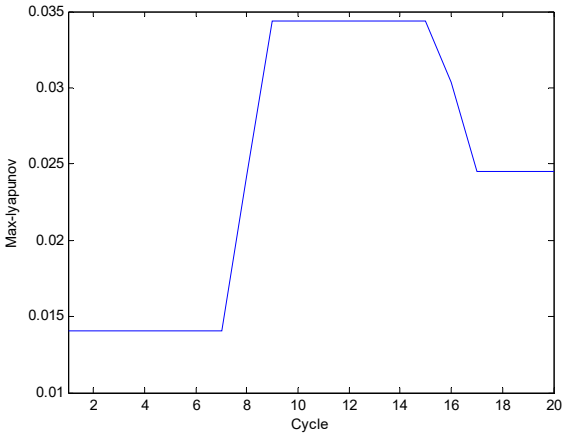


Figure 3 The largest Lyapunov exponent of the power load sequence

(4) Saturated correlation dimension

According to the correlation dimension, it can be distinguished whether the time series has random or chaotic characteristics. Now find the correlation dimension for the load sequence, and the embedding dimension is taken from 2 to 8. The result is shown in Figure 4. It can be seen from the figure that the correlation dimension of the sequence is saturated with the increase of the embedding dimension, so it is verified that the power load sequence has certain chaotic characteristics.

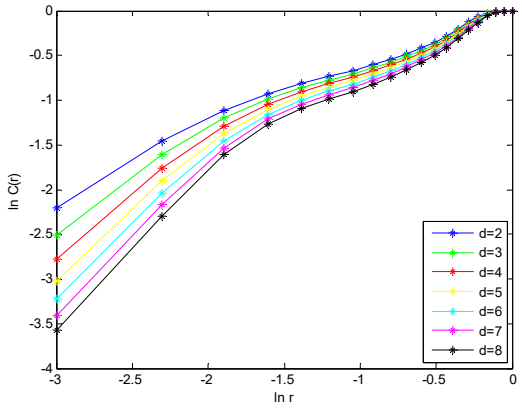


Figure 4 Saturation correlation dimension diagram of power load sequence

5 Forecast examples and results analysis and comparison

The two load sequences in this article are now predicted, and the prediction experiment is carried out on a Lenovo laptop. CPU is Intel Core (TM) i5-2430m, 2.40GHz,

memory is 4GB, software platform is matlab 2010a. In the power load sequence prediction experiment, the learning samples are from the 201st time to the 500th time, and the test samples are from the 501st time to the 800th time.

In order to fully demonstrate the improvement effect of the proposed combined kernel function on the GPM model and the advantages of GPM over the traditional forecasting model, this article first predicts the power load together with three single kernel functions and multiple combined kernel functions under the same experimental parameters, that is, the final prediction The kernel function used is: SE, RQ, Ma, SE+RQ, SE+Ma, RQ+Ma and RQ+SE+Ma; Then GPM and traditional models are used to predict load under common parameters. The traditional models involved in the comparison are Kernel-Regression (K-R), Ridge-Regression (R-R) and GP models. Among them, K-R is a kernel-based regression prediction model. By adjusting the optimal window width, the prediction result with the smallest error can be obtained gradually. R-R is a biased estimation regression model, through improved least squares estimation method, to obtain more reliable prediction results. As the basis of GPM model, GP has been widely used in various forecasts. The following two indicators are used to quantitatively evaluate the pros and cons of the prediction results:

（1） Root mean square error ($RMSE$):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[y_p(i)-y_t(i)\right]^2} \tag{11}$$

（2） Determination coefficient (R^2)

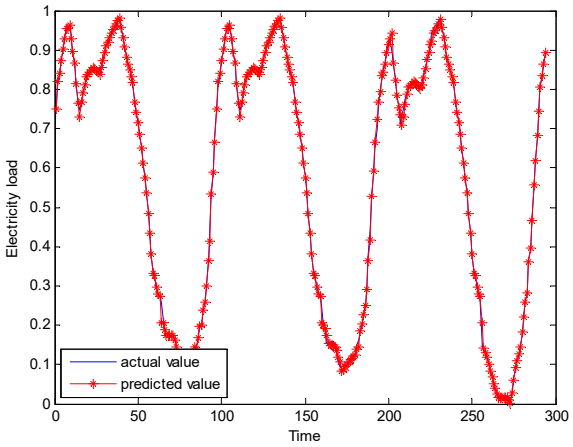
$$R^2 = 1 - SSR/SST \tag{12}$$

$$SSR = \sum_{i=1}^{n}[y_p(i)-y_t(i)]^2 \tag{13}$$

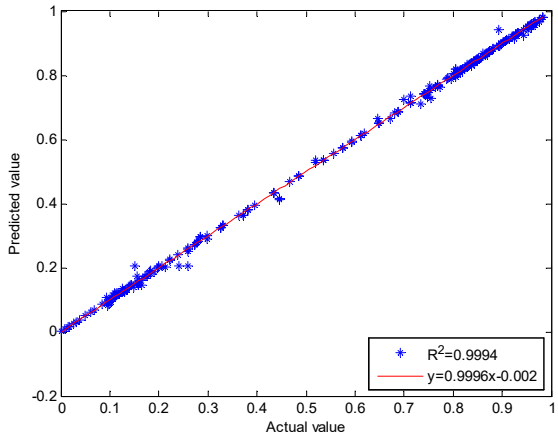
$$SST = \sum_{i=1}^{n}\left[y_p(i)-y_m\right]^2 \tag{14}$$

n equations (11) to (14), $y_p(i)$ Forecast the results for the time series, $y_t(i)$ Is the time series sample to be predicted, y_m Is the mean of the time series samples to be predicted. According to the different meanings represented by the two indicators, R^2 $RMSE$ The larger and smaller the prediction effect, the better. In the phase space reconstruction link, because the traditional mutual information method and pseudo-neighborhood method are relatively time-consuming, this paper uses a grid traversal search to obtain the optimal parameters. In the grid search for the optimal parameters, the optimal discriminating criteria are equations (11) and (14) $RMSE$ and R^2 . Is satisfied $RMSE$ The smallest, Simultaneously R^2 biggest d and τ , Is the optimal parameter.

In the GPM model, the main undetermined parameter is the number of modes, plus the phase space reconstruction parameters and, there are three parameters that need to be determined. The best prediction parameter obtained through grid traversal search is d =6, τ =1, $C=3$. After setting all parameters, predict the electric load sequence. The blue line in Figure 5(a) is the true value of the sequence, and the red line is the predicted value obtained. Therefore, the higher the fit of the blue and red double lines, the better the prediction effect. The abscissa of the dot diagram in Figure 5(b) represents the true value of the humidity sequence, and the ordinate represents the predicted value. The more blue dots in the figure are concentrated on the main diagonal, the better the prediction effect.



(a) Comparison curve between the real value and the predicted value of the sequence



(b) Dot plot of the real and predicted values of the sequence

Figure 5 GPM model prediction results of power load

6Conclusion

In this paper, Gaussian Process Mixture (GPM) model is used for power load forecasting, and its real load sequence is used for forecasting experiments. GPM prediction uses an iterative learning algorithm for hard partitioning of hidden variables posterior, which improves the prediction efficiency of the model. GPM is a kernel-based machine learning model. When the kernel function changes, the prediction effect will also change, and when the sequence distribution characteristics are complex, the combined kernel function is more

comprehensive than a single kernel. Therefore, in this paper, we have conducted an in-depth study in the direction of kernel functions, combining three common single kernel functions (SE, RQ, and Ma) to form a new combined kernel function, and verifying its improvement effect in experiments. The following conclusions can be drawn through experiments:

(1) Power load has strong nonlinearity, non-stationarity, chaotic characteristics and certain short-term predictability;

(2) For the phase space reconstruction ameter embedding dimension and time delay τ, generally speaking, with the increase and decrease, the prediction accuracy wll gradually increase, but the value of should not be too large;

(3) There is no obvious rule for the number of modes of GPM. But for the power load sequence in this paper, it is largely affected by the three periods of morning, middle and evening. Through the traversal search, the optimal number of modes in prediction is;

(4) In ower load forecasting, the SE+RQ+Ma combined kernel is the best forecasting effect, and the GPM forecasting effect based on the optimal kernel function is better than traditional forecasting models. Therefore, the selection of the combined kernel function largely depends on the sequence. When the sequence is different, the selection of the combined kernel will change, but the prediction effect is better than that of a single kernel.

Finally, GPM can adaptively select the optimal combination kernel function when facing different load sequences

References

1. Marie Bessec, Julien Fouquau, Short-run electricity load forecasting with combinations of stationary wavelet transforms, In European Journal of Operational Research, Volume 264, Issue 1, 2018, Pages 149-164.
2. Hisashi Takeda, Yoshiyasu Tamura, Seisho Sato, Using the ensemble Kalman filter for electricity load forecasting and analysis, In Energy, Volume 104, 2016, Pages 184-198.
3. Yang Liu, Wei Wang, Noradin Ghadimi, Electricity load forecasting by an improved forecast engine for building level consumers, In Energy, Volume 139, 2017, Pages 18-30.
4. Trest V. Mixtures of Gaussian processes[C]//Advances in Neural Information Processing Systems 13. Cambridge: MIT Press,2000:654-660.
5. ZHOU Ya-tong,CHEN Zi-yi,MA Jin-wen. From Gaussian Process to the Mixture of Gaussian Processes: A Surver[J].Journal of Sigal Processing, 2016,(08):960-972.
6. Chen Ziyi, Ma Jinwen, Zhou Yatong. Hard-cut EM algorithm for mixtures of Gaussian Processes. Lecture Notes in Computer Science, 10th Int. Conf. on Intelligent Computing Methodologies (ICIC), August 3-6, Taiyuan, 2014, 2: 68-75
7. Xu Chong, Liu Baoguo, Liu Kaiyun, et al. Slope Angle Intelligent Design Based on Gaussian Process With Combinatorial Kernal Function[J]. Rock and Soil Mechanics, 2010,31(3):821-826.
8. Huang Fang, Liu Yuanjun, Chen Bo. Prediction Model of Network Traffic Based on Combined Kernel Function Gaussian Regression[J]. Computer Engineering and Applications, 2015,51(19):93-97.
9. Zong Wenting, Wei Zhinong, Sun Guoqiang, et al. Short-term Load Interval Prediction Based on Improved Gaussian Process Regression Model[J]. Proceedings of the CSU-EPSA, 2017, 29(8):22-28.
10. Sun Bin, Yao Haitao, Liu Ting. Short-term Wind Speed Forecasting Based on Gaussian Process Regression Model[J]. Proceedings of the CSEE, 2012, 32(29):104-109.