Multiple kernel support vector machine short-term load forecasting based on multi-source heterogeneous integration of load factors

Jun Gao[1], Meng Nie[1], Ying Zhen[1], Qianhong Wu[2,a], Yanda Wu[1], Pengli Qiao[1] and Pei Li[1]

[1]State Grid Shandong electric power company Heze power supply company, Heze, 274000, China
[2]Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education, Shanghai Jiao Tong University, Shanghai, 200240, China

Abstract. In this paper we present a novel method developed from multiple kernel function for short term load forecasting to integrate multi-source heterogeneous load factors in big data. Nine kinds of load factors were selected as multi-source heterogeneous factors. In the proposed method, three algorithms (the sample distribution method, single variable method and rank space diversity method) were adopted to establish the optimal multiple kernel functions to integrate the load factors. Experimental results show that the average relative error of multiple kernel SVM is smaller than single kernel SVM, and the accuracy of multiple kernel SVM model based on double layer multiple kernel learning algorithm and l_p norm is the highest. Therefore, multiple kernel SVM can deal with the multi-source heterogeneous data in the load forecasting effectively, and the speed and accuracy of load forecasting can be improved by parallel processing.

Keywords: multiple kernel function; load forecasting; multi-source heterogeneous;load factors; big data; SVM model; parallel processing.

1 Introduction

High accuracy of load forecasting is important for the safe and stable operation of power system.In recent years, big data has been widely concerned, smart grid is seen as one of the important technologies of big data applications in the energy field [1].With the development of smart grid, communication network technology and sensor technology [2], factors that affect the load forecasting have more quantity and varieties, more complex distribution characteristics. Therefore, the traditional load forecasting methods needs to be improved .

Big data in smart grid is decentralized, diverse, complex and of high value[3], and the three former characteristics make the big data multi-source heterogeneous. Factors that affect the load forecast including history load data, meteorological data (temperature, rainfall, atmospheric pressure, relative humidity, wind direction, wind speed, etc), geographic data, holidays, the price of electricity, and so on.The different construction time, research and development unit, technology adoption and specific business requirements of these factors' stroage system resultingin different data storage mode, data types and update frequency[4]. Further more, the factors have different physical meaning, dimension

---

a Corresponding author : mailtowqh@163.com

and statistical characteristics, so the factorsare multi-source heterogeneous. Load forecasting based on the multi-source heterogeneous load factors is of great significance to improve the accuracy of prediction.

Under the environment of big data, the existing load forecasting methods such as time seriesmethod [5], regression analysis[6-7], neural network[8-9], grey prediction[10], fuzzy theory[11-12]and support vector machine(SVM)[13-14]have limitations in arithmetic speed and multi-source heterogeneous load factors.In terms of arithmetic speed, when the data volume is very large, locally weighted linear regression method requires to search the nearest k points for each sample point and the single computer operation time may cost days; neural network and fuzzy theory also need longer training time.In terms of multi-source heterogeneous loadfactors, classical time series method considers less load factors, mostly only using historical load data to predict. Most SVM used in load forecasting is based on single kernel. It puts all features into a vector, which is the input of a single kernel function with fixed form and parameters. This way appears simple and inefficient when load factors are derived from different data sources and have different kinds of characteristic. Some literatures have done some research to improve the speed of load forecasting. Wang *et.al*[2]proposed a parallel load forecasting method based on random forest algorithm and shorten the time of load forecasting. Zhang *et.al* [15] parallelized locally weighted linear regression model based on a cloud computing platform and improve the prediction speed. The literatures are mainly concentrated on improving speed of the load forecasting, rarely considering multi-source heterogeneous characteristics of the load factors, so this needs deep exploration.

This paper is organized as follows: Section 2 describes multiple kernel (MK) SVM model; Section 3 proposes an optimal multi kernel function selection method and a parallel MK SVM load prediction algorithm flow; Section 4 demonstrates the proposed method on short term load forecasting platform based on Hadoop cluster. Finally, conclusions of our paper are provided in the last section.

2 Multiple kernel SVM prediction model

2.1 Multiple kernel functions

Traditional single kernel SVM has been widely applied to load forecasting in power system [16], but it's not accurate enough when dealing with multi-source heterogeneous load factors. MK learning is to create the corresponding kernel function for each type of data, then combine the kernel functions based learning methods, which can achieve integrative analysis of heterogeneous data and a better prediction.

MK functions have two basic forms: linear and nonlinear[17]. Linear MK function is as follows :

$$K_{MKL}(x, y) = \sum_{h=1}^{m} \lambda_h K_h(x, y) \tag{1}$$

where $K_h(x, y)$ is single kernel function, $\lambda_h \in [0,1]$ is kernel weight, m is the number of single kernel functions. And nonlinear MK function is as follows:

$$K_{MKL}(x, y) = \lambda_1 K_1(x, y) \times \left(\lambda_2 K_2(x, y) + \lambda_3 K_3(x, y)\right) + \lambda_4 K_4(x, y) + \ldots\ldots \tag{2}$$

The commonly used single kernel functions include linear, polynomial, RBF and sigmoid. There are two options to choose single kernel function, one is to select a kernel function with different parameters, the other is to select different kinds of kernel function for different types of data. RBF is a local kernel function, the data points which are close to the sample point has great influence on the value of the kernel function, resulting in better learning ability and poor generalization performance[18].Sigmoid is a global kernel function, the data points that are far away from the sample

point can influence the value of the kernel function, resulting in better generalization performance and poor learning ability[18].The curves are shown in figure 1.The MK function combined with local and global kernel functions with different ability of deduction can foster strengths and circumvent weaknesses, and achieve a SVM prediction model with better performance[19].
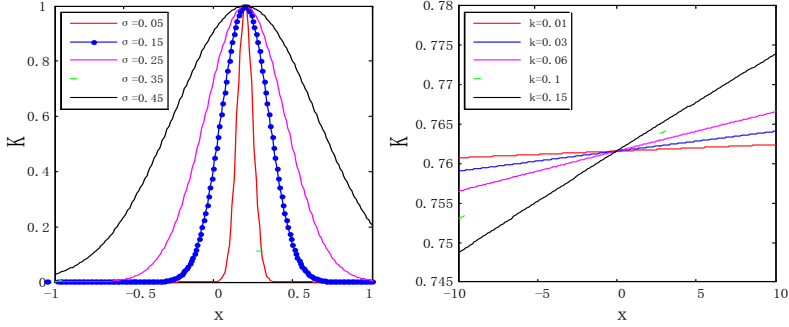
Figure 1. Local kernel function RBF and global kernel function Sigmoid

When the single kernel function of traditional SVM prediction model is replaced by a MK function, the model becomes a MK learning model. If the nonlinear MK function is adopted, the relevant optimization problem is not convex and it's difficult to get the local optimal solution [19], so linear MK function is used in this paper.

2.2 MK-SVM model based on l_1 and $l_p$norm

The MK-SVM model based on l_1 and l_p $(1 < p < +\infty)$ normare as follows[17,20], respectively:

$$\min_{\lambda} \quad \max_{\alpha,\hat{\alpha}} \sum_{i=1}^{N} y_i (\alpha_i - \hat{\alpha}_i) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \hat{\alpha}_i) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) K_{MKL1}(x_i, x_j)$$

$$s.t. \quad \sum_{i=1}^{N} (\alpha_i - \hat{\alpha}_i) = 0$$

$$0 \leq \alpha_i, \hat{\alpha}_i \leq C, i = 1, 2, ..., N$$

$$\lambda_h \geq 0, h = 1, 2, ..., m$$

$$\|\lambda\|_1 = 1$$

(3)

$$\min_{\lambda} \quad \max_{\alpha,\hat{\alpha}} \sum_{i=1}^{N} y_i (\alpha_i - \hat{\alpha}_i) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \hat{\alpha}_i) - \left\| \frac{1}{2} \sum_{h=1}^{m} \lambda_h \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) K_{MKLp}(x_i, x_j) \right\|_{p^*}$$

$$s.t. \quad \sum_{i=1}^{N} (\alpha_i - \hat{\alpha}_i) = 0$$

$$0 \leq \alpha_i, \hat{\alpha}_i \leq C, i = 1, 2, ..., N$$

$$\lambda_h \geq 0, h = 1, 2, ..., m$$

(4)

where $K_{MKL1}(x_i, x_j) = \sum_{h=1}^{m} \lambda_h K_h(x_i, x_j)$, $K_{MKLp}(x_i, x_j) = \sum_{h=1}^{m} \lambda_h K_h(x_i, x_j)$, $p^* = p/(p-1)$; x and y are input and output of training samples, respectively; N and m are the number of training samples and single kernel function of MK, respectively; λ is kernel weights; ε is accuracy; α and $\hat{\alpha}$ are lagrangian multipliers; C is penalty factor.

The final regression estimation functions based on is l_1 and l_p $(1 < p < +\infty)$ norm are:

$$f(x) = \sum_{i=1}^{N} \left(\alpha_i^* - \hat{\alpha}_i^* \right) K_{MKL1}\left(x_i, x \right) + b \tag{5}$$

$$f(x) = \sum_{i=1}^{N} \left(\alpha_i^* - \hat{\alpha}_i^* \right) K_{MKLp}\left(x_i, x \right) + b \tag{6}$$

where $b \in R$ is threshold value, the rest parameters have the same meaning with (3) and (4).

2.3 MK learning optimization algorithm

There are two ways of computing weights of the kernel function and the Lagrange multipliers of MK-SVM prediction model.

2.3.1 Single layer MK learning

The weights of the kernel function λ are given, then adopt the conventional algorithm to train SVM and get Lagrange multipliers α , $\hat{\alpha}$. The simplest way is the average allocation method, which gives the same weight to each kernel function, but the prediction is not better:

$$K_{MKL} = \frac{1}{m} \sum_{h=1}^{m} K_h\left(x, y \right) \tag{7}$$

2.3.2 Double layer MK learning

One layer is to optimize kernel weight λ , the other layer is to optimize Lagrange multipliers α , $\hat{\alpha}$, the procedure is conducted alternatively until the stop criterion is meet. The optimal kernel weight λ^* and Lagrange multipliers α^* , $\hat{\alpha}^*$ are obtained.

3 Parallel MK-SVM load forecasting

3.1 The constitution of MK function for multi-source heterogeneous load factors

The constitution of MK function needs to be determined when MK-SVM is used to estimate the regression function. This paper adopts sample distribution method, single variable method and rank space diversity method to establish the optimal MK function to integrate multi-source heterogeneous load factors, as shown in figure 2. Assuming the set of multi-source heterogeneous load factors is $\{V_d\}$, $d=1,2,...Q$, $V_d = \left[v_1, v_2, ..., v_T \right] \in R^{1 \times T}$. The procedure is as follows:

(1) Sample distribution method: the characteristic of each type of load factor V_d is depicted in the corresponding coordinate, then select the corresponding optimal single kernel function $K_{d1}\left(v_i, v_j \right)$ according to the distribution characteristics of the samples. Establish the MK function:

$$K_1\left(v_i, v_j \right) = \sum_{d=1}^{Q} K_{d1}\left(v_i, v_j \right) \tag{8}$$

(2) Single variable method: compare the prediction results, choose the single kernel function with the highest prediction accuracy $K_{d2}\left(v_i, v_j \right)$ as a part of MK function. Establish the MK function:

$$K_2\left(v_i,v_j\right)=\sum_{d=1}^{Q}K_{d2}\left(v_i,v_j\right) \tag{9}$$

If the constitution of $K_1\left(v_i,v_j\right)$ and $K_2\left(v_i,v_j\right)$ is the same, then $K_1\left(v_i,v_j\right)$ is the optimal MK function K_{MKL}, otherwise, weighting method is used to form a new MK function:

$$K_3\left(v_i,v_j\right)=aK_1\left(v_i,v_j\right)+bK_2\left(v_i,v_j\right)=a\sum_{d=1}^{Q}K_{d1}\left(v_i,v_j\right)+b\sum_{d=1}^{Q}K_{d2}\left(v_i,v_j\right)@\sum_{d=1}^{Q}K_{d3}\left(v_i,v_j\right) \tag{10}$$

where a and b are weight coefficients, $a+b=1$, $a>0$, $b>0$.
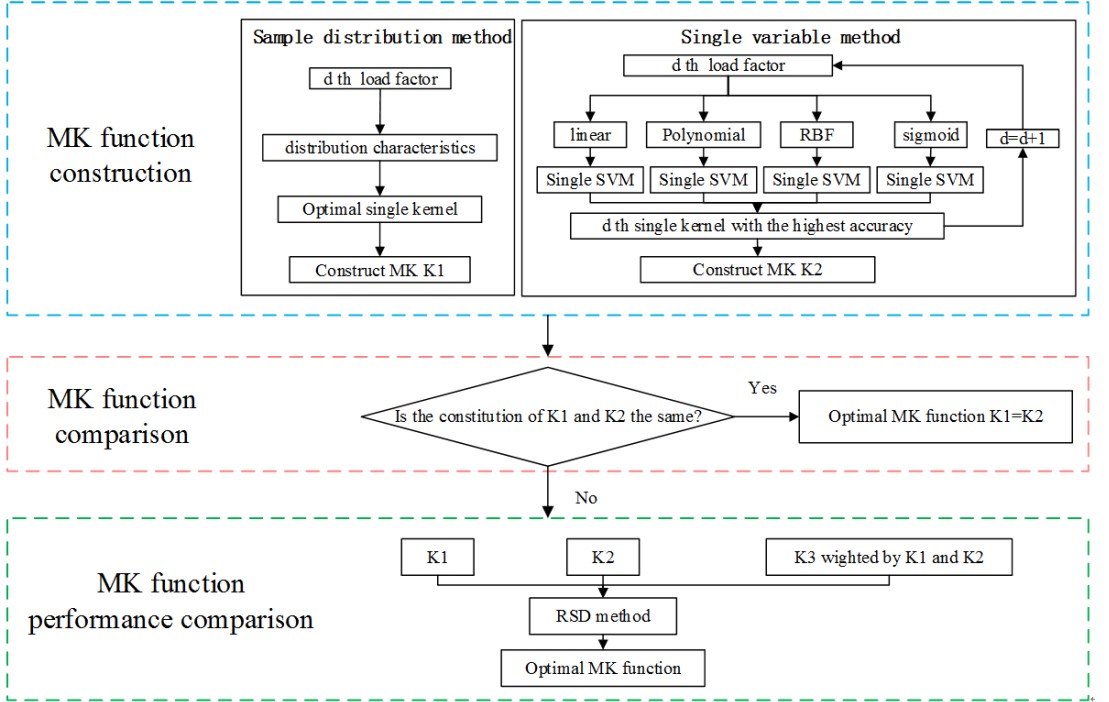
Figure 2.The diagram of selecting multiple kernel function

(3) Rank space diversity method [21]: calculate kernel matrix and *RSD* of $K_{d1}\left(v_i,v_j\right)$, $K_{d2}\left(v_i,v_j\right)$ and $K_{d3}\left(v_i,v_j\right)$ respectively:

$$A_{d,s}=\begin{bmatrix} K_{d,s}\left(v_1,v_1\right) & \cdots & K_{d,s}\left(v_1,v_T\right) \\ \vdots & \ddots & \vdots \\ K_{d,s}\left(v_T,v_1\right) & \cdots & K_{d,s}\left(v_T,v_T\right) \end{bmatrix} \tag{11}$$

$$RSD_s\left(A_{1,s},A_{2,s},...,A_{Q,s}\right)=\frac{1}{Q}\sum_{d=1}^{Q}\left\{ \begin{array}{l} 2\times rank\left(\left[A_{d,s}^{T}\ A_{1,s}^{T}\cdots A_{d-1,s}^{T}A_{d+1,s}^{T}\cdots A_{Q,s}^{T}\right]^{T}\right) \\ -rank\left(A_{d,s}\right)-rank\left(\left[A_{1,s}^{T}\cdots A_{d-1,s}^{T}A_{d+1,s}^{T}\cdots A_{Q,s}^{T}\right]^{T}\right) \end{array} \right\} \tag{12}$$

where $d=1,2,...Q$, $s=1,2,3$.The greater the value of *RSD*, the greater the rank of the kernel matrix [21], the better the performance of the MK function. Therefore, compare the value of

RSD_1, RSD_2 and RSD_3, select the corresponding kernel function with the maximum value as a part of optimal MK function K_{MKL} for each multi-source heterogeneous factor.

3.2 Parallel MK-SVM load forecasting

Experiments show that with the same sample quantity and stand-alone operation, the arithmetic speed of MK-SVM is slower than single kernel SVM. To make use of MK-SVM's ability of integrating multi-source heterogeneous data mean while improve the arithmetic speed, Map Reduce parallel load forecasting based on MK-SVM learning is proposed, for the purpose of meeting the speed and accuracy in big data of smart grid. The specific process is shown in figure 3.
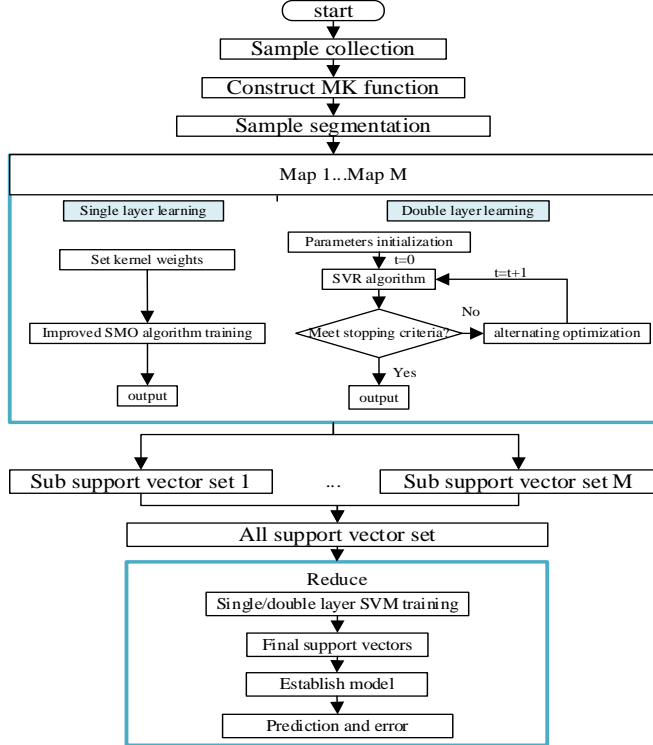
Figure 3.The diagram of parallel multiple kernel SVM load forecasting

(1) Collect the original data, and construct the training sample set with N training samples and the test sample set with S test samples;

(2) Establish MK functions according to figure 2;

(3) Sample set segmentation: divide the training sample set into M sub training sets, each sub training set contains about N/M training samples, then upload sub training sets to the HDFS file system;

(4) Parallel training of M map parts: each map part uses single or double layer MK learning algorithm to train SVM model after obtain the sub training sets in (3), the Value is the sub support vector sets;

(5) A reduce function obtains all the support vectors, single or double layer MK learning algorithm is used again to train SVM model for all the support vectors; establish the final MK-SVM load forecasting model and apply the model to test sample set, calculate errors.

4 Simulation example

4.1 Simulation system

To meet the high demand of speed and accuracy, a big data management platform based on Hadoop cluster is established, achieving data loading, data processing, load forecasting and data visualization. The platform consists of four parts: data integration layer, load forecasting layer, result visualization layer and user management layer, as shown in figure 4.
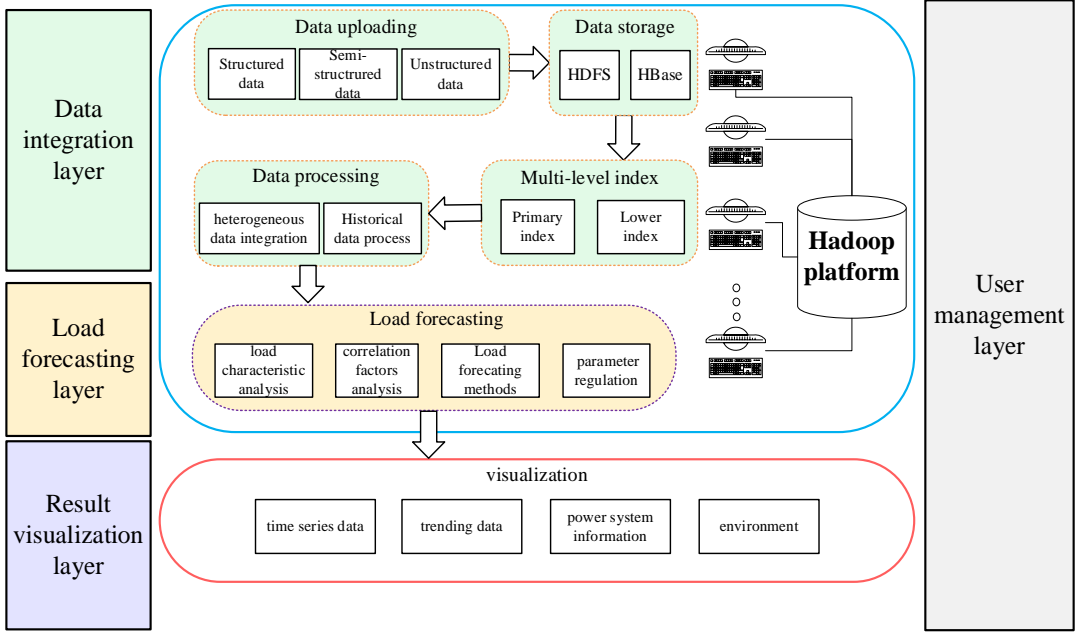
Figure 4. Big data management platform based on Hadoop cluster

(1) Data integration layer: this layer contains four modules: a) data loading module: the collected multi-source heterogeneous data (historical load data, meteorological data, GIS data, traffic data, economic and demographic data, electricity price data, and so on) are uploaded into a data storage module; b) data storage module: HBase stores small semi-structured and unstructured documents and complex variable structured data, HDFS file system stores large data files of these three kinds; c) the multi-level comprehensive index module: find the required data quickly according to search criteria; d) data processing module: process and integrate multi-source heterogeneous data.

(2) Load forecasting layer: this layer is the core part of the platform, it achieves correlation analysis between short-term load and various factors, makes the load forecasting algorithm MapReduce parallelized.

(3) Result visualization layer: it presents the real-time analysis results.

(4) User management layer: it realizes the operation of the platform to be safe, reliable, efficient.

The experiments are conducted on the above platform. The experimental data is from Heze City of Shandong Province, China. The frequency of collecting all data are 15 minutes. The training sample set with a total of 4608 training samples is from June 16 to August 2, 2014, and the test sample set with a total of 96 samples is from August 3, 2014.

4.2 Evaluation index of simulation results

Relative error:

$$E_i = \frac{\hat{L}_i - L_i}{L_i} \times 100\% \tag{13}$$

Average relative error:

$$e = \frac{1}{96} \sum_{i=1}^{96} \left| \frac{\hat{L}_i - L_i}{L_i} \right| \times 100\% \tag{14}$$

where $\hat{L}_i$ is the predicted load, L_i is the actual load.

4.3 Multi-source heterogeneous feature and MK function selection

This paper selects nine kinds of multi-source heterogeneous data. The training sample set is $\{(X_i, Y_i), i = 1, 2...N\}$, X represents multi-source heterogeneous load factors, which is: historical load L, temperature T, air pressure P, relative humidity H, wind direction WD, wind speed WS, rainfall R, electricity price J, holidays Ho. Figure 5 and table 1 are the results of sample distribution method and single variable method, respectively.
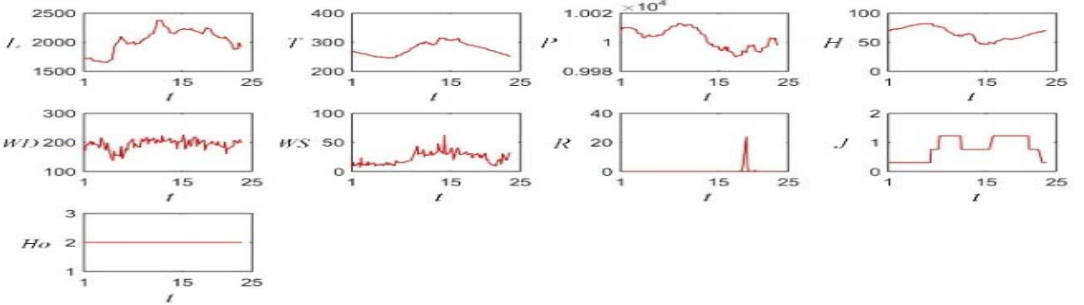
Figure 5. Sample distribution of multi-source heterogeneous load factors

Figure 5 shows that the distribution characteristic of historical load, temperature, air pressure, relative humidity, rainfall and electricity price is local, so it is more suitable to choose local kernel function. The distribution characteristic of wind direction, wind speed as well as the holidays is global, so global kernel function is better.

Table 1 shows that the predicting average relative error is the smallest using RBF for historical load, temperature, air pressure, relative humidity, rainfall and price. It indicates that the RBF can describe the distribution of these heterogeneous factors and their relationship with load. Mean while the optimal parameters are different, so RBFs with different parameters are parts of MK function. However, for wind direction, wind speed and holidays, sigmoid kernel has the minimum average relative error, so sigmoid functions with different parameters are also parts of MK function.

The conclusion obtained from figure 5 and table 1 is the same, therefore MK function is composed of RBFs and sigmoid functions with different parameters. At the same time, RBF is a local kernel function and sigmoid is a global kernel function, the combination of the two is complementary.

Table 1. The prediction of single factor and single kernel SVM

historical load *L*			temperature *T*			air pressure *P*		
Kernel function	Optimal parameters	e%	Kernel function	Optimal parameters	e%	Kernel function	Optimal parameters	e%
Linear		4.096925	Linear		4.842431	Linear		10.29210
Polynomial	0.1, 0.1, 1	4.156350	Polynomial	0.1, 0.1, 1	4.862370	Polynomial	0.01, 0.1,1	10.29799
RBF	2	3.515335	RBF	0.1	4.824133	RBF	10	10.02478
Simoid	0.1, 0.1	4.167938	Simoid	0.1, 0.1	4.860679	Simoid	1, 0.1	10.31133
relative humidity *H*			wind direction *WD*			wind speed *WS*		
Kernel function	Optimal parameters	e%	Kernel function	Optimal parameters	e%	Kernel function	Optimal parameters	e%
Linear		11.66816	Linear		10.29288	Linear		25.59834
Polynomial	0.001, 0.1, 1	10.09789	Polynomial	0.0001,0.1,1	10.30254	Polynomial	0.01,0.1, 1	25.59838
RBF	0.0001	9.89350	RBF	0.01	10.29330	RBF	0.0001	25.84783
Simoid	0.001, 0.1	9.97596	Simoid	0.1, 0.1	10.27861	Simoid	0.0001,0.1	25.57078
rainfall *R*			electricity price *J*			holidays *Ho*		
Kernel function	Optimal parameters	e%	Kernel function	Optimal parameters	e%	Kernel function	Optimal parameters	e%
Linear		10.28483	Linear		7.849361	Linear		10.31298
Polynomial	1, 1, 0.1	10.28483	Polynomial	3, 2, 0.1	7.269531	Polynomial	0.0001,1,0.1	10.29725
RBF	100	10.25560	RBF	3	7.262701	RBF	0.0001	10.30304
Simoid	1, 0.1	10.28483	Simoid	0.01,0.1	7.882567	Simoid	0.0001,0.1	10.29718

4.4 Simulation cases and results

This part compares the prediction accuracy of different model, algorithm and arithmetic speed by the following six cases.

Case 1:single kernel SVM model, ignoring the heterogeneous characteristics of the load factors. The predicted results are shown in figure 6(a) and table 2.

Table 2. Relative error of load forecasting of single kernel SVM

Single kernel	Linear	Polynomial	RBF	Sigmoid
e%	3.1604	3.0193	2.7928	2.6439

From figure 6(a)and table 2: the average relative error of the sigmoid is the smallest, and the second is the RBF. According to the MK function selected in section 4.3, the selected multi-source heterogeneous load factors in this paper are closely connected with RBF and sigmoid, so the corresponding average relative error is smaller. However, the accuracy is not ideal at valley and peak load, and the time after the second peak.

Case 2: MK-SVM model based on $l_1$norm and single layer learning. MK function is composed of nine single kernels, six are RBFs with different parameters, three are sigmoid kernels with different parameters, and the average weight is 1/9.The predicted results are shown in figure 6(b) and table 3.

Table 3. Average relative error of load forecasting based on single layer learning and $l_1$norm

Kernel function	RBF	Sigmoid	RBF and Sigmoid
e%	2.7928	2.6439	2.4460

From figure 6(b)and table 3: compared with single SVM, the average relative error of MK-SVM based on single layer learning is smaller, the accuracy is a bit better at the time after the second peak, but the error at valley and peak load is still big.

Case 3:MK-SVM model based on$l_1$norm and double layer learning. The weight of the kernel function and the Lagrange multiplier are solved by the double layer iterative optimization, and the kernel function weights are shown in table 4, the predicted results are shown in figure 6(c) and table 5.

Table 4. Kernel weight

RBF 1	RBF 2	RBF 3	RBF 4	RBF 5	RBF 6	Sigmoid1	Sigmoid2	Sigmoid3
0.000003	0.000002	0.365470	0.082664	0.300144	0.000003	0.084448	0.002137	0.000204

Table 5. Multiple kernel SVM forecasting relative error based on double layer learning and $l_1$norm

	RBF	Sigmoid	RBF and Sigmoid single layer learning	RBF and Sigmoid double layer learning
Max error %	8.50	7.96	8.75	4.53
e%	2.7928	2.6439	2.4460	1.6935

From figure 6(c) and table 5: the average relative error is decreased obviously, the accuracy has been significantly improved at peak load and the time after the second peak.The maximum relative error drops from 8.5% to 4.53%. The simulation results show that MK-SVM model based on$l_1$norm and double layer learning is better than the single kernel SVM model. Local kernel RBF and global kernel sigmoid are combined in MK function to effectively deal with the multi-source heterogeneous load factors, optimal weights are solved by algorithm, so load forecasting accuracy is improved.

Case 4: MK-SVM model with different MK based on $l_1$norm and double layer learning. Because the established MK function is the same by sample distribution method and single variable method in section 4.3, this case is set specifically to verify the feasibility of the proposed scheme section 3.1. The predicted results are shown in table 6:

Table 6. Different multiple kernel SVM forecasting relative error based on double layer learning and $l_1$norm

method	Kernel function	RSD	e%
Single kernel	Sigmoid	27	2.6439
Sample distribution method K_1	RBF and Sigmoid	42.3	1.6935
Single variable method K_2	Polynomial and Sigmoid	32	2.4369
weighting method K_3	RBF, Polynomial and Sigmoid	36.6	2.2545

From table 6: All the average relative error of SVM model with different MK are lower than single kernel function SVM. RSD obtained by MK of sample distribution method is the biggest, because local distributed RBFs that can describe historical load, temperature, pressure, relative humidity, rainfall and price are specifically replaced by global distributed polynomial. Simulation results show that the bigger RSD, the smaller the errors, therefore MK function selection proposed in section 3.1 is feasible.

Case 5:MK-SVM model based on l_p norm and double layer learning. The results are shown in table 7:

Table 7. Multiple kernel SVM forecasting relative error based on double layer learning and l_p norm

Norm p value	1	8/7	6/5	4/3	3/2	2
e%	1.6935	1.6322	1.6214	1.6285	1.6194	1.6105

From table7: the average relative error of MK $p(1 < p < +\infty)$ norm is smaller than $l_1$norm, this is because the kernel weights of $l_1$norm is constrained to $\|\lambda\|_1 = 1$, so the solution is sparse, which behaves poorly in load forecasting.

Case 6: Comparison between parallel computing and standalone computing. The computation time is almost the same with less kernel functions. But when kernels functions become more, the computation time of parallel is much less than standalone computing. The result is shown in figure 7.
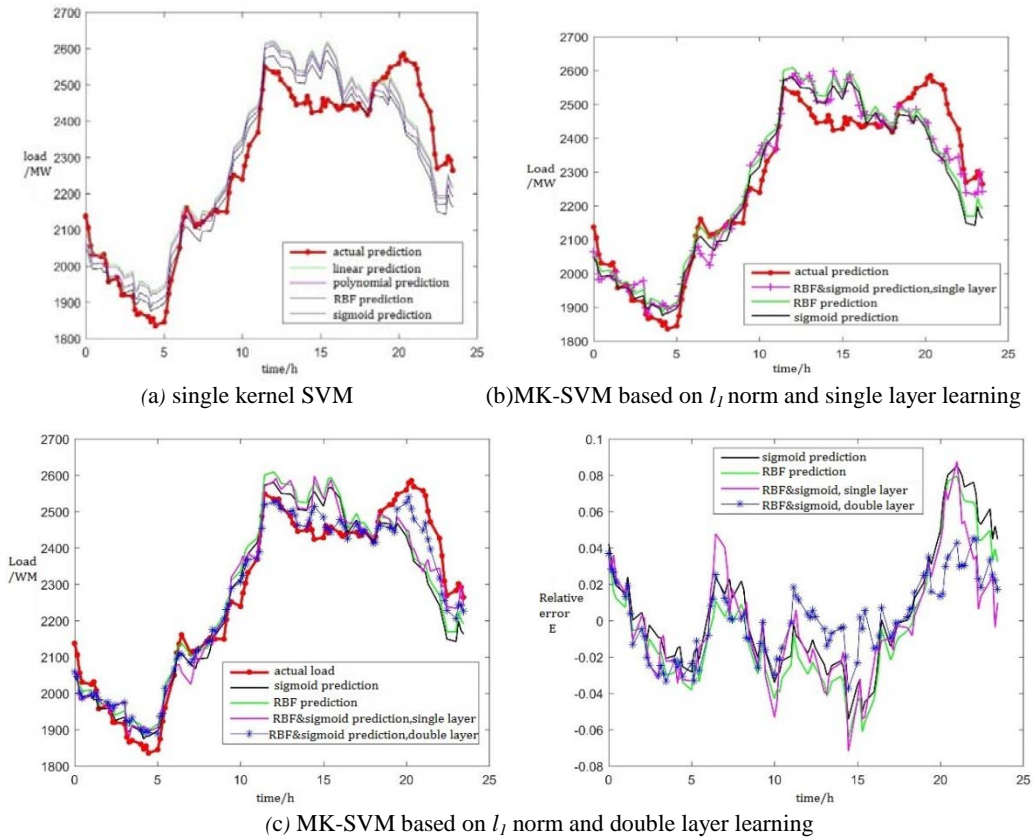
(a) single kernel SVM   (b)MK-SVM based on l_1 norm and single layer learning

(c) MK-SVM based on l_1 norm and double layer learning

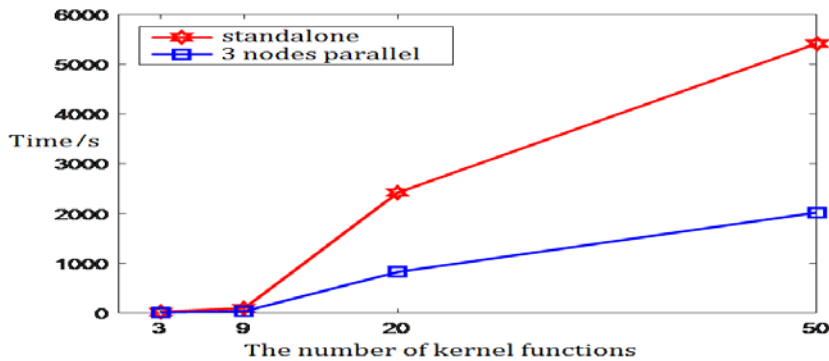**Figure 6**. SVM forecasting curve and relative error

Figure 7. Consume time contrast between one machine and Hadoop cluster

5 Conclusions

(1) MK-SVM model of load forecasting is proposed to integrate multi-source heterogeneous load factors in the big data environment of smart grid, due to the lack of multi-source heterogeneity processing capabilities in existing load forecasting algorithms. Parallel MK-SVM load forecasting simulations are conducted on big data management platform based on Hadoop cluster, arithmeticspeed and accuracy has been improved.

20

(2) The simulations show that the average relative error of single kernel SVM and MK-SVM is 2.64% and 1.69%, respectively. The accuracy of MK-SVM based on l_p norm and double layer learning is the highest. The computation time of parallel and standalone computing with 20 kernel functions is 826s and 2421s, respectively.

(3) Compared with single kernel SVM, the parameter selection range of MK-SVM is much wider, which makes it much easier to determine. For example, if the MK function is composed of 100 RBFs,100 values are prepared to be chosen as the MK function parameters, thus reducing the difficulty of parameter selection.

References

1. J. Wang, Z. Ji, M. Shi, F. Meng, C. Zhu and D. Zhang. "Scenario Analysis and Application Research on Big Data in Smart Power Distribution and Consumption Systems," Proceedings of the CSEE, **35**: 1829-1836(2015)
2. D. Wang, Z. Sun. "Big Data Analysis and Parallel Load Forecasting of Electric Power User Side," Proceedings of the CSEE, **35**: 527-537(2015)
3. D. Zhang, X. Miao, L. Liu. "Research on Development Strategy for Smart Grid Big Data," Proceedings of the CSEE, **35**: 2-11(2015)
4. H. Geng, M. Zhang, Y. Zhang. "A multi-dimensional decision-making analysis and visualization method for solving multi-source and heterogeneous datasets," Information Technology, **37**: 49-53(2013)
5. K. Wan, R. Liu. "Application of Interval Time-Series Vector Autoregressive Model in Short-Term Load Forecasting," Power System Technology, **37**: 49-53(2013)
6. F. Zhang, T. Han. "Short-term Load Forecasting Based on Partial Least-squares Regression," Power System Technology, **27**: 36-40(2003)
7. G. Zhong. "Applications of Partial Least-squares Regression Analysis in Short-term Load Forecasting" (Hebei University of Engineering, 2011)
8. S. Ge, O. Jia, H. Liu. "A Gray Neural Network Model Improved by Genetic Algorithm for Short-term Load Forecasting in Price-Sensitive Environment," Power System Technology, **36**: 224-229(2012)
9. Q. Pang. "A Rough Set-Based Neural Network Load Forecasting Algorithm and Its Application in Short-Term Load Forecasting," Power System Technology, **34**: 168-173(2010)
10. R. Jiao, C. Su, B. Lin. "Short-term Load Forecasting by Grey Model With Weather Factor-Based Correction," Power System Technology, **37**: 720-725(2013)
11. F. Shu. "Short-term Load Forecasting of Power System Based on RBF Neural Network and Fuzzy Theory,"(XI'AN University of Technology, 2008)
12. Bakirtzis A G, Theocharis J B. "Short term load forecasting using fuzzy neural networks," Power System, IEEE Transactions on, **10**:1518-1524(1995)
13. M. Zeng, C. Lv, K. Tian. "Least Squares-support Vector Machine Load Forecasting Approach Optimized by Bacterial Colony Chemotaxis Method,"Proceedings of the CSEE, **31**: 93-99(2011)
14. X. Li, J. Zhang. "Short-term Load Forecasting for Similar Days Based on Support Vector Machine and Dempster-Shafer Theory,"Power System Technology, **34**: 143-147(2010)
15. S. Zhang, B. Zhao, F. Wang. "Short-term Power Load Forecasting Based on Big Data," Proceedings of the CSEE, **35**: 37-42(2015)
16. WANG Ben, LENG Beixue, ZHANG Xihai, et al. Application Profiles of Support Vector Machine in Short-term Load Forecasting[J]. Proceedings of the CSU-EPSA, 2011, 23(4): 115-121.
17. X. Shao. "The Research and Application of Multiple Kernel Prediction Model Based on Statistical Learning Theory,"(Central South University, 2013)
18. C. Zhao."The Choice of Funds Based on Support Vector Machine with Mixed-kernel Function," (South China University of Technology, 2012)

19. Y.Cao. "The Applications of Hybrid-kernel Function Support Vector Machine in Credit Evaluation,"(Harbin Institute of Technology, 2007)
20. Sonnenburg S, Rätsch G, Schäfer C. "Large scale multiple kernel learning," The Journal of Machine Learning Research, **7**: 1531-1565(2006)
21. L. Luo. "Research on Kernel Selection of Support Vector Machine,"(Xiamen University, 2007)