A Model for Micro-grid Short-term Load Forecasting Based on SR-LS-SVR

Yu-xian ZHANG and Yu-bao GAO
School of Electrical Engineering, University of Technology Shenyang, Shenyang, China

Keywords: Micro-grid, Net load, Short-term load forecasting, SVR.

Abstract. Considering the impact of distributed power on the net load, the short-term load forecasting model established by SR-LS-SVR from the global perspective of the distribution network. The model parameters are quickly optimized by fine-grained parallel PSO. The accuracy, robustness and efficiency of the model are verified by the Ausgrid smart grid measurement data.

Introduction

With the development of distribution networks, more and more distributed power sources are connected to the distribution network in the form of microgrids[1]. Compared with large-scale distribution systems, the net load of microgrids shows more non-smoothness[2]. The net load refers to the difference between the power consumption of each type of load in the microgrid and the distributed power generated in the microgrid[3]. If the net load is calculated according to the individual prediction results of each part of the microgrid, it will inevitably lead to cumulative error to reduce the prediction accuracy and to affect the economic dispatch of the distribution network.

In this article, a combined forecasting approach is proposed to improve the forecasting accuracy and efficiency of net load. Based on the measured data of Ausgrid, the sparse robust LS-SVR (SR-LS-SVR) is used to establish the net load forecasting model which parameters are quickly optimized by running the fine-grained parallel PSO on the GPU.

Proposed Combined Forecasting Model

The separate forecasting approach needs two forecasting models at least which inevitably leads to the accumulative error to reduce the forecasting performance[4]. The proposed combined forecasting approach can address the disadvantages. In this article, we use the SR-LS-SVR theory to propose a self-updating net load forecasting model. The model not only update the training data but also quickly adjust the model parameters and then make predictions point by point. Fig. 1 shows the flowchart of the proposed model.

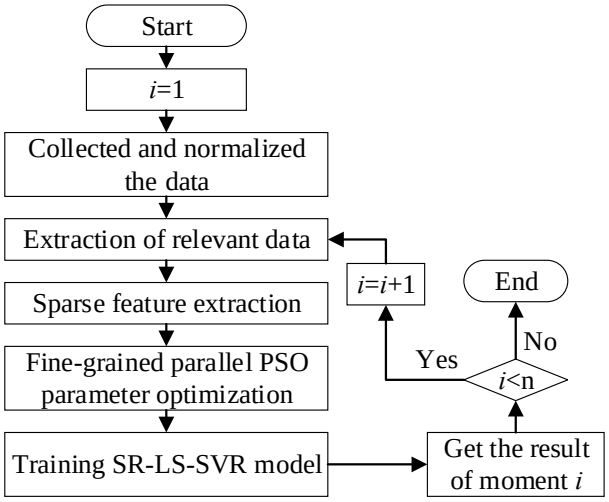

Figure 1. Flowchart of the self-updating forecasting model

Improved Forecasting Algorithms

By replacing the inequality constraints with equality constraints, LS-SVR greatly reduces the computational complexity compared to SVR, meanwhile, it also introduces two potential problems: (1) The evaluation criterion of LS-SVR objective function makes all training points determine the decision function. For data with outliers and non-normally distributed noise, it will lead to large prediction errors and make the decision function lack of robustness. (2) All training sample points contribute to the decision function in LS-SVR but the decision function in SVR depends only on a small number of support vectors so that the solution of LS-SVR lacks sparseness compared to the solution of SVR[5]. For the above two problems of LS-SVR, an improved LS-SVR algorithm with both sparsity and robustness is adopted.

Sparse Improvement

From the solution process of LS-SVR, $\boldsymbol{\omega}$ is the linear combination of the input vector in the high-dimensional feature space and the sparsity of the solution can be obtained to some extent by finding the approximate base of input vector in the feature space[0].The training data set $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$ is mapped to the high dimensional feature space by the radial basis function $\varphi(\cdot)$ and the mapping set is $A=\{\varphi(\boldsymbol{x}_i)\}_{i=1}^N$. From the matrix theory, if the mapping set is linearly related there is at least one $\varphi(\boldsymbol{x}_q)=\sum_{i=1,i\neq q}^{n}\lambda_i\varphi(\boldsymbol{x}_i)$, where $\lambda_i\in\mathbf{R}$. Although the radial basis function $\varphi(\cdot)$ cannot be accurately expressed, it can be expressed indirectly through its kernel form $K(\boldsymbol{x}_q,\boldsymbol{x}_q)=\sum_{i=1,i\neq q}^{N}\sum_{j=1,j\neq q}^{N}\lambda_i\lambda_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)$. The solution process of the approximate maximal irrelevant group of mapping set *A* is as follows:

(1) Initialize the maximal irrelevant group $B=\{\phi\}$. In the set $S=\{1,2,\mathrm{L},N\}$, select the data i=1 and put it in *B*.

(2) Select $i=i+1$ in S, and calculate $\min_{\lambda} G(\lambda)=\left(\varphi(\boldsymbol{x}_i)-\sum_{i\in B}\lambda_i\varphi(\boldsymbol{x}_i)\right)^T\left(\varphi(\boldsymbol{x}_i)-\sum_{i\in B}\lambda_i\varphi(\boldsymbol{x}_i)\right)$

(3) If $\min_{\lambda} G(\lambda)<\varepsilon$, then $\varphi(\boldsymbol{x}_i)$ can be linearly represented by $\{\varphi(\boldsymbol{x}_i)|i\in B\}$, and data i is discarded; if $\min_{\lambda} G(\lambda)\geq\varepsilon$, then $\varphi(\boldsymbol{x}_i)$ cannot be represented linearly by $\{\varphi(\boldsymbol{x}_i)|i\in B\}$, and i is placed in set *B*.

(4) If $i\leq n$ is not met, go to (2); otherwise, iterate.

The sparse data set $\Psi=\{\boldsymbol{x}_k, y_k\}_{k=1}^r$ consists of the number in the set *B* in the data sample corresponding to the original data set and *r* is the number of samples after sparse. Ψ is mapped to $\Phi=[\varphi(\boldsymbol{x}_1),\varphi(\boldsymbol{x}_2),\cdots,\varphi(\boldsymbol{x}_r)]$ by $\varphi(\cdot)$. Because Ψ is a maximally irrelevant group of *A*, then

$$\boldsymbol{\omega}=\sum_{k=1}^{r}\beta_k\varphi(\boldsymbol{x}_k)=\boldsymbol{\Phi\beta} \tag{1}$$

Replace $\boldsymbol{\omega}$ In the optimization goal of LS-SVR with equation (1)

$$\begin{aligned}&\min_{\beta,b} F(\boldsymbol{\beta},b,\boldsymbol{e})=\frac{1}{2}\left(\gamma\sum_{i=1}^{r}\|e_i\|^2+(\boldsymbol{\Phi\beta})^T(\boldsymbol{\Phi\beta})\right)\\&\text{s.t. } y_i=(\boldsymbol{\Phi\beta})^T\varphi(\boldsymbol{x}_i)+b+e_i, i=1,2,\mathrm{L},r\end{aligned} \tag{2}$$

Robust Improvement

In order to improve the robust performance of the sparse LS-SVR algorithm described above, the weighting factor v_i is introduced to the error term e_i in equation (2), thereby obtaining the following optimization problem:

$$\min_{\beta,b} F(\boldsymbol{\beta},b,\boldsymbol{e}) = \frac{1}{2}\left(\gamma\sum_{i=1}^{r} v_i \|e_i\|^2 + (\boldsymbol{\Phi\beta})^T(\boldsymbol{\Phi\beta})\right) \tag{3}$$

$$\text{s.t. } y_i = (\boldsymbol{\Phi\beta})^T \varphi(\boldsymbol{x}_i) + b + e_i, i = 1,2,\text{L}\ ,r$$

After introducing the Lagrangian operator

$$L(\boldsymbol{\beta},b,\boldsymbol{e},\boldsymbol{\alpha}) = F(\boldsymbol{\beta},b,\boldsymbol{e}) - \sum_{i=1}^{r}\alpha_i\left((\boldsymbol{\Phi\beta})^T\varphi(\boldsymbol{x}_i) + b + e_i - y_i\right) \tag{4}$$

Where: $\boldsymbol{\alpha} \in \mathbf{R}^r$ is a Lagrangian multiplier. Eliminate e and α according to the optimal conditions:

$$\begin{bmatrix} \boldsymbol{\Phi}^T\sum_{i=1}^{r}\varphi(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_i)^T\boldsymbol{\Phi} - -\frac{1}{\gamma\boldsymbol{v}}\boldsymbol{\Phi}^T\boldsymbol{\Phi} & \boldsymbol{\Phi}^T\sum_{i=1}^{r}\varphi(\boldsymbol{x}_i) \\ \sum_{i=1}^{r}\varphi(\boldsymbol{x}_i)\boldsymbol{\Phi}^T & N \end{bmatrix} \cdot \begin{bmatrix} \beta \\ b \end{bmatrix} = \boldsymbol{\alpha} \in \mathbf{R}^r \tag{5}$$

Where $\boldsymbol{v} = \text{diag}(v_1, v_2, \ldots, v_r)$, v is determined by the IGGIII weight function[7].

Fine-grained Parallel PSO Parameter Optimization

The value of penalty factor γ and kernel parameter σ has a great influence on the prediction results of SVM model, which determines the performance of SVM model. Usually, the cross-validation method is adopted in the optimization algorithm and the loss function is repeatedly called to adjust the parameters to optimize the performance of the SVM model.

The idea of PSO algorithm optimizing SVM is that to treat the parameters γ and σ to be selected in the SVM as two particles. The square loss function of the k-fold cross-validation method constitutes the objective function to calculate the fitness value and then update these two particles continuously yield the optimal values of γ and σ. The fitness evaluation of each particle is independent of each other and for each particle K-fold cross-validation needs to perform k independent operations. It is a typical Single Instruction Multiple Data (SIMD) structure.

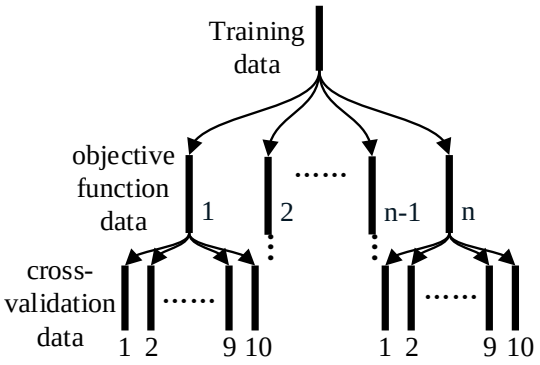

Figure 2. Fine-grained parallel optimization data distribution

According to these two independences, the objective function of PSO is parallelized. Figure 3 shows the fine-grained parallel optimization data distribution diagram which expresses the data distribution process in the optimization process. First, the training data is distributed to the threads of n particles, then the objective function data is composed with the initialized particles and distributed training data on the thread. According to the criterion of cross-validation, the objective function data is segmented into the data required for each folding operation, then the fitness value of the PSO algorithm is calculated concurrently.

Case Study

In order to verify the effectiveness of the proposed net load forecasting method, this article uses the measured data released by the Australian distribution company Ausgrid in 2014. The data set contains the load and power data of 300 households in Sydney with a sampling interval of half an

hour. A single home distribution structure, including household load *gc*, controllable load *cl*, and PV power *gg*, where *cl* is the power company controlling the all-electric hot water system. *d*(*j*) is the net load and its power balance equation is

$$d(j) = gc(j) + cl(j) - gg(j) \tag{6}$$

Where, $d(j)>0$ means that the microgrid supplies power to the house, $d(j)<0$ means that the house feeds the microgrid, and *j* denotes the time index.

The root mean square error (RMSE) and the mean absolute error percentage (MAPE) are used to evaluate the prediction results.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(L_i - \hat{L}_i\right)^2}{n}} \tag{7}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{L_i - \hat{L}_i}{L_i}\right| \times 100\% \tag{8}$$

Where, L_i and $\hat{L}_i$ are the actual load and preload at time *i*. *n* is the predicted number of points.

In order to verify the accuracy of the proposed model, LS-SVR and SR-LS-SVR were used to make a short-term forecasting of the net load on March 31, 2013. The method proposed in this article can better predict the trend of the net load as a whole in Figure 3. From the data in Figure 3, the MAPE of the SR-LS-SVR model is 11.04% which is significantly smaller than the MAPE of SVR that indicates the effectiveness of the proposed method. By analyzing the error, the MAPE of this test is larger than the MAPE of the traditional load forecast. The reason is that the absolute error percentage of the predicted point is far greater than 100% when the net load power is near zero resulting in a larger overall MAPE.

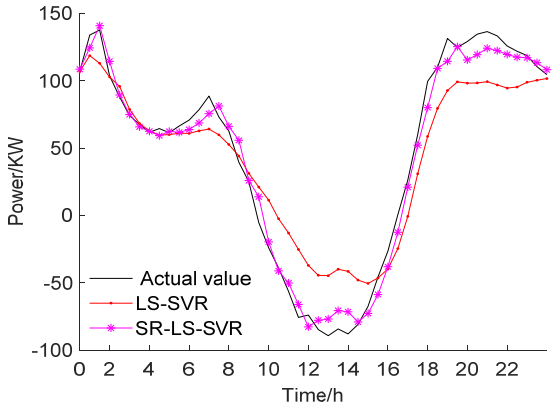

Figure 3. Forecasting results of one-day

Table 1. Compared errors of one-day

Model	*RMSE*/KW	*MAPE*/%	e_{max} /KW
LS-SVR	16.87	19.17	35.37
SR-LS-SVR	9.38	11.04	19.41

For the robustness of this method, the above two models were used to perform 30 days of daily load forecasting. The 30-day MAPE curve is shown in Figure 4. The 30-day average MAPE of the two models was 15.56% and 10.72%, respectively. Analysis of Figure 4 shows that the SR-LS-SVR model has better robustness and improves the stability of the prediction results that verifies the effectiveness of the proposed method.

SR-LS-SVR has high robustness, but its running time is greatly increased. In order to quickly optimize the model parameters in the prediction process, a fine-grained parallel PSO is proposed. The standard PSO and parallel PSO optimization time comparison is shown in Table 2. The

acceleration ratios of 4.47 and 9.35 are obtained respectively which verifies the optimization efficiency of fine-grained parallel PSO.

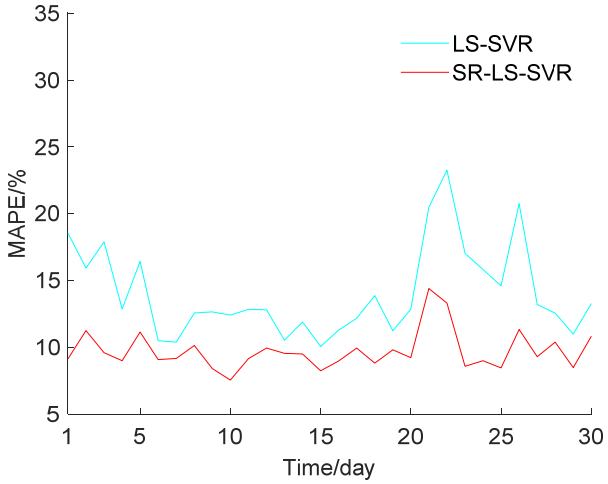

Figure 4. MAPE curve for 30-day

Table 2. Compared time of model running

Model	Standard PSO	Parallel PSO	speed-up
LS-SVR	324.31	72.55	4.47
SR-LS-SVR	29512.37	3156.42	9.35

Summary

A large number of distributed power sources are connected to the power distribution system through the microgrid. The uncertainty of these distributed power and the uncertainty of the load interact that making the uncertainty of the microgrid increasingly prominent. In this paper, the net load forecasting of the microgrid is carried out from the layer of global control of the distribution network. The experimental results show that the combined prediction model and parallel parameter optimization using SR-LS-SVR show good predictive performance in the PV microgrid instance.

References

[1] Yang Xinfa, Su Jian, Lü Zhipeng, et al. Overview on Micro-grid Technology[J]. Proceedings of the CSEE, 2014(1): 57-70.

[2] Zhong Qing, Sun Wen, YU Nanhua, et al. Load and Power Forecasting in Active Distribution Network Planning[J]. Proceedings of the CSEE, 2014, 34(19): 3050-3056.

[3] Li, Yong, Wen, et al. A combined forecasting approach with model self-adjustment for renewable generations and energy loads in smart community[J]. Energy, 2017, 216-2 27.

[4] Jin L, Cong D, Guangyi L, et al. Short-term net feeder load forecasting of microgrid considering weather conditions[C]// In Proceedings of Energy Conference, Cavate, IEEE, May 2014: 1205-1209.

[5] Rousseeuw P J, Leroy A M. Robust Regression and Outlier Detection[J]. Techno metrics, 2005, 31(2): 260-261.

[6] Gan Liangzhi, Sun Zonghai, Sun Youxian. Sparse least squares support vector machine[J]. Journal of Zhejiang University (Engineering Science), 2007, 41 (2):245-248.

[7] Xu Q, Zhang J, Jiang C, et al. Weighted quantile regression via support vector machine[J]. Expert Systems with Applications, 2015, 42(13):5441-5451.