ICIC Express Letters
Part B: Applications
Volume **7**, Number **5**, May **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **977–984**

SHORT-TERM LOAD FORECASTING VIA INTEGRATED INCREMENTAL EXTREME SUPPORT VECTOR REGRESSION APPROACH

Min Jiang[1], Lin Sun[1], Long Zhang[2], Jun Kong[1,3] and Tiejiang Yuan[3]

[1]Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education)
Jiangnan University
No. 1800, Lihu Avenue, Wuxi 214122, P. R. China
minjiang@jiangnan.edu.cn; sunlinzhr@163.com

[2]China State Grid Electric Power Research Institute
Xinjiang Electric Power Company
Urumqi 830011, P. R. China

[3]College of Electrical Engineering
Xinjiang University
No. 14, Shengli Road, Urumqi 830046, P. R. China

Received October 2015; accepted January 2016

Abstract. *In this paper, a novel short-term load forecasting (STLF) model based on integrated incremental extreme support vector regression (II-ESVR) approach is presented. Firstly an incremental ESVR (IESVR) model is proposed to adapt to incremental learning which can preserve prior knowledge while learning from new data. The IESVR model can avoid storing large amounts of historical load data and retraining the whole dataset, which will reduce the system consumption of computation cost and memory storage. Then an integrated IESVR network is constructed to handle the instability of IESVR method. As verified by the experiments, the IESVR model has very fast incremental learning speed with much lower computational cost. We show attractive experimental results to highlight the system efficiency and stability by using our integrated IESVR approach to forecast short-term power load.*
Keywords: Short-term load forecasting (STLF), Incremental extreme support vector regression (IESVR), Incremental learning, Integrated network

1. **Introduction.** Load forecasting has always been an important issue and played a vital role in energy management system (EMS), especially in short-term load forecasting (STLF). The reliable operations of EMS such as system scheduling, system maintenance and economic dispatch can be carried out more efficiently with accurate information [1]. During the past few decades, researchers have paid much attention to apply batch learning methods in STLF such as neural network (NN) methods [1], support vector machine (SVM) methods [2] and extreme learning machine (ELM) methods [3]. Recently, Liu et al. [4] proposed a new nonlinear SVM method, called extreme support vector machine (ESVM). ESVM can produce better generalization performance than SVM and ELM in almost all of the time and run much faster than other nonlinear SVM algorithms with comparable accuracy. Zhu et al. [5] also extended ESVM model into a regression model, named ESVR. Actually, in real life applications, the STLF problems can naturally be viewed as large-scale incremental learning rather than batch learning and the power data are often collected continuously in real time.

Based on the batch learning methods, many incremental sequential algorithms have been presented to meet the actual application demand. Stochastic gradient descent back-propagation (SGD-BP) is one of the most common-used BP algorithms in real sequential learning applications [6]. However, SGD-BP suffers a lot from slow training error convergence with the large-scale data. Montana and Parrella [7] proposed an online SVR method to deal with the new coming data, but it only can be used to learn one by one. Incremental SVR (ISVR) method [8] is innovatively developed for large-scale traffic flow prediction by using kernel correlation matrix. Liang et al. [9] and Li et al. [10] introduced online sequential ELM (OS-ELM) and structure-adjustable OS-ELM (SAO-ELM) methods respectively which have much faster learning speed and produce better generalization performance than other sequential learning algorithms. However, OS-ELM and SAO-ELM may be over-fitting when the number of hidden nodes is relatively large.

Motivated by the batch-learning process of ESVR, we extend the ESVR into an incremental learning model (IESVR) to deal with large-scale power data from actual power management systems. It can work for chunk-by-chunk (with fixed or varying size) and efficiently reduce the computation cost and memory storage. In order to deal with the instability of the modified IESVR method, an integrated IESVR (II-ESVR) network is presented. The solutions of IESVR and II-ESVR method are proposed for large-scale online power load forecasting applications to avoid retraining all the data while adding new data. The experiments show that the IESVR method tends to have better performance than traditional OS-ELM method with faster training speed. The constructed II-ESVR network significantly outperforms state-of-the-art methods in stability.

This paper is organized as follows. Section 2 briefly introduces ESVR theory and extends an incremental learning algorithm for ESVR. Section 3 proposes an integrated network to deal with the unstable issue of IESVR. In Section 4, the proposed approach is applied for STLF and evaluation performances are discussed. Conclusions based on this study are drawn and highlighted in Section 5.

2. **Incremental Extreme Support Vector Regression Method.**

2.1. **Review of extreme support vector machine for regression.** Here we briefly review the model of ESVR. Similar to ELM theory, ESVR is a novel single hidden layer feedforward network (SLFN) model combined SVM [11] and ELM [12]. For the given sample data $A = \{(x_i, y_i)|x_i \in R^m\}_{i=1}^N$, here x_i is an $m \times 1$ input vector and y_i is the target value. The decision function of ESVR is $y = \Phi(a, x) * \beta + b\boldsymbol{e}$ aiming at approximating any continuous target function, and the hidden weights can be generated randomly similar to ELM [12].

$$\underset{(\beta,b,\xi)\in R^{L+1+N}}{\text{Minmize}} \frac{C}{2}\|\xi\|^2 + \frac{1}{2}\left\|\begin{bmatrix}\beta \\ b\end{bmatrix}\right\|^2 \tag{1}$$

$$s.t.\ \ \Phi(a, x) * \beta + b\boldsymbol{e} - y = \xi \tag{2}$$

where C and ξ respectively represent the user-defined regularization parameter and the training error.

Based on the Lagrangian theory, the goal of ESVR is equivalent to dealing with the least square problem [13], and the Lagrange formula can be defined as follows:

$$L_{ESVR} = \frac{C}{2}\|\xi\|^2 + \frac{1}{2}\left\|\begin{bmatrix}\beta \\ b\end{bmatrix}\right\|^2 - \lambda^T\left(\Phi(a, x) * \beta + b\boldsymbol{e} - y - \xi\right) \tag{3}$$

where $\lambda \in R^N$ is the Lagrange multiplier according to the LSSVM theory [13]. The KKT optimality condition theory is applied and can get the final expression of β and b as:

$$W = \begin{bmatrix}\beta \\ b\end{bmatrix} = H_\Phi^T\lambda = \left(\frac{I}{C} + H_\Phi^T H_\Phi\right)^{-1} H_\Phi^T \mathrm{y} \qquad (H_\Phi = [\Phi(a, x)\ \ \boldsymbol{e}]) \tag{4}$$

It can be clearly seen that the ESVR model is developed from SVM and ELM, and takes advantages of SVM and ELM. Compared to nonlinear SVM, ESVR leads to a simplified and extremely fast approach with lower computational cost. However, many real-life application problems can be more naturally viewed as online rather than batch learning problems since the training data may arrive continuously. ESVR is inappropriate to deal with the incremental dataset due to its highly intensive computation.

2.2. **Incremental extreme support vector regression method.** In this section, we proposed an incremental learning algorithm for ESVR which is capable of adding new data to generate an alternative updating regression model. For the whole training process, the initialization and incremental learning process will be conducted to update the output weight matrix W. Assume given a sequential training data set A, we first extract a chunk of training set $A_0 = \{(x_i, y_i) | x_i \in R^m\}_{i=1}^{N_0}$ and $N_0 > L$ (representing the number of hidden nodes) to initialize the ESVR model, and the target of ESVR is to minimize $\left\| H_{0\Phi} W_0^T - Y_0 \right\|$ and finally get the optimal output weight $W_0 = \begin{bmatrix} \beta_0 \\ b \end{bmatrix} = {\Omega_0}^{-1} \Lambda_0$, where $Y_0 = [y_1 \cdots y_{N_0}]^T$, $\Omega_0 = I/C + H_{0\Phi}^T H_{0\Phi} \in R^{(L+1)\times(L+1)}$ and $\Lambda_0 = H_{0\Phi}^T Y_0 \in R^{(L+1)\times 1}$.

Now consider the second chunk of training dataset $A_1 = \{(x_i, y_i)\}_{i=N_0+1}^{N_0+N_1}$, where N_1 denotes the number of the training data in this chunk. Considering both chunks of A_0 and A_1, the solution of ESVR model aims at minimizing $\left\| H_{1\Phi} W_1^T - Y_1 \right\|$ and get the output weight matrix W_1:

$$W_1 = \begin{bmatrix} \beta_1 \\ b \end{bmatrix} = {\Omega_1}^{-1} \begin{bmatrix} H_{0\Phi} \\ H_{1\Phi} \end{bmatrix}^T \begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} \tag{5}$$

For incremental learning, we have to express W_1 as a recursive function of W_0 and A_1 instead of a function of the dataset A_0, and then Ω_1 can be calculated as:

$$\Omega_1 = \frac{I}{C} + \begin{bmatrix} H_{0\Phi} \\ H_{1\Phi} \end{bmatrix}^T \begin{bmatrix} H_0 \\ H_1 \end{bmatrix} = \Omega_0 + {H^T}_{1\Phi} H_{1\Phi} \tag{6}$$

Combining the aforementioned Equations (6) and (7), W_1 can be expressed as:

$$W_1 = W_0 + \Omega_1^{-1} \Lambda_1 \qquad \left(\Lambda_1 = H_{1\Phi}^T \left(Y_1 - H_{1\Phi} W_0\right)\right) \tag{7}$$

Generalizing the previous conclusions, as new data coming in chunk by chunk, a recursive algorithm is presented for updating all the parameters same with the solution of batch ESVR. When $(k+1)$th chunk of dataset is received as: $A_{k+1} = \{(x_i, y_i)\}_{i=\left(\sum_{j=0}^{k} N_j\right)+1}^{\sum_{j=0}^{k+1} N_j}$, where $k > 0$ and N_j denotes the number of training dataset in the jth chunk, and N_j can be a fixed or varying value. Therefore, we have the following conclusions:

$$\Omega_{k+1} = \Omega_k + H_{(k+1)\Phi}^T H_{(k+1)\Phi} \tag{8}$$

$$\Lambda_{k+1} = H_{(k+1)\Phi}^T \left(Y_{k+1} - H_{(k+1)\Phi} W_k\right) \tag{9}$$

Then the equations for updating W_{k+1} can be written as:

$$W_{k+1} = W_k + \Omega_{k+1}^{-1} \Lambda_{k+1} \tag{10}$$

Now we have transformed the solution of an ESVR model into an iterative model. For IESVR, all we need to store in memory is the random input-weight matrix: $a \in R^{L\times(m+1)}$, a relatively small matrix $\Omega_0 = R^{(L+1)\times(L+1)}$ and $W_0 = R^{(L+1)\times N_0}$, when the new coming dataset A_{k+1} is received, according to the ESVR algorithm and Equation (4), it will retrain the whole dataset by taking $2(L+1)^2 \sum N_i$ and $2(L+1) \sum N_i$ operations to compute $H_{(k+1)\Phi}^T H_{(k+1)\Phi}$ and $H_{(k+1)\Phi}^T Y_{k+1}$, while for IESVR algorithm, it only takes $2(L+1)^2 N_i$ operations to compute $H_{(k+1)\Phi}^T H_{(k+1)\Phi}$ and $2(L+1)N_i$ operations to compute $H_{(k+1)\Phi}^T Y_{k+1}$ referring to Equations (9) and (10), and it can be seen clearly that the computation cost of IESVR algorithm will only increase linearly consistent with the number of new dataset.

Following the initialization phase, the incremental learning phase can work on chunk-by-chunk (with fixed or varying size) as desired without reserving historical data. Therefore, IESVR algorithm allows us to handle arbitrarily large-scale dataset by successively adding new data, which can be very useful when the training data cannot be obtained in one time or the training data is too large to store in memory.

3. **Short-Term Load Forecasting Based on Integrated IESVR Method.** According to IESVR learning theory, the much fast learning speed and excellent generalization capacity of IESVR make it very suitable for STLF. However, the random generated parameters of IESVR and the uncertainty of new coming data can form a crux in the stability of the forecasting results [2]. In real STLF applications, this issue will make the forecasting results unacceptable. Considering to enhance the stability, a novel framework is presented for tackling the instability of STLF problems. The basic idea is to adopt the boosting learning strategy and construct an integrated network (II-ESVR) to solve this problem.

Under the integrated network, for the given training dataset $A = \{(x_1, y_1), \ldots, (x_N, y_N)\}$, the II-ESVR contains a series of component IESVR predictors $\{f_1, \ldots, f_P\}$ with the same structure and parameters, and the output of the kth component IESVR is denoted as $f_k(x_i)$. The learning theory for each IESVR model is described in Section 2. In terms of the unstable issue of each single IESVR model, we apply the Mean-Value method in II-ESVR, a special case of weighted algorithm. For each component predictors $\{f_1, \ldots, f_P\}$, a set of weights $\{w_1, \ldots, w_P\}$ should correspond with every predictors where $\sum_{k=1}^{P} w_k = 1$ ($w_k \geq 0$) and $w_1 = w_2, \ldots = w_P$. Then we combine the component predictors $\{f_1, \ldots, f_P\}$, and get the average of the output of each IESVR model as the final result of II-ESVR network. It can be defined as $F(x_i) = \frac{1}{P}\sum_{k=1}^{P} w_k f_k(x_i)$.

We expect that II-ESVR network works better than single IESVR model because the randomly generated parameters make each single IESVR model work in the random region. Therefore, each single IESVR model may have different adaptive capacities to the new data. When the data come into the integrated network sequentially, some of IESVR models may adapt faster and better to new data than others. The integrated IESVR network can effectively mitigate the instability problems of single IESVR model. In the II-ESVR network, each single IESVR is trained independently, and all operations can be performed in parallel. Therefore, parallel computation mode can be applied to reducing computation cost. Specifically, we construct the integrated network by utilizing incremental learning model IESVR, which can efficiently handle both online learning and large-scale problem even with millions of data.

The proposed approach aims at forecasting the day-ahead power load based on large-scale power load data. Thus, the selection of input vector has a great impact on the accuracy of forecasting model [1,3]. The weekly and monthly load demand patterns of ISO-New England (www.iso-ne.com) are shown in Figure 1. From Figure 1(a), it can be easily learned that the changes of power load in 24 hours for different days in a week share a similar pattern. Besides, from Figure 1(b), the power loads of four weeks in a month also show periodical trend changes. Hence, to capture the most critical factors, we defined the input variables including the hour (h) of the day (d) in the week (w), working day or holiday (0 or 1), real-time temperature (t) and corresponding power load $P(w, d, h)$, $P(w, d, h)$, $P(w, d, h)$, and $P_{avg}(w, d-1)$.

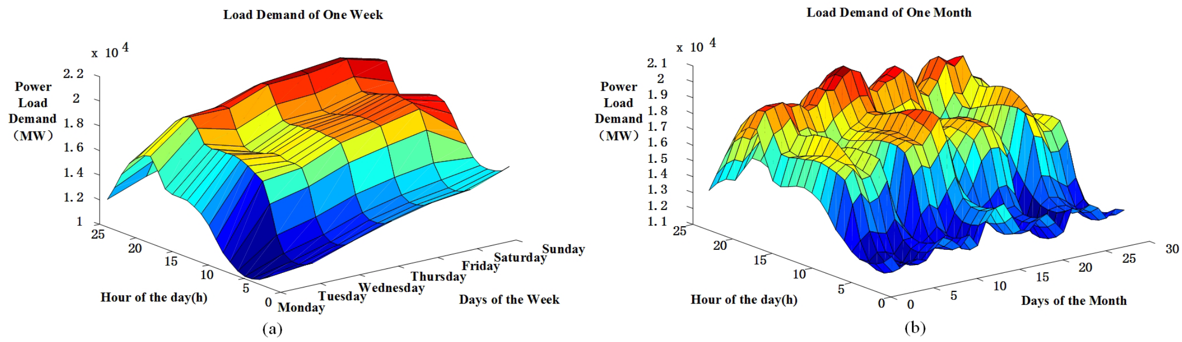

FIGURE 1. The weekly and monthly load demand in ISO-New England

4. Experiment.

4.1. Experimental dataset and environments. In this section, we are aiming at forecasting day-ahead hourly load and proposing comparative investigations on the ISO-New England hourly power load dataset and the Australian hourly power load dataset. More details about the datasets can be found in Table 1. In the experiments, the parameter C is selected from 2^{-25} to 2^{-25} in II-ESVR and IESVR, and the hidden node L is selected from 100 to 1000 with step 100 in all ELM models [9]. For SGD-BP, the number of hidden nodes is selected from 30 to 60 with step 10. All the experimental data should be normalized into $[0, 1]$ before the training process. All the experiments are conducted by ten rounds to obtain an average performance evaluation.

TABLE 1. Details of the power load datasets

Dataset	#Training data	#Testing data	#Total data
$ISO - NewEngland$	35064	17544	52608
$Australian$	52607	35041	87648

4.2. Validation on IESVR and integrated IESVR networks. The parameter P (IESVR number) in II-ESVR network is vital to the whole system considering the computational cost, complexity and stability of the network. Figure 2 shows the validation MAPE with varying single ESVR number. It can be observed that the MAPE will converge to a stable state with the number increasing. Considering the balance of precision and computation cost, we choose $P = 5$ with MAPE $= 1.30\%$.

It also illustrates the stability of II-ESVR and IESVR. Experiments on II-ESVR and single IESVR model are both carried out 10 rounds. II-ESVR network presents convincing improvement upon accuracy and stability. Besides, it should be mentioned that the computation time of II-ESVR is around 5 times higher than that of a single IESVR model. However, since each IESVR model is trained independently, all operations can be performed in parallel, thus reducing the total computing time.

4.3. Generalization performance on datasets. The forecasted residual errors (MAE) of II-ESVR, IESVR, OS-ELM and EOS-ELM models are compared on the actual load of ISO-England from Jun-01 to Jun-15 in 2008 and the actual load of Australia Jun-01 to Jun-15 in 2008. It can be clearly seen in Figure 3 that OS-ELM and EOS-ELM have shown serious deviations from the actual load in some period. However, II-ESVR is much stable in both two prediction jobs. More detail results can be obtained from Table 2. Six methods are evaluated with measure criteria of MAEs, RMSEs and MAPEs. Two activation functions of Sigmoid and Sin are used, and linear kernel is used in Online-SVR

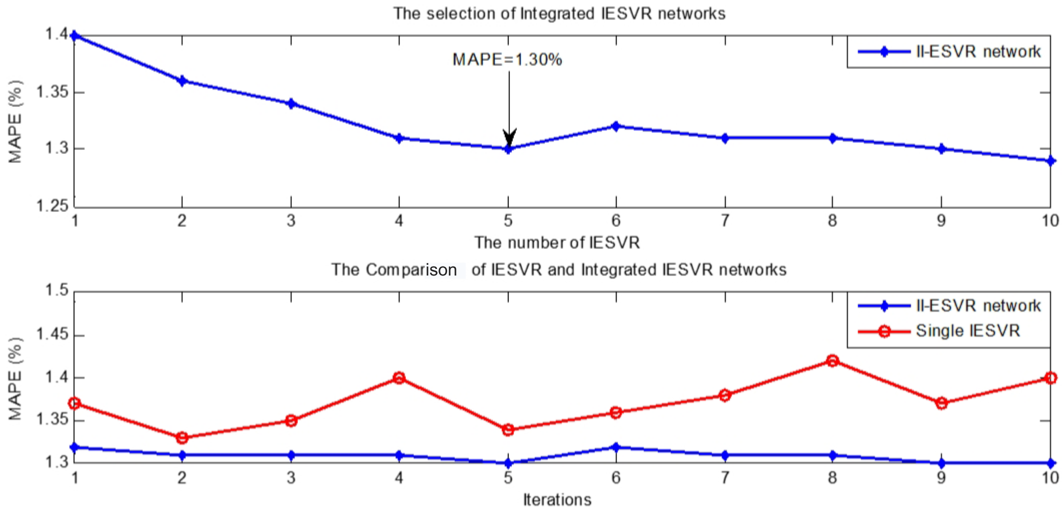

FIGURE 2. The selection of parameter P in II-ESVR network

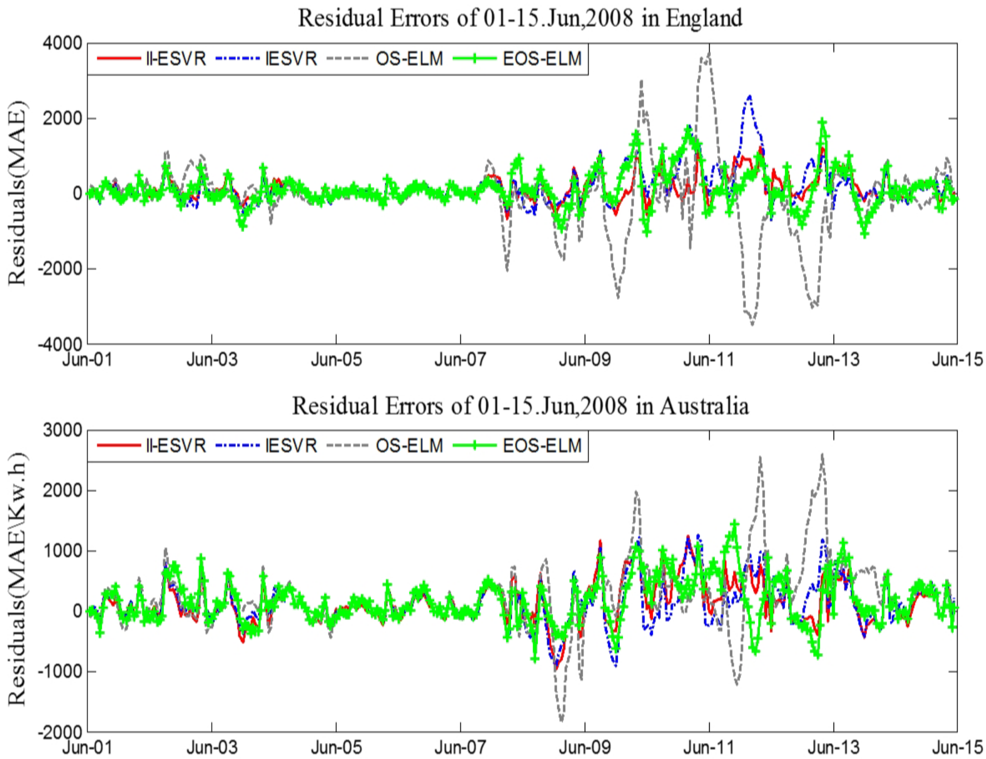

FIGURE 3. The forecasted residual errors of Jun-01 to Jun-15, 2008 in England and Jun-01 to Jun-15, 2008 in Australia

considering the computational cost. Experiment results show that II-ESVR has achieved best performance in most of these measures on the two public power load dataset. In all the experiments, IESVR wins the champion in training time while II-ESVR achieves highest precision. As for II-ESVR network consisting of 5 single IESVR models, the computation time of II-ESVR is around 5 times of a single IESVR, which is exactly acceptable by the STLF problem. Compared to OS-ELM and EOS-ELM, our II-ESVR approach has better generalization and stable performance. SGD-BP suffers much from local optimization and only works well for one by one, while II-ESVR can work well for chunk by chunk as desired size. Besides, II-ESVR can work for non-differentiable activation functions as well. Online-SVR only can work for one by one and has much

Table 2. Comparison of the experiment results between the six algorithms

Dataset	Activation	Algorithm	Node	Time (s)	MAE	RMSE	MAPE
ISO-New England	Sig	OS-ELM	400	15.51	236.14	330.05	1.58 ± 0.22
		EOS-ELM	400	88.65	222.55	321.80	1.52 ± 0.10
		IESVR	800	**7.29**	205.42	285.29	1.40 ± 0.06
		II-ESVR	800	36.75	**195.05**	**273.96**	$\mathbf{1.32\pm0.02}$
		SGD-BP	50	104.40	224.31	320.35	1.55 ± 0.16
	Sin	OS-ELM	400	16.06	231.20	331.80	1.55 ± 0.16
		EOS-ELM	400	90.17	215.82	314.56	1.45 ± 0.10
		IESVR	800	**7.05**	206.75	289.70	1.40 ± 0.05
		II-ESVR	800	36.25	**195.56**	**275.14**	$\mathbf{1.32\pm0.03}$
	linear	Online-SVR	–	>10 h	220.32	320.65	1.52
Australian	Sig	OS-ELM	500	22.85	146.43	196.15	1.74 ± 0.40
		EOS-ELM	500	108.56	142.28	188.25	1.70 ± 0.25
		IESVR	800	**8.07**	113.04	151.02	1.32 ± 0.16
		II-ESVR	800	42.11	**109.62**	**146.51**	$\mathbf{1.27\pm0.03}$
		SGD-BP	50	107.36	114.86	155.45	1.31 ± 0.02
	Sin	OS-ELM	500	20.70	179.30	231.68	2.08 ± 0.60
		EOS-ELM	500	102.35	170.26	221.65	1.92 ± 0.36
		IESVR	800	**7.94**	127.60	164.85	1.47 ± 0.11
		II-ESVR	800	40.75	**120.92**	**156.26**	$\mathbf{1.39\pm0.05}$
	linear	Online-SVR	–	>10 h	141.80	189.18	1.70

higher computation cost. The results indicate that Online-SVR is unsuitable for online STLF application. In summary, II-ESVR can implement STLF applications efficiently without sacrificing the accuracy and computation cost.

5. **Conclusion and Future Work.** In this paper, an incremental learning model for STLF based on II-ESVR approach has been proposed. By combining incremental learning techniques and integrated network strategy in ESVR, the proposed method can handle the problem of training incremental data and large-scale data, and it can efficiently improve the forecasting accuracy and overcome instability problem of single IESVR model by leveraging the advantage of II-ESVR model. Another benefit of the proposed model is that it has the capability to learn the variation trend of the power load incrementally without reserving the historical data. The experiment results reveal that the stability and generalized performance of II-ESVR approach are superior over the other competitive incremental learning algorithms. The proposed II-ESVR method can be efficiently applied in large-scale power load problem and dramatically reduce the computational cost and memory storage of energy management systems. Each single IESVR is trained independently, therefore, parallel computing technique will be applied to reducing computation cost in the future.

Acknowledgment. This project was supported by National Natural Science Foundation of China (61362030, 61201429), Xinjiang and Uygur Autonomous Regions University Science and Research Key Project (XJEDU2012I08).

REFERENCES

[1] Q. Hao, S. Dipti and K. Abbas, Short-term load and wind power forecasting using neural network-based prediction intervals, *IEEE Trans. Neural Networks and Learning Systems*, vol.25, no.2, pp.303-315, 2014.

[2] A. Adel, H. Mohammad and A. A. Akbar, Short term load forecasting using a hybrid intelligent method, *Knowledge-Based Systems*, vol.76, pp.139-147, 2015.

[3] S. Li, P. Wang and L. Goel, Short-term load forecasting by wavelet transform and evolutionary extreme learning machine, *Electric Power Systems Research*, vol.122, pp.96-103, 2015.

[4] Q. Liu, Q, He and Z. Shi, Extreme support vector machine classifier, *Proc. of the 12th Pacific Asia Conference on Advances in Knowledge Discovery and Data Mining*, Japan, 2008.
[5] W. Zhu, M. Jun and L. Qing, Robust regression with extreme support vectors, *Pattern Recognition Letters*, vol.45, no.1, pp.205-210, 2014.
[6] R. Meir and V. E. Maiorov, On the optimality of neural-network approximation using incremental algorithms, *IEEE Trans. Neural Networks*, vol.11, no.323, pp.65-81, 2000.
[7] G. Montana and F. Parrella, Learning to trade with incremental support vector regression experts, *Lecture Notes in Artificial Intelligence*, vol.5271, pp.591-598, 2008.
[8] H. Su, L. Zhang and S. Yu, Short-term traffic flow prediction based on incremental support vector regression, *The 3rd International Conference on Natural Computation*, vol.1, pp.640-645, 2007.
[9] N. Liang, G. B. Huang and P. Saratchandran, A fast and accurate online sequential learning algorithm for feedforward networks, *IEEE Trans. Neural Networks*, vol.17, no.6, pp.1411-1424, 2006.
[10] G. Li, M. Liu and M. Dong, A new online learning algorithm for structure-adjustable extreme learning machine, *Computers & Mathematics with Applications*, vol.60, no.3, pp.377-389, 2010.
[11] C. Cortes and V. Vapnik, Support vector networks, *Machine Learning*, pp.248-259, 2003.
[12] G. B. Huang, H. M. Zhou, X. Ding and R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics*, vol.42, no.2, pp.513-529, 2012.
[13] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, Least squares support vector machines, *World Scientific*, Singapore, 2002.