ICIC Express Letters
Part B: Applications
Volume **8**, Number **12**, December **2017**

ICIC International ©2017 ISSN 2185-2766
pp. **1653–1661**

SOFT-MEASUREMENT MODELING WITH MULTI-SCALE EVOLUTIONARY *V*-SUPPORT VECTOR MACHINE FOR ETHYLENE PRODUCT CONCENTRATION

Xueying Jiang[1], Chengli Su[1,*], Huiyuan Shi[1,2,*]
Ping Li[1,2] and Jiangtao Cao[1]

[1]School of Information and Control Engineering
Liaoning Shihua University
No. 1, Dandong Road, Fushun 113001, P. R. China
[*]Corresponding authors: shy723915@126.com; huiyuan_shi@lnpu.edu.cn

[2]School of Automation
Northwestern Polytechnical University
No. 1, Dongxiang Road, Xi'an 710129, P. R. China

Received June 2017; accepted August 2017

Abstract. *For the low predictive accuracy and poor stability of SVM incremental learning, a modeling method of multi-scale evolutionary v-support vector machine (ME-v-SVM) is proposed. Firstly, the global optimal model with multi-scale v-SVM is obtained by solving multi-scale quadratic programming problem, which can approximate the uneven distributed samples on multiple scales and overcome the poor predictive ability. Secondly, an evolutionary method based on the SVM incremental learning process is proposed, which solves the "Explosion" problem of training sample number caused by new samples constantly joining into training set and improves the stability of the SVM incremental learning process. Finally, the ME-v-SVM method is applied to the soft-measurement of the ethylene product concentration. The experimental results show that the proposed method has a better model performance than SVM and M-v-SVM method.*
Keywords: Multi-scale learning, Evolutionary SVM, Ethylene product concentration, Soft-measurement

1. **Introduction.** Ethylene is an important raw material in chemical industry and its concentration directly determines the quality of subsequent chemical products. However, in practical production, the concentration of ethylene product cannot be measured directly by the hardware sensor. It is usually tested offline in the laboratory. Unfortunately, offline analysis is too costly and often has 2-4 hours time-delay [1], which cannot meet the requirements of real-time control. An accurate and stable soft-measurement method [2] is required to enhance productive efficiency and reduce productive cost. Obviously, the traditional mechanism model is difficult to adapt the strong nonlinear process. Hence, many soft-measurement methods based on data-driven intelligent algorithm, such as genetic algorithm [3], neural network [4] and other intelligent algorithms [5], are proposed. However, these methods exceedingly depend on data, which cannot obtain satisfactory effect using limited samples.

Support vector machine (SVM) [6] is a machine learning algorithm based on Vapnik's structural risk minimization theory. It is one of the most popular methods in modeling research and has been successfully applied in practical processes. However, the standard SVM still has some bottlenecks. For example, the parameter ε used to control regression accuracy is hard to select. v-SVM [7] overcomes this defects because the parameter v is used to control the ε tube which simplifies the process of setting ε.

Above-mentioned SVM models based on single kernel are hard to get a satisfactory regression precision for predicting the uneven distribution samples. In [8], multi-scale

learning method was proposed to adapt the uneven distribution sample utilizing a liner combination of kernels with different widths, which enhances the approximate ability of SVM. The multi-scale kernel SVM [9,10] has been successfully applied to adaptive time sequence predictive question. Because the model can eliminate the regressive residuals on different scales, it has high predictive precision. Although the multi-scale kernel method can effectively improve model accuracy, it still belongs to offline modeling strategy without self-learning ability. The off-line modeling is very difficult to adapt the complex changes of working conditions. Some incremental learning methods [11] were proposed to obtain the constant change of model parameters by the variation of prediction error or the features of new samples. However, when new samples are continuously joined into the training set, model training speed will be badly affected.

In this paper, a new SVM with the combination of SVM incremental learning and multi-scale theory is proposed, called multi-scale evolutionary v-support vector machine (ME-v-SVM). Firstly, a multi-scale v-SVM (M-v-SVM) regression model is obtained by solving quadratic programming (QP) problem on different scales rather than linear combination of the different kernel scales. Secondly, a new "evolution" principle is proposed based on SVM incremental learning process. In statistics, the sample similarity is generally judged by Euclidean distance [12] and angle [13]. For SVM incremental learning process, the old sample most similar to new sample is replaced by the new sample. This method keeps quantity of training sample constant and avoids the unstable learning speed.

2. Multi-Scale Evolutionary v-SVM.

2.1. Multi-scale v-support vector machine.

A new support vector machine based on the multi-scale learning method is proposed. Taking the double-scale model as the example, the steps of the proposed algorithm are as follows.

Firstly, for a given training set $\{(x_i, y_i)_{i=1,2,\ldots,n}\}$, the samples in smooth region are fitted using large-scale kernel. The large-scale regression model is described as follows:

$$f_1(x) = w_1 K_1(x, x_i) + b_1 \tag{1}$$

where $K_1(x, x_i)$ is large-scale kernel, and x_i is a sample of the training set x.

Secondly, the samples in the region of severe changes are fitted using small-scale kernel. The objective is $\{(x_i, y_i) - f_1(x_i)\}_{i=1,2,\ldots,n}$. The small-scale regression model is expressed as:

$$f_2(x) = w_2 K_2(x, x_i) + b_2 \tag{2}$$

Finally, the double-scale model is viewed as:

$$f(x) = w_1 K_1(x, x_i) + w_2 K_2(x, x_i) + b_1 + b_2 \tag{3}$$

Based on the multi-scale learning idea and SVM principle, the solution of Equation (3) is obtained by solving the following optimization problem.

$$\min \frac{1}{2} \|w_1\|^2 + \frac{1}{2} \|w_1\|^2 + \frac{C_1}{n} \sum_{i=1}^{n} (\xi_{1i} + \xi_{1i}^*) + \frac{C_2}{n} \sum_{i=1}^{n} (\xi_{2i} + \xi_{2i}^*) + C_1 v_1 \varepsilon_1 + C_2 v_2 \varepsilon_2 \tag{4}$$

$$\text{s.t.} \begin{cases} w_1 \phi_1(x_i) + b_1 - y_i \le \varepsilon_1 + \xi_{1i} \\ y_i - w_1 \phi_1(x_i) - b_1 \le \varepsilon_1 + \xi_{1i}^* \\ w_2 \phi_2(x_i) + b_2 - (y_i - w_1 \phi_1(x_i) - b_1) \le \varepsilon_2 + \xi_{2i} \\ (y_i - w_1 \phi_1(x_i) - b_1) - w_2 \phi_2(x_i) - b_2 \le \varepsilon_2 + \xi_{2i}^* \end{cases} \tag{5}$$

where C_1 and C_2 are penalty factors; ξ_{1i}, ξ_{1i}^*, ξ_{2i}, ξ_{2i}^* are slack variables; $\phi(x)$ is the nonlinear mapping about input data to high-dimensional feature space. v_1, v_2, ε_1, ε_2 are

the parameters in scales 1 and 2 respectively. The Lagrange function of Equation (4) is structured as follows:

$$\begin{aligned}\min L = &\frac{1}{2}\|w_1\|^2 + \frac{1}{2}\|w_1\|^2 + \frac{C_1}{n}\sum_{i=1}^{n}(\xi_{1i}+\xi_{1i}^*) + \frac{C_2}{n}\sum_{i=1}^{n}(\xi_{2i}+\xi_{2i}^*)\\ &+ C_1v_1\varepsilon_1 + C_2v_2\varepsilon_2 - \sum_{i=1}^{n}\alpha_{1i}(\varepsilon_1+\xi_{1i}+y_i-w_1\phi_1(x_i)-b_1)\\ &- \sum_{i=1}^{n}\alpha_{1i}^*(\varepsilon_1+\xi_{1i}^*-y_i+w_1\phi_1(x_i)+b_1)\\ &- \sum_{i=1}^{n}\alpha_{2i}(\varepsilon_2+\xi_{2i}-(y_i-w_1\phi_1(x_i)-b_1)-w_2\phi_2(x_i)-b_2)\\ &- \sum_{i=1}^{n}\alpha_{2i}^*(\varepsilon_2+\xi_{2i}^*+w_2\phi_2(x_i)+b_2-(y_i-w_1\phi_1(x_i)-b_1))\end{aligned} \tag{6}$$

where α_{1i}^* and α_{2i}^* are Lagrange multiplier vectors. By the Karush-Kuhn-Tucker (KKT) condition, the partial derivatives with regard to w_1, w_2, b_1, b_2, ε_1, ε_2, ξ_{1i}^*, ξ_{2i}^* are calculated as:

$$\begin{aligned}\frac{\partial L}{\partial w_1} &= w_1 + \sum_{i=1}^{n}(\alpha_{1i}-\alpha_{1i}^*+\alpha_{2i}-\alpha_{2i}^*)\phi_1(x_i) = 0,\\ \frac{\partial L}{\partial w_2} &= w_2 + \sum_{i=1}^{n}(\alpha_{2i}-\alpha_{2i}^*)\phi_2(x_i) = 0\end{aligned} \tag{7}$$

$$\frac{\partial L}{\partial b_1} = \sum_{i=1}^{n}(\alpha_{1i}-\alpha_{1i}^*+\alpha_{2i}-\alpha_{2i}^*) = 0, \quad \frac{\partial L}{\partial b_2} = \sum_{i=1}^{n}(\alpha_{2i}-\alpha_{2i}^*) = 0 \tag{8}$$

$$\frac{\partial L}{\partial \varepsilon_1} = C_1v_1 - \beta_1 - \sum_{i=1}^{n}(\alpha_{1i}+\alpha_{1i}^*) = 0, \quad \frac{\partial L}{\partial \varepsilon_2} = C_2v_2 - \beta_2 - \sum_{i=1}^{n}(\alpha_{2i}+\alpha_{2i}^*) = 0 \tag{9}$$

$$\frac{\partial L}{\partial \xi_{1i}} = \frac{C_1}{n} - \sum_{i=1}^{n}\alpha_{1i} = 0, \quad \frac{\partial L}{\partial \xi_{1i}^*} = \frac{C_1}{n} - \sum_{i=1}^{n}\alpha_{1i}^* = 0 \tag{10}$$

$$\frac{\partial L}{\partial \xi_{2i}} = \frac{C_2}{n} - \sum_{i=1}^{n}\alpha_{2i} = 0, \quad \frac{\partial L}{\partial \xi_{2i}^*} = \frac{C_2}{n} - \sum_{i=1}^{n}\alpha_{2i}^* = 0 \tag{11}$$

Substituting the aforementioned derivative equations into Equation (6), we obtain the dual question of Equation (4) as follows:

$$\begin{aligned}\max\ W = &\sum_{i=1}^{n} y_i\left(\alpha_{1i}^*-\alpha_{1i}+\alpha_{2i}^*-\alpha_{2i}\right) - \frac{1}{2}\sum_{i,j=1}^{n}\Big[\left(\alpha_{1i}^*-\alpha_{1i}+\alpha_{2i}^*-\alpha_{2i}\right)\left(\alpha_{1j}^*-\alpha_{1j}\right.\\ &\left.+\alpha_{2j}^*-\alpha_{2j}\right)K_1(x_i,x_j) + \left(\alpha_{2i}^*-\alpha_{2i}\right)\left(\alpha_{2j}^*-\alpha_{2j}\right)K_2(x_i,x_j)\Big]\end{aligned} \tag{12}$$

$$\text{s.t.}\ \begin{cases}\sum_{i=1}^{n}(\alpha_{1i}^*-\alpha_{1i}+a_{2i}^*-\alpha_{2i}) = 0,\ \sum_{i=1}^{n}(\alpha_{2i}-a_{2i}^*) = 0\\ 0 \le \alpha_{1i}^* \le \frac{C_1}{n},\ 0 \le \alpha_{2i}^* \le \frac{C_2}{n},\ \sum_{i=1}^{n}(a_{1i}+a_{1i}^*) \le C_1v_1,\ \sum_{i=1}^{n}(a_{2i}+a_{2i}^*) \le C_2v_2\end{cases} \tag{13}$$

In the dual Equation (12), the function $K(x_i, x_j)$ is the kernel matrix which is equivalent to the dot product $< \phi(x_i), \phi(x_j) >$. The radial basis function (RBF) is selected as kernel

function. The final regression model $f(x)$ of double-scale v-SVM depends on the Lagrange multipliers a_{1i}^* and a_{2i}^* as follows.

$$f(x) = \sum_{i=1}^{n} \left(\alpha_{1i}^* - \alpha_{1i} + \alpha_{2i}^* - \alpha_{2i}\right) K_1(x, x_i) + \sum_{i=1}^{n} (\alpha_{2i}^* - \alpha_{2i}) K_2(x, x_i) + b \tag{14}$$

$$b = \frac{1}{n} \sum_{i=1}^{n} [y_i - f(x_i)] \tag{15}$$

It is not hard to see that the regression model (14) is not the simple liner combination on different scales (kernel width). Multi-scale learning models can be extended on the basis of aforementioned derivation.

2.2. **An evolutionary method based on the SVM incremental learning process.** This section presents a kind of "evolution" incremental learning method which can improve the self-learning ability of multi-scale v-SVM. To keep the training sample size constant, the idea of evolution is incorporated to SVM incremental learning process. When the new sample is added to training set, old sample will be replaced by the new sample. The process can be described as follows.

Given a training set: $N = \{(x_{ij}, y_{ij})\}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, t$, where n is sample number and t is the dimensional character of sample. For example, there are two samples $x_k = (x_{k1}, x_{k2}, \ldots, x_{kt})$ and $x_l = (x_{l1}, x_{l2}, \ldots, x_{lt})$, $k \neq l$. The difference of distance variance can be estimated by Euclidean distance as follows.

$$dist(x_k, x_l) = \sqrt{\sum_{j=1}^{t} (x_{kj} - x_{lj})} \tag{16}$$

where x_{kj} is the jth character of x_k and x_{lj} the jth character of x_l.

The angle variance is computed by cosine similarity:

$$sim(x_k, x_l) = \cos\theta = \frac{x_k \cdot x_l}{\|x_k\| \cdot \|x_l\|} \tag{17}$$

When the new sample was added to training set, the distance between new sample and training samples is calculated:

$$dist(c, N_i) = \sqrt{\sum_{j=1}^{t} (c_j - N_{ij})^2} \tag{18}$$

where c is the new sample, N the training set. According to Euclidean distance, these training samples close to c may be a higher similarity with c. Hence, we define Equation (19) as follows:

$$dist(c, N_i) < \eta \tag{19}$$

where η is potential factor whose value could determine the size of potential set. The larger η means a greater potential size of the set, and vice-versa. The training samples corresponding to Equation (19) are defined as potential set $\{P_i\}_{i=1,2,\ldots,g}$ in which P_i is potential vector. The angel similarity between c and P_i is analyzed as follows.

$$sim(c, P_i) = \frac{c \cdot P_i}{\|c\| \cdot \|P_i\|} \tag{20}$$

$sim(c, P_M)$ is as the maximum of $\{sim(c, P_i)\}$:

$$sim(c, P_M) = \max\{sim(c, P_i)\}_{i=1,\ldots,g} \tag{21}$$

As the description of Equation (16) to Equation (21), P_M is the most similar sample with c. The proposed multi-scale evolutionary v-SVM (ME-v-SVM) algorithm is summarized as follows.

Step 1: Train multi-scale v-SVM on set N.

Step 2: Add new sample c to training set N, and generate new training set $N^* = \{N, c\}$.

Step 3: Calculate sample P_M which is the most similar to c, and get rid of it from N^*. Enable N^* and N with the same number of samples.

Step 4: Train multi-scale v-SVM on N^* and obtain a new ME-v-SVM model.

Step 5: Return Step 2 to Step 4.

3. Soft-Measurement for Ethylene Product Concentration.

3.1. **General description.** Ethylene distillation tower (C-1440 tower) is key equipment for ethylene productive unit. The process is shown in Figure 1.

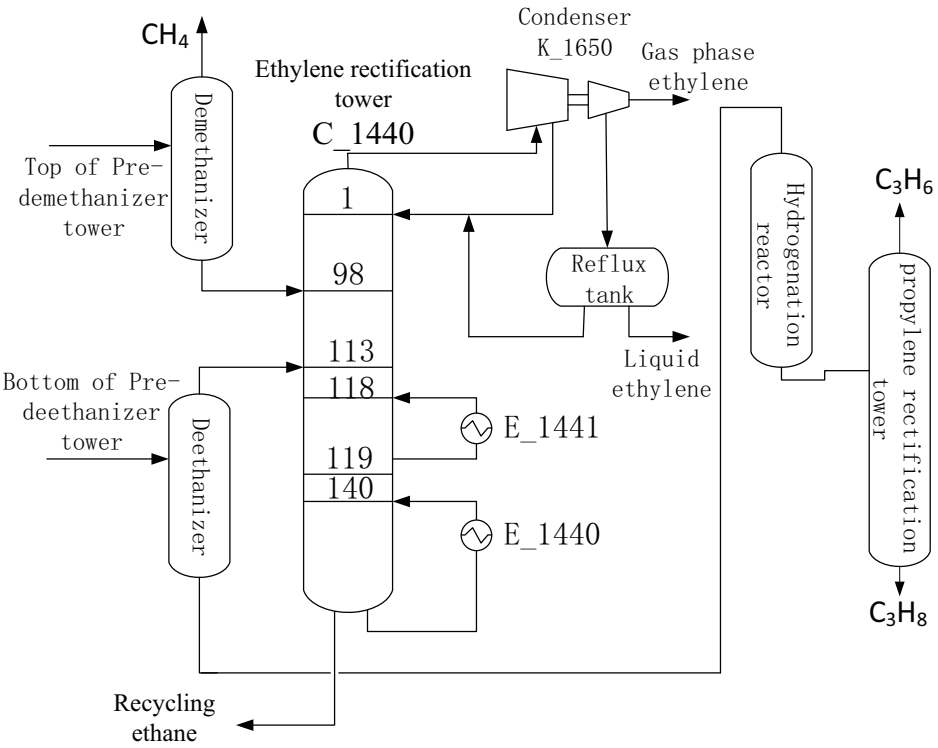

FIGURE 1. Process flow chart for C-1440 tower

C-1440 tower is composed of tower shell, layer plate, reboiler and other important equipment. The main component of feed is C_2. The feed, from demethanizer bottom and deethanizer top, is respectively sent into 113th and 98th layer plate of C-1440 tower. The gas component rises. Products in the top of tower are mainly ethylene fractions. The liquid component flows downwards, and its main component is ethane. When liquid component flows downwards to 119th layer plates, some of the liquid components are produced and fed into the middle reboiler (E-1411) to be heated through propylene heating medium. The upper gas is then returned to the C-1440 tower from the 118th layer plate. The remaining liquid components continue to flow downwards to the bottom reboiler (E-1440) by the same method in middle reboiler. The vapors for tower top are condensed by condenser (K-1650) with propylene refrigerant and fed to reflux vessel (V-1690). Then the product is sent into the first layer plate of C-1440 tower by the pump (V-1690).

3.2. **Hardware system architecture.** The control system of the ethylene distillation tower is CENTUM VP distributed control system (DCS). The system achieves process control through the field control station (FCS). The industrial implementation of soft-measurement modeling method is developed in an upper computer based on OPC technology. The upper computer adopts APC-ISYS software, while the lower computer is

CENTUM VP software which is connected with the ethylene distillation tower. The DCS system communicates with the upper computer through the OPCServer. The proposed ME-v-SVM method is compiled in VbScript of the lower computer. The ethylene product concentration can be calculated by the proposed soft-measurement modeling method. The result is sent back to the FCS by OPCServer. The hardware system architecture is shown in Figure 2 as follows.

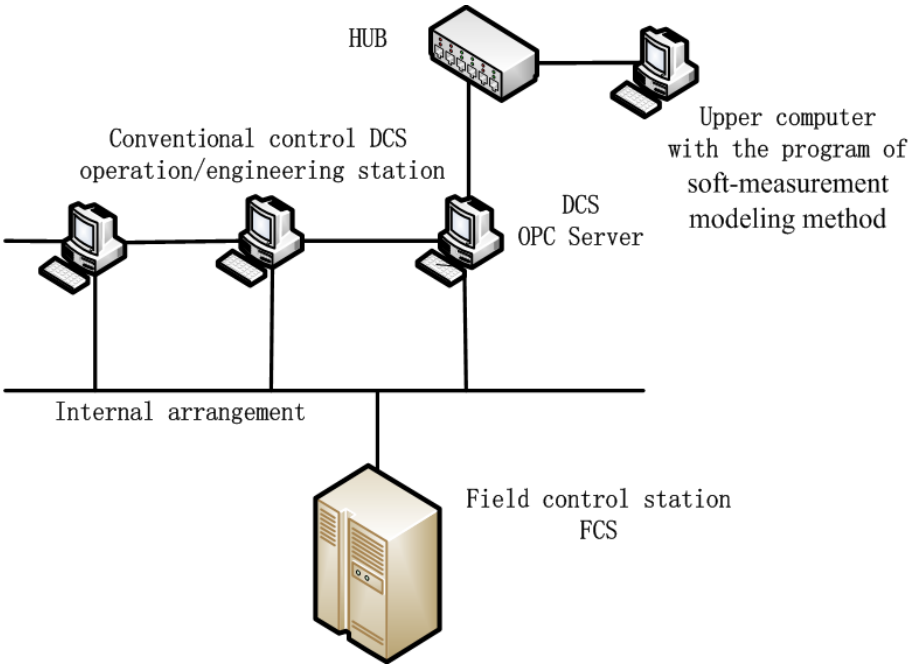

FIGURE 2. Hardware system architecture

3.3. **Soft-measurement modeling.** Based on the mechanism and operation experience for ethylene distillation process, 9 process variables defined as auxiliary variables including the feed, the temperature of 118th sensitive plate, temperature of the flow of feed tower bottom, the flow of tower top, reflux, pressure of tower top, temperature of tower top, temperature of tower kettle, liquid level of tower bottom. These auxiliary variables are as the input of soft-measurement model. Ethylene product concentration is as output. Process of soft-measurement modeling with ME-v-SVM for ethylene product concentration is shown in Figure 3.

The 1100 samples are collected from field C-1440 tower. We use firstly 1000 samples as training set to build ethylene concentration model, and then utilize the surplus 100 samples to test the performance of the model. Here three methods including ME-v-SVM, M-v-SVM and SVM are compared through establishing the soft measurement model of ethylene products. MSE, MAE and MAXEE are employed for accuracy evaluation. For Jackknifing, all the samples in the benchmark seek out orderly and the prediction model is examined by training residual samples between training set and test set. According to this process, it rules out the "memory" effect. Therefore, we use the Jackknife test to choose parameters where $C = 35$, $v = 0.02$, $\eta = 0.5$, $\delta = 3.1$.

The prediction results and error of ethylene concentration on different soft-measurement methods are shown in Figures 4-6. Theses figures illustrate that ME-v-SVM has better prediction results and smoother error for ethylene concentration compared to SVM and M-v-SVM. From Figure 6, we can see that there is a smaller error between the initial data and predicted data using the ME-v-SVM which has stronger generalization ability and higher prediction accuracy.

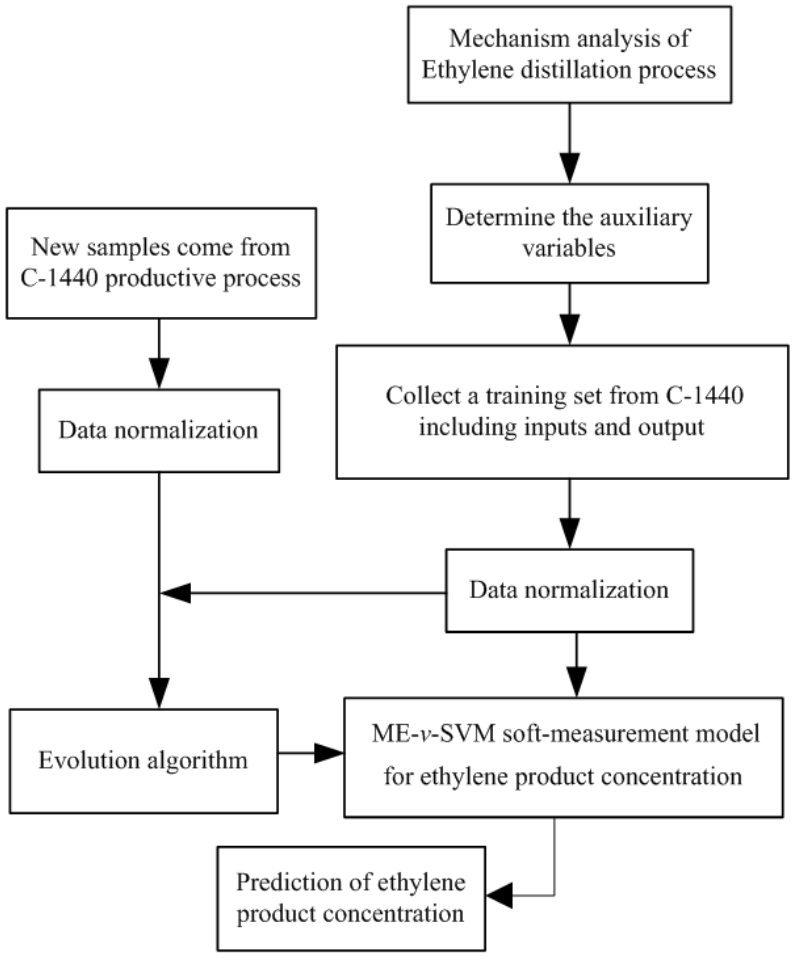

FIGURE 3. Process of soft-measurement modeling with ME-v-SVM

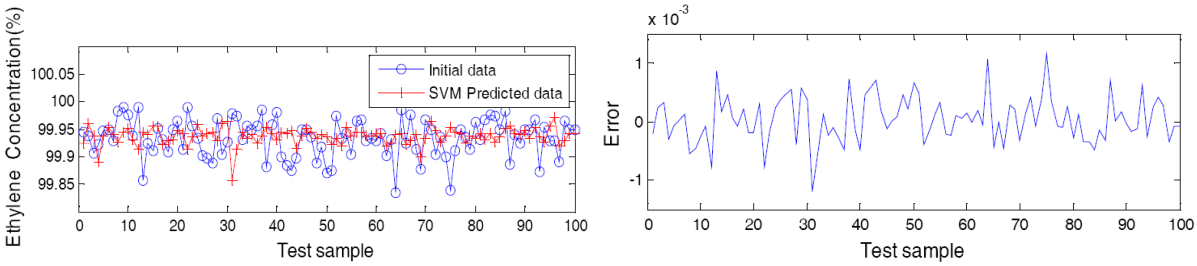

FIGURE 4. Predictive results with SVM

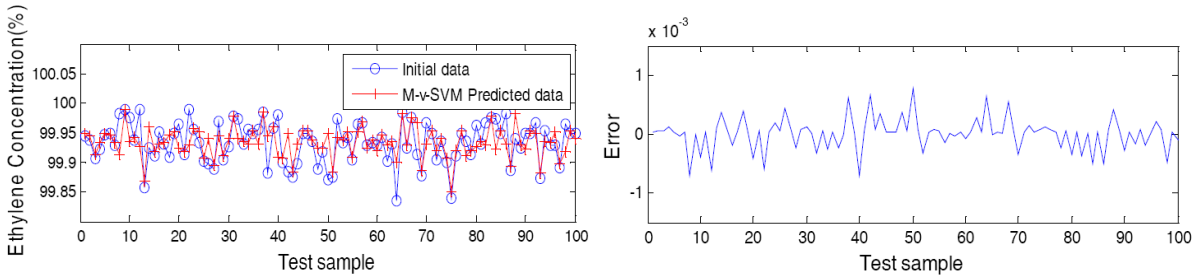

FIGURE 5. Predictive results with M-v-SVM

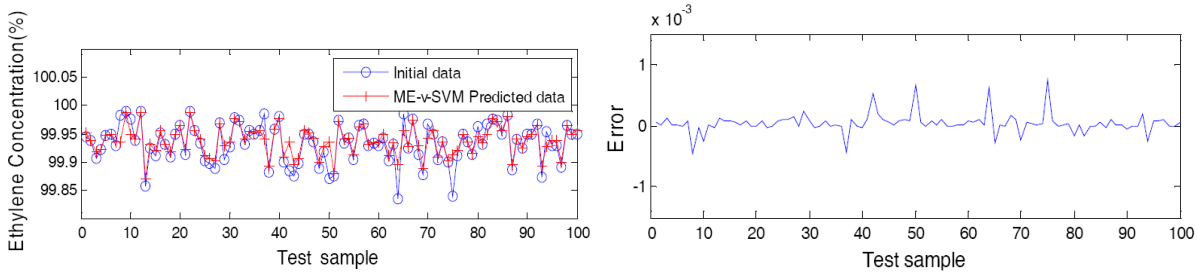

FIGURE 6. Predictive results with ME-v-SVM

TABLE 1. Predictive error with various methods incremental learning

Algorithm	MSE	MAXE	MAE
SVM	1.592×10^{-7}	1.32×10^{-3}	3.153×10^{-4}
M-v-SVM	7.243×10^{-8}	8.72×10^{-4}	1.719×10^{-4}
ME-v-SVM	3.028×10^{-8}	8.63×10^{-4}	0.944×10^{-4}

TABLE 2. Time consumption with various methods incremental learning

Algorithm	1st	10th	50th	100th
SVM	2.22s	2.36s	2.77s	3.47s
M-v-SVM	3.14s	3.44s	4.15s	5.24s
ME-v-SVM	3.39s	3.39s	3.39s	3.39s

The predictive error and predictive time on different methods are listed in Table 1 and Table 2. In Table 1, MSE, MAXE and MAE are 1.592×10^{-7}, 1.32×10^{-3}, 3.153×10^{-4} with the SVM respectively. Using the M-v-SVM method, MSE, MAXE and MAE are 7.243×10^{-8}, 8.72×10^{-4}, 1.719×10^{-4} respectively. From the information, M-v-SVM method has better performance than SVM. However, employing ME-v-SVM method, the better results can be seen. MSE, MAXE and MAE are 3.028×10^{-8}, 8.63×10^{-4} and 0.944×10^{-4} respectively. Three errors compared with SVM decreased by 80.9%, 34.8% and 70.1%.

Table 2 shows a time analysis for ethylene concentration. Clearly, with the increase of prediction times, the time consumptions of SVM and M-v-SVM are constantly prolonged. From the 1st to 100th prediction, the time consumptions of SVM and M-v-SVM increase by 1.25s and 2.10s, respectively. However, it is worth mentioning that ME-v-SVM keeps the same time consumption at each prediction time.

4. **Conclusion.** In this work, a new ME-v-SVM is proposed to solve the defects of SVM incremental learning algorithm, which combines multi-scale learning method and evolution learning method. Experimental results in ethylene distillation product quality show that ME-v-SVM has high model accuracy, good adaptable ability, and stable performance. Based on the multi-scale v-SVM method, we shall try to improve the structure of the SVM model or to enhance the training speed of the model in the future research.

Acknowledgment. This paper is supported by the National Natural Science Foundation of China (No. 61673199), the Scientific, National Science Foundation of Liaoning Province (No. 2013020024), and the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No. ICT1600354).

REFERENCES

[1] M. M. Zhang and X. G. Liu, A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction, *Chemometrics & Intelligent Laboratory Systems*, vol.126, pp.83-90, 2013.

[2] Z. Tang, H. Zhang and J. Li, Adaptive soft measurement, method for the furnace exit gas temperature based data-driven method, *Proc. of 2016 IEEE Conference on Control and Decision*, Yinchuan, China, pp.4681-4685, 2016.

[3] M. Huang, Y. Ma and J. Wan, A sensor-software based on a genetic algorithm-based neural fuzzy system for modeling and simulating a wastewater treatment process, *Applied Soft Computing*, vol.27, pp.1-10, 2015.

[4] C. A. Duchanoy, M. A. Moreno-Armendáriz and L. Urbina, A novel recurrent neural network soft sensor via a differential evolution training algorithm for the tire contact patch, *Neurocomputing*, vol.235, pp.71-82, 2017.

[5] Z. Cheng and X. Liu, Optimal online soft sensor for product quality monitoring in propylene polymerization process, *Neurocomputing*, vol.149, pp.1216-1224, 2015.

[6] V. N. Vapnik, An overview of statistical learning theory, *IEEE Trans. Neural Networks*, vol.10, no.5, pp.988-999, 1999.

[7] B. Scholkopf, A. J. Smola and R. C. Williamson, New support vector algorithms, *Neural Computation*, vol.12, no.5, pp.1207-1245, 2000.

[8] J. Bao, Y. Chen and L. Yu, A multi-scale kernel learning method and its application in image classification, *Neurocomputing*, vol.257, pp.16-23, 2017.

[9] D. Zheng, J. Wang and Y. Zhao, Non-flat function estimation with a multi-scale support vector regression, *Neurocomputing*, vol.70, no.1, pp.420-429, 2006.

[10] R. S. Jia, H. M. Sun and Y. J. Peng, Automatic event detection in low SNR microseismic signals based on multi-scale permutation entropy and a support vector machine, *Journal of Seismology*, vol.21, no.4, pp.735-748, 2017.

[11] B. Gu, V. S. Sheng and Z. Wang, Incremental learning for v-support vector regression, *Neural Networks*, vol.67, pp.140-150, 2015.

[12] O. Kisi and C. Ozkan, A new approach for modeling sediment-discharge relationship: Local weighted linear regression, *Water Resources Management*, vol.31, no.1, pp.1-23, 2017.

[13] L. Yi and J. Chen, Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes, *Journal of Process Control*, vol.23, no.6, pp.793-804, 2013.