Preprints, 10th IFAC International Symposium on
Advanced Control of Chemical Processes
Shenyang, Liaoning, China, July 25-27, 2018

Selective Ensemble Least Square Support Vector Machine with Its Application

J. Tang, J. F. Qiao*. Z. Liu, Z. W. Wu, X. J. Zhou.**
G. Yu, J. J. Zhao*,******

** Faculty of Information Technology, Beijing University of Technology, Beijing, China, (e-mail: freeflytang@ bjut.edu.cn).*
*** State Key Laboratory of Synthetical Automation for Process Industries, Northeaster University, Shenyang, 110004, China, (e-mail: liuzhuo@ise.edu.cn)*
**** State Key Laboratory of Process Automation in Mining & Metallurgy, Beijing, China, (e-mail: yugang@bgrimm.com)*
*****Beijing Key Laboratory of Process Automation in Mining & Metallurgy, Beijing,China, (e-mail: zhao_jj@bgrimm.com)}*

Abstract: Kernel-based modeling methods have been used widely to estimate some difficulty-to-measure quality or efficient indices at different industrial applications. Least square support vector machine (LSSVM) is one of the popular ones. However, its learning parameters, i.e., kernel parameter and regularization parameter, are sensitive to the training data and the model's prediction performance. Ensemble modeling method can improve the generalization performance and reliability of the soft measuring model. Aim at these problems, a new adaptive selective ensemble (SEN) LSSVM (SEN-LSSVM) algorithm is proposed by using multiple learning parameters. Candidate regularization parameters and candidate kernel parameters are used to construct many of candidate sub-sub-models based on LSSVM. These sub-sub-models based on the same kernel parameter are selected and combined as candidate SEN-sub-models by using branch and bound-based SEN (BBSEN). By employing BBSEN at the second time, these SEN-sub-models based on different kernel parameters are used to obtain the final soft measuring model. Thus, multiple kernel and regularization parameters are adaptive selected for building SEN-LSSVM model. UCI benchmark datasets and mechanical frequency spectral data are used to validate the effectiveness of this method.

Keywords: Selective ensemble modeling, soft measuring, least square support vector machine (LSSVM), learning parameters selection.

1. INTRODUCTION

Some difficulty-to-measure production quality and quantity indices cannot be directly measured by hardware sensors or calculated by first principal models [1]. Manual off-line laboratory analysis methods or experienced practitioners inference approaches are constantly used in several practical industries [2 , 3]. These methods limit the operation optimization and control of such industrial processes [4]. Data-driven-based soft sensing techniques are used as alternative approaches for measuring these process parameters in broad fields [5]. The commonly used methods include artificial neural networks and support vector machines (SVM). The least square support vector machines (LSSVM) can simplify the quadratic program (QP) problem of the SVM to solve a set of linear equations in the sense of sub optimum. However, the learning parameters, i.e., kernel parameter and regularization parameter, are data depended. Some optimization methods have been used to address such problem [6]. Many GA-based approaches address learning parameters' identification problem [7 , 8 , 9]. Recently, some new optimization algorithms are proposed to estimate the unknown parameters [10,11]. However, long time has to be used to search sub optimum solution and only single-model is constructed.

By combining several sing-models with linear or nonlinear method, ensemble modeling can improve the generalization, validity and reliability of the prediction model. The first problem in ensemble modeling is to construct ensembles. The normal used methods include sampling the training examples, manipulating the input features, manipulating the output targets, and injecting randomness. Different modeling data should select different ensemble methods. The optimized weighting coefficients calculation approach of neural network ensemble model is proposed with the assumption that the rows and columns of correlation co-efficient are linear independent [12]. Selective ensemble (SEN) modelling based on "re-sample training samples" ensemble construction method validates that ensemble many of the available sub-models can obtain better performance than an ensemble all of them [13]. Branch and bound (BB) algorithm is always used to find optimal solutions of various optimization problems. Thus, BB-based SEN (BBSEN) method is used to select sub-models and calculate the weighting coefficients [3]. One or more ensemble construction approaches can also be used in one SEN modelling algorithm for improving prediction performance further [14]. Aim at mill load parameters forecasting problem based on multi-source multi-scale frequency spectral data, selective information fusion strategy by integrating SEN and adaptive weighted fusion (AWF) is proposed, which are used to selective fuse the interesting mechanical sub-signals [15,16,17].

Only LSSVM-based SEN modelling methods is focused in this study. In [18], fuzzy c-means cluster and LSSVM are integrated to obtain training sub-samples and sub-models, and partial least squares (PLS) is applied as the combination strategy. Most of LSSVM ensemble learning methodologies are based on some heuristic algorithms. For example, evolutionary programming (EP)-based asymmetric weighted LSSVM ensemble model is constructed. Recently, a multi-level approach in an ensemble of LSSVM by using genetic algorithm is proposed, which considers the input feature selection, ensemble sub-model construction, and their selection and combination together [19]. However, how to adaptive select their learning parameters isn't addressed.

Motivated by the above problems, a new adaptive SEN-LSSVM by using learning parameters from candidate ones is proposed. At first, candidate regularization parameters and kernel parameters are used to construct many of candidate sub-sub-models based on LSSVM. Then, these sub-sub-models based on the same kernel parameter are selected and combined as candidate SEN-sub-models by using BBSEN. Finally, by employing BBSEN at the second time, these SEN-sub-models based on different kernel parameters are selected and combined. UCI benchmark datasets and mechanical frequency spectral data are used to validate the effectiveness of the proposed method.

2. RELATE WORKS

2.1 Least Square Support Vector Machine (LSSVM) and Its Learning Parameter Selection

We have the following function estimation problem

$$y(\boldsymbol{x}) = \boldsymbol{W}^{\mathrm{T}}\Phi(\boldsymbol{x}) + b \tag{1}$$

where $\Phi(*)$ maps $\{\boldsymbol{x}_l\}_{l=1}^{k}$ to a higher dimensional feature space, $\mathbf{W}$ is the weight vector, and b is the bias. The following problem would be solved with LSSVM.

$$\begin{cases} \min\limits_{W \cdot b} \quad J_{\mathrm{LSSVM}} = \dfrac{1}{2}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} + \dfrac{1}{2}C_{\mathrm{LSSVM}}\sum\limits_{l=1}^{k}\xi_l^2 \\ s.t.: \quad y_l = \boldsymbol{W}^{T}\Phi(\boldsymbol{x}) + b + \xi_l \end{cases} \tag{2}$$

where J_{LSSVM} is the objective function, C_{LSSVM} is a regularization parameter that used to decide trade-off between model complexity and approximation accuracy, and ξ_l is the approximation error. The Lagrange form of Eq. (2) is

$$\begin{aligned} \mathrm{L}(\boldsymbol{W}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}) = & \frac{1}{2}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} + \frac{1}{2}C_{\mathrm{LSSVM}}\sum_{l=1}^{k}\xi_l^2 \\ & -\sum_{l=1}^{k}\alpha_l[\boldsymbol{W}^{\mathrm{T}}\Phi(\boldsymbol{x}) + b + \xi_l - y_l] \end{aligned} \tag{3}$$

where $\boldsymbol{\alpha} = \{\alpha_l\}_{l=1}^{k}$ are Lagrange multipliers.

The solution is given by solving the $(k+1)\times(k+1)$ linear equation $\mathbf{A\Theta} = \mathbf{Y}'$ with $\mathbf{A} = \begin{bmatrix} 0 & \tilde{1}^{\mathrm{T}} \\ \tilde{1} & \mathbf{\Omega} + \dfrac{1}{C_{\mathrm{LSSVM}}}\mathbf{I} \end{bmatrix}$, $\mathbf{\Theta} = \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix}$, $\mathbf{Y}' = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}$, $\tilde{1} = [1, 1, \Lambda, 1]^{\mathrm{T}}$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \Lambda, \alpha_k]^{\mathrm{T}}$, and $\mathbf{y} = [y_1, y_2, \Lambda, y_k]^{\mathrm{T}}$. In which, kernel $\mathbf{\Omega}$ follows Mercer's condition. It is realized by kernel trick with the selected kernel type and kernel parameter, i.e.,

$$\Omega = f\left(K_{\mathrm{LSSVM}}, \{\boldsymbol{x}_l\}_{l=1}^{k}\right) \tag{4}$$

where K_{LSSVM} is the kernel parameter.

Therefore, two learning parameters, i.e., K_{LSSVM} and C_{LSSVM}, should be selected based on characteristics of the modelling data. The single LSSVM model can be denoted as:

$$\hat{y} = f_{\sin}(K_{\mathrm{LSSVM}}, C_{\mathrm{LSSVM}}; \{\boldsymbol{x}_l\}_{l=1}^{k}) \tag{5}$$

2.2 Ensemble Modelling and Weighting Coefficient Calculation

The prediction output of the lth sample based on J_{En} ensemble sub-models is given by

$$\bar{f}(x_l) = \sum_{j=1}^{J_{\mathrm{En}}}\alpha_j f_{j\ \ j}^{\mathrm{Sin}}\left(x_l\right) \tag{6}$$

where α_j is the weighting coefficient.

Defining misfit function of $f_j(x_l)$ as $m_j(x_l) \equiv y_l - f_j(x_l)$, Eq. (5) is rewritten as:

$$\bar{f}(x_l) = \sum_{j=1}^{J_{\mathrm{En}}}\alpha_j f_j^{\mathrm{Sin}}\left(x_l\right) = y_l + \sum_{j=1}^{J_{\mathrm{En}}}\alpha_j m_j(x_l) \tag{7}$$

By defining symmetric correlation matrix $C_{js} \equiv E[m_j(x)m_s(x)]$, the optimum α_j is calculated by minimizing MSE of $\bar{f}(x_l)$, i.e.,

$$\begin{aligned} \boldsymbol{\alpha}_{\mathrm{opt}} &= \arg\min\left(\mathrm{MSE}[\bar{f}(x_l)]\right) \\ &= \arg\min\left(\sum_{j,s}\alpha_j\alpha_s C_{js}\right) \end{aligned} \tag{8}$$

Then, langrange multipliers method can be used with the constraint $\sum\alpha_j = 1$. The $\alpha_{\mathrm{opt},j^*}$ is used to denote the j^*th variable of $\boldsymbol{\alpha}_{\mathrm{opt}}$ [12],

$$\alpha_{\mathrm{opt},j^*} = \frac{\sum_j C_{sj}^{-1}}{\sum_{j^*}\sum_j C_{j^*j}^{-1}} \tag{9}$$

It shows that the optimal weights of these ensemble sub-models can be calculated with the above analytical solution method. However, it is difficult to be realized for the practical industrial problem because of the undependable assumption among prediction outputs of ensemble sub-models.

2.3 SEN-based LSSVM Method

By only selecting some ensemble sub-model, the SEN-based LSSVM model can be denoted as:

$$\overline{f}_{\mathrm{SEN}}(x_l)=\sum_{j=1}^{J_{\mathrm{SEN}}}\alpha_j f_j^{\mathrm{Sin}}(K_{\mathrm{LSSVM}},C_{\mathrm{LSSVM}};x_l) \quad (10)$$

It shows that the following problems should be solved in SEN-LSSVM: (i) Select J_{SEN} ensemble sub-models from J_{En} candidate ones; (ii) Calculate weighting coefficients $\{\alpha_j\}_{j=1}^{J_{\mathrm{Sen}}}$; (iii) Select kernel parameter K_{LSSVM}; (IV) Select regularization parameter C_{LSSVM}; and (V) Maintain diversities among different ensemble sub-models.
Therefore, except the former two problems, the challenge is how to make the later three questions to be solved jointly. It is focus of this study.

3. PROPOSED ADAPTIVE SEN-LSSVM METHOD

Based on the above analysis, a new adaptive SEN-LSSVM modelling strategy is proposed based on ensemble construction by using candidate learning parameters. We denote $\{(\boldsymbol{x}_l,y_l)\}_{l=1}^{k}$ as the modelling data, $K_{\mathrm{LSSVM}}^{j_{\mathrm{ker}}}$ as the j_{ker}th candidate kernel parameter, $C_{\mathrm{LSSVM}}^{j_{\mathrm{reg}}}$ as the j_{reg}th candidate regularization parameter, and $\{\hat{y}_l\}_{l=1}^{k}$ as prediction outputs of the adaptive SEN-LSSVM model. The same candidate regularization parameters are used in term of maintain simplify and diversities among different sub-sub-models. Thus, all the candidate kernel and regularization parameters can be represented as $\{K_{\mathrm{LSSVM}}^{j_{\mathrm{ker}}}\}_{j_{\mathrm{ker}}=1}^{J_{\mathrm{ker}}}$ and $\{C_{\mathrm{LSSVM}}^{j_{\mathrm{reg}}}\}_{j_{\mathrm{reg}}=1}^{J_{\mathrm{reg}}}$ with number J_{ker} and J_{reg}.

Taken the j_{ker}th kernel parameter as example, the candidate sub-sub-models with different regularization parameters are constructed. This process is shown as,

$$\left.\begin{array}{l}\{(\boldsymbol{x}_l,y_l\}_{l=1}^{k}\\ \{C_{\mathrm{LSSVM}}^{j_{\mathrm{reg}}}\}_{j_{\mathrm{reg}}=1}^{J_{\mathrm{reg}}}\\ K_{\mathrm{LSSVM}}^{j_{\mathrm{ker}}}\end{array}\right\}\xrightarrow{\mathrm{LSSVM}}\{f_{j_{\mathrm{ker}}j_{\mathrm{reg}}}^{\mathrm{Sub\text{-}sub}}(\cdot)\}_{j_{\mathrm{reg}}=1}^{J_{\mathrm{reg}}}. \quad (11)$$

The prediciotn ouput of the $j_{\mathrm{reg}}^{\mathrm{can}}$th sub-sub-model is denoted as $\hat{y}_{j_{\mathrm{ker}}j_{\mathrm{reg}}}^{\mathrm{Sub\text{-}sub}}$, i.e.,

$$\hat{y}_{j_{\mathrm{ker}}j_{\mathrm{reg}}}^{\mathrm{Sub\text{-}sub}}=f_{j_{\mathrm{ker}}j_{\mathrm{reg}}}^{\mathrm{Sub\text{-}sub}}\left((\boldsymbol{x}_l)_{l=1}^{k},K_{\mathrm{LSSVM}}^{j_{\mathrm{ker}}},C_{\mathrm{LSSVM}}^{j_{\mathrm{reg}}}\right) \quad (12)$$

Then, BBSEN approach is used to select the optimum ensenble sub-sub-models and calculate their weighting coefficients. This process is shown as:

$$\left.\begin{array}{l}\{\hat{y}_{j_{\mathrm{ker}}j_{\mathrm{reg}}}^{\mathrm{Sub\text{-}sub}}\}_{j_{\mathrm{reg}}=1}^{J_{\mathrm{reg}}}\\ \\ \{\boldsymbol{y}_l\}_{l=1}^{k}\end{array}\right\}\xrightarrow{\mathrm{BBSEN}}\left\{\begin{array}{l}\left\{f_{j_{\mathrm{ker}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}(\cdot)\right\}_{j_{\mathrm{reg}}^{\mathrm{sel}}=1}^{J_{\mathrm{reg}}^{\mathrm{sel}}}\\ \{w_{j_{\mathrm{ker}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}\}_{j_{\mathrm{reg}}^{\mathrm{sel}}=1}^{J_{\mathrm{reg}}^{\mathrm{sel}}}\end{array}\right. \quad (13)$$

Based on Eq. (13), the prediction output of the obtained SEN-sub-model $f_{j_{\mathrm{ker}}J_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}(\cdot)$ can be represented as:

$$\hat{y}_{j_{\mathrm{ker}}}^{\mathrm{SEN\text{-}sub}}=f_{j_{\mathrm{ker}}J_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}(\cdot)=\sum_{j_{\mathrm{reg}}^{\mathrm{sel}}=1}^{J_{\mathrm{reg}}^{\mathrm{sel}}}w_{j_{\mathrm{ger}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}\cdot\hat{y}_{j_{\mathrm{ker}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}. \quad (14)$$

By repeating the above procedure J_{ker} times, all the SEN-sub-models with different kernel parameters are obtained. Their prediction outputs can be denoted as $\{\hat{y}_{j_{\mathrm{ker}}}^{\mathrm{SEN\text{-}sub}}\}_{j_{\mathrm{ker}}=1}^{J_{\mathrm{ker}}}$. Then, we use BBSEN at the second time to obtain the final adaptive SEN-LSSVM model $f^{\mathrm{SEN}}(\cdot)$. This process is shown as,

$$\left.\begin{array}{l}\{\hat{y}_{j_{\mathrm{ker}}}^{\mathrm{SEN\text{-}sub}}\}_{j_{\mathrm{ker}}=1}^{J_{\mathrm{ker}}}\\ \\ \{\boldsymbol{y}_l\}_{l=1}^{k}\end{array}\right\}\xrightarrow{\mathrm{BBSEN}}\left\{\begin{array}{l}\left\{f_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}(\cdot)\right\}_{j_{\mathrm{ker}}^{\mathrm{sel}}=1}^{J_{\mathrm{ker}}^{\mathrm{sel}}}\\ \{w_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\}_{j_{\mathrm{ker}}^{\mathrm{sel}}=1}^{J_{\mathrm{ker}}^{\mathrm{sel}}}\end{array}\right. \quad (15)$$

Based on Eq. (11)-(15), the prediction output of the final soft measuring model $f^{\mathrm{SEN}}(\cdot)$ can be represented as:

$$\begin{aligned}\hat{y}&=f^{\mathrm{SEN}}(\cdot)\\&=\sum_{j_{\mathrm{ker}}^{\mathrm{sel}}=1}^{J_{\mathrm{ker}}^{\mathrm{sel}}}w_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\cdot\hat{y}_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\\&=\sum_{j_{\mathrm{ker}}^{\mathrm{sel}}=1}^{J_{\mathrm{ker}}^{\mathrm{sel}}}w_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\cdot f_{j_{\mathrm{ker}}^{\mathrm{sel}}J_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\\&=\sum_{j_{\mathrm{ker}}^{\mathrm{sel}}=1}^{J_{\mathrm{ker}}^{\mathrm{sel}}}w_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\cdot\left(\sum_{j_{\mathrm{reg}}^{\mathrm{sel}}=1}^{J_{\mathrm{reg}}^{\mathrm{sel}}}w_{j_{\mathrm{ger}}^{\mathrm{sel}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}\cdot\hat{y}_{j_{\mathrm{ker}}^{\mathrm{sel}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}\right)\\&=\sum_{j_{\mathrm{ker}}^{\mathrm{sel}}=1}^{J_{\mathrm{ker}}^{\mathrm{sel}}}w_{j_{\mathrm{ker}}^{\mathrm{sel}}}^{\mathrm{SEN\text{-}sub}}\cdot\left(\sum_{j_{\mathrm{reg}}^{\mathrm{sel}}=1}^{J_{\mathrm{reg}}^{\mathrm{sel}}}w_{j_{\mathrm{ger}}^{\mathrm{sel}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}\cdot f_{j_{\mathrm{ker}}^{\mathrm{sel}}j_{\mathrm{reg}}^{\mathrm{sel}}}^{\mathrm{Sub\text{-}sub}}\left((\boldsymbol{x}_l)_{l=1}^{k},K_{\mathrm{LSSVM}}^{j_{\mathrm{ker}}},C_{\mathrm{LSSVM}}^{j_{\mathrm{reg}}}\right)\right)\end{aligned}$$

. (14)

It shows that the number of Sub-sub-model and SEN-sub-model are $J_{\mathrm{ker}}\times J_{\mathrm{reg}}$ and J_{ker}, respectively. The BBSEN method is used $(J_{\mathrm{ker}}+1)$ times to build the SEN-sub-models and the final adaptive SEN-LSSVM model. Obviously, the SEN-sub-models and the final SEN-LSSVM model can select the regularization and kernel parameters adaptively.

4. EXPERIMENTAL STUDY

Three datasets are used to validate the proposed methods. Each datasets are divided into two parts, which are training and testing ones, respectively. The popular radius basis function (RBF) is used in this paper. Two groups of candidate kernel and regularization parameters, i.e., Can_1={[0.01,0.03,0.05,0.07, 0.09,0.1,0.3,0.5, 0.7,0.9,1,3,5, 7,9,10,30, 50,70,100, 300,500,700,900,1000], [1,10,100,200, 400, 800, 1600, 3200, 6400, 12800]} and Can_2= {[1,10,100,500, 1000,2000, 4000, 6000, 8000, 10000, 20000, 40000, 60000,8000],[400,800, 1600,3200,6400,12800, 25600,51200, 102400]} are used in this study.

In this study, the proposed adaptive SEN-LSSVM method is mainly compared with the linear/non-linear latent structure modelling methods (partial least squares (PLS), kernel PLS (KPLS)) and one ensemble modelling approach (AGA-DNNE

[20]) methods. The low UCI and high dimensional mechanical frequency spectral datasets are used in this study.

4.1 Simulation Results by Using UCI Data

Two low-dimensional UCI benchmark datasets are used to validate the proposed method, which are Boston housing data and Concrete compressive strength data.

Boston housing data: The inputs include (1) Per capita crime rate by town (CRIM); (2) Proportion of residential land zoned for lots over 25,000 sq. ft. (ZN); (3) Proportion of non-retail business acres per town (INDUS) ; (4) Charles River dummy variable (CHAS) ; (5) Nitric oxide concentrations (NOX) ; (6) Average number of rooms per dwelling (RM) ; (7) Proportion of owner-occupied units built prior to 1940 (AGE); (8) Weighted distances to five Boston employment centers (DIS); (9) Index of accessibility to radial highways (RAD); (10) Full-value property-tax rate per $10,000 (TAX); (11) Pupil-teacher ratio by town (B); (12)lower status of the population (LSTAT);(13) Median value of owner-occupied homes in $1000s (MEDV). The output is housing values in the suburbs of Boston. And the data size is 506.

Concrete compressive strength data: The inputs include (1) Cement; (2) Blast furnace slag; (3) Fly ash; (4) Water; (5) Superplasticizer; (6) Coarse aggregate; (7) Fine aggregate in concrete per cubic meters of the various ingredients of concrete placement; (8) Conserved days. The output is concrete compressive strength. And the data size is 1030.

To Boston housing data with two groups candidate kernel and regularization parameters, the SEN-sub-models' prediction errors (RMSEs,) are shown in Fig. 1.

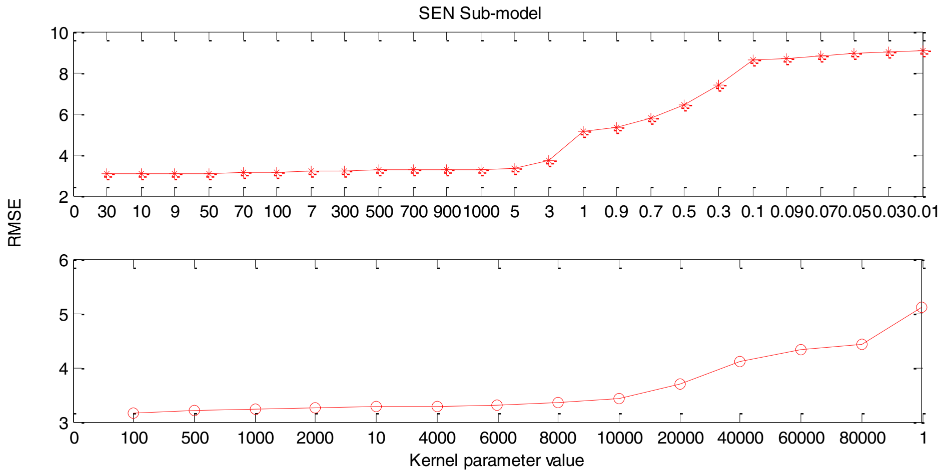

Fig. 1. SEN-sub-models' prediction errors of Boston housing data

It shows the two best SEN-sub-models select kernel parameter 30 and 1 from different candidate ones. To the SEN-sub-model with kernel parameter 30, it selects regularization parameters 10, 800, 400, 200 and 100 to construct ensemble sub-sub-models with RMSE 3.051. To SEN-sub-model with kernel parameter 1, regularization parameters 102400, 3200, 1600, 800 and 400 are used to build sub-sub-models with RMSE 3.162. To the Concrete compressive strength data, similar results can be obtained as that in Fig. 1. They are omitted in here.

Table 1. Statistical results of based on different data

		Boston housing data		Concrete compressive strength data	
		Can-1	Can-2	Can-1	Can-2
RMSE	Sen SENsub	3.027	3.158	7.220	7.163
	SENsub	3.051	3.162	7.291	7.436
	EnSENsub	4.160	3.284	8.968	8.270
SEN-sub number of SenSENsub		14	9	11	3

With these candidate SEN-sub-models, the final adaptive SEN-LSSVM model is constructed. The statistical results of the final SEN model (SenSENsub), the best SEN-sub-model (SENsub) and ensemble all SEN-sub-model (EnSENsub) are show in Table 1. The prediction curves are shown in Fig. 2-3.

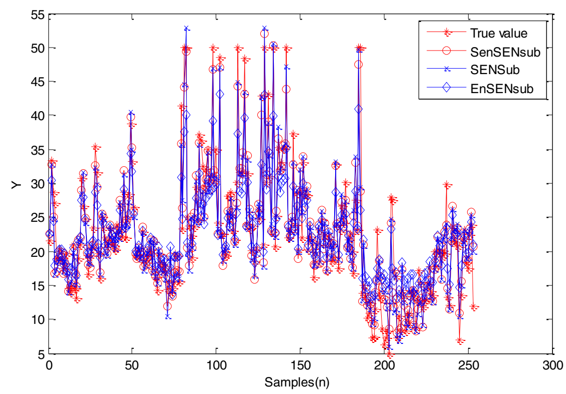

Fig. 2. Prediction curves of Boston housing data

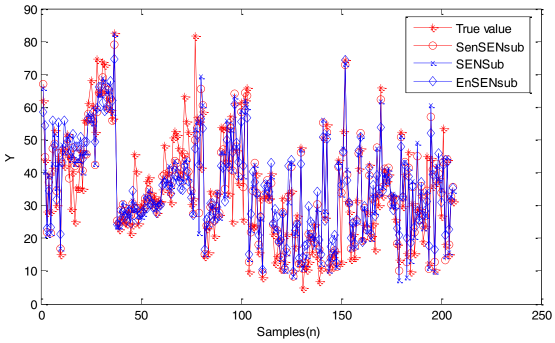

Fig. 3. Prediction curves of concrete compressive strength data

The above results show that the proposed method can model two UCI benchmark datasets effectively. The proposed method has the best prediction performance among three approaches. Its structure is adaptive selected.

4.2 Simulation Results by Using Frequency Spectral Data

Mechanical frequency spectral data from a laboratory scale ball mill is used to validate the proposed method. The data size is 26×15000, half of which is for training and half is for testing. Same as that in [20], charge volume ratio (CVR) is the only process parameter modelled in this study.

Considering the high dimensional of the frequency spectral data, PCA is the used to make feature extraction. The relation between the former 10 principal components (PCs) and the cumulate contribution ratio is shown in Fig. 4.

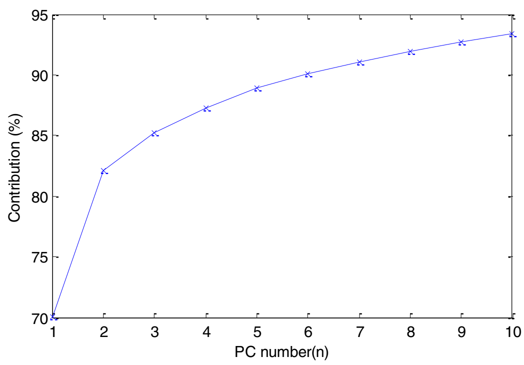

Fig. 4. Relation between the former 10 PCs and cumulate contribution ratio contribution

Fig. 4 shows that the first PC captures 70% information of the 15000 input frequency features. It also shows to select suitable PC number is important for construct effective soft sensor model. With the candidate learning parameters in "Can_2", the adaptive SEN-LSSVM models are constructed with PC number from 1 to 10. The best testing prediction results among SenSENsub, SENsub and EnSENsub models are shown in Fig. 5.

Fig. 5 shows that the suitable value is 8 for the SEN-LSSVM model based on these frequency spectral data. By using PC number as 8, the prediction curves of different LSSVM-based soft measuring models are shown in Fig. 6.

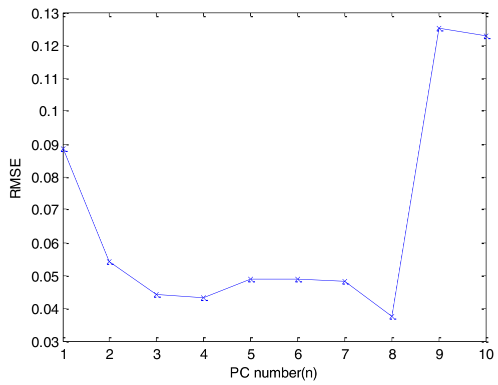

Fig. 5. Prediction errors based on different PCs number

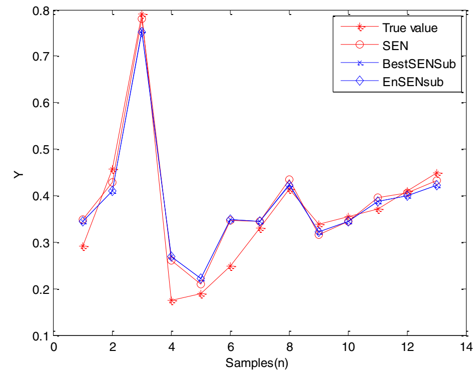

Fig. 6. Prediction curves of frequency spectral data

5.3 Comparative Results

The proposed method is compared with the PLS, KPLS, RVFL and AGA-DNNE approaches. Popular radius kernel function is used in KPLS. The number of latent variables (LVs) and kernel LVs (KLVs) are decided by the leave-one-out cross-validation. The number of hidden nodes of the RVFL is set as two times of the original inputs' number. The input features and learning parameters of DNNE are selected by using AGA. For RVFL and AGA-DNNE, the modelling process is repeated 100 times to overcome their randomness.

Table 2. Statistical results of comparative methods

	Boston housing data	Concrete compressive strength data	Frequency spectral data
PLS	4.681	10.92	0.04902
KPLS	3.195	8.179	0.05312
RVFL	4.740±0.3223	10.66±0.7308	4.208±0.5078
AGA-DNNE	3.272±0.1399	8.303±0.3590	0.06176±0.01019
This paper	3.027	7.163	0.03590

Table 2 shows that the proposed method has the best prediction performance without disturbance. As the input weights and bias of RVFL and AGA-DNNE methods are random initialization, their prediction errors are disturbed in a certain range. PLS/KPLS methods use the extracted latent variables to construct linear or nonlinear model. More researches should be done in the further study.

5. CONCLUSIONS

Selective ensemble (SEN) modeling method based on least square support vector machine (LSSVM) by adaptive using multiple candidate learning parameters, i.e., adaptive SEN-LSSVM, is proposed in this study. The contributions include: (1) Learning parameters of LSSVM are adaptive selected by using SEN modeling strategy; (2) Multiple learning parameters-based adaptive layer SEN-LSSVM algorithm is proposed at the first time; (3) The structure of double layer SEN model are adaptive determined during the modeling process. Simulation results based on UCI benchmark and mechanical frequency spectral datasets validate the effectiveness of this method.

ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China (61573364, 61703089), State Key Laboratory of Synthetical Automation for Process Industries (PAL-N201605), State Key Laboratory of Process Automation in Mining & Metallurgy and Beijing Key Laboratory of Process Automation in Mining & Metallurgy (BGRIMM-KZSKL-2017-07)

REFERENCES

Kadlec, P., Gabrys, B., and Strandt, S. (2009). Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 33(4), 795-814.

Wang, W., Chai, T.Y., and Yu, W. (2012). Modeling component concentrations of sodium aluminate solution via hammerstein recurrent neural networks. *IEEE Transactions on Control Systems Technology*, 20(4), 971-982.

Tang, J., Chai, T.Y., Yu, W., and Zhao, L.J. (2013). Modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information. *IEEE Transactions on Automation Science & Engineering*, 10(3), 726-740.

Chai, T.Y. (2009). Challenges of optimal control for plant-wide production processes in terms of control & optimization theories. *Acta Automatica Sinica*, 35(6), 641-649.

Kano, M., and Fujiwara, K. (2013). Virtual sensing technology in process industries : trends & challenges revealed by recent industrial applications. *Journal of Chemical Engineering of Japan*, 46(1), 1-17.

Tang, J., Chai, T.Y., Yu, W., and Zhao, L.J. (2012). Feature extraction and selection based on vibration spectrum with application to estimate the load parameters of ball mill in grinding process, *Control Engineering Practice*, 20(10), 991-1004

Nambiar, V.P., Khalil-Hani, M. Marsono, M. N., and Sia, C. W. (2014). Optimization of structure and system latency in evolvable block-based neural networks using genetic algorithm. *Neurocomputing*, 145(18), 285-302.

Chen, J., Yang, J., Zhao, J., Xu, F., Shen, Z., and Zhang, L. (2015). Energy demand forecasting of the greenhouses using nonlinear models based on model optimized prediction method. *Neurocomputing*, 174(PB), 1087-1100.

Guo, T., Han, L., He, L., and Yang, X. (2014). A ga-based feature selection and parameter optimization for linear support higher-order tensor machine. *Neurocomputing*, 144(1), 408-416.

Hu, J., Wang, Z., Liu, S., and Gao, H. (2016). A variance-constrained approach to recursive state estimation for time-varying complex networks with missing measurements. Automatica, 64, 155-162.

Hu, J., Wang, Z., Shen, B., and Gao, H. (2013). Quantised recursive filtering for a class of nonlinear systems with multiplicative noises and missing measurements. *International Journal of Control*, 86(86), 650-663.

Hansen, L.K. and Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 12(10), 993-1001.

Zhou, Z.H., Wu, J., and Tang, W. (2002.). Ensembling neural networks: many could be better than all, *Artificial Intelligence*, 137(1-2), 239-263,

Tang, J., Qiao, J.F., Wu, Z.W., and Chai, T.Y. (2018). Vibration and acoustic frequency spectra for industrial process modeling using selective fusion multi-condition samples and multi-source features. *Mechanical Systems & Signal Processing*, 99:142-168.

Tang, J., Chai, T.Y., Cong, Q.M., Zhao, L.J., Liu, Z., and Yu, W. (2015). Modeling Mill Load Parameters Based on Selective Fusion of Multi-scale Shell Vibration Frequency Spectrum. *Control Theory & Application*, 32(12): 1582-1591.

Tang, J., Chai,T.Y., Cong, Q.M., Yuan, B.C., Zhao, L. J., Liu, Z., and Yu, W. (2014) Soft sensor approach for modeling mill load parameters based on EMD and selective ensemble learning algorithm, *Acta Automatica Sinica*, 40(9), 1853-1866,.

Tang, J., Yu, W., Chai, T.Y., Liu, Z., and Zhou, X.J. (2016). Selective ensemble modeling load parameters of ball mill based on multi-scale frequency spectral features and sphere criterion, *Mechanical Systems & Signal Processing*, 66-67, 485-504.

Lv, Y., Liu, J., and Yang, T. (2013). A novel least squares support vector machine ensemble model for NOx, emission prediction of a coal-fired boiler. *Energy*, 55(1), 319-329.

Padilha, C.A.D.A., Barone, D.A.C., and Neto, A.D.D. (2016). A multi-level approach using genetic algorithms in an ensemble of Least Squares Support Vector Machines. *Knowledge-Based Systems*, 106(C), 85-95.

Tang, J., Qiao, J.F., Zhang, J., Wu, Z.W., Chai, T.Y., and Yu, W. (2018). Combinatorial optimization of input features and learning parameters for decorrelated neural network ensemble-based soft measuring model. *Neurocomputing,* 275: 1426-1440.