Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Adaptive Anti-Over-Fitting Soft Sensing Method Based on Local Learning

Weiming Shao, Xuemin Tian, Honglong Chen

College of Information and Control Engineering, China University of Petroleum, Qingdao, 266580, China (E-mail: swmupc@126.com; tianxm@upc.edu.cn; chenhl@upc.edu.cn)

Abstract: Local learning based soft sensing methods are effective in dealing with process nonlinearities as well as time varying characteristics. In this paper, an anti-over-fitting method is proposed for appropriate online local model adaptation. The proposed method is based on the weighted sum of the predicted errors for the newest few samples, the weights of which are determined adaptively. Moreover, to reduce the online computational load and memory cost, we propose two adaptive process states division schemes which consider the influence of both the variance and mean value of the predicted residual. Two case studies on continuous stirred tank reactor and debutanizer column demonstrate the effectiveness of the proposed soft sensing scheme.

Keywords: Soft sensing, local learning, model adaptation, over-fitting, process states division.

1. INTRODUCTION

In industrial plants, soft sensors are valuable alternatives for estimating important quality variables under some circumstances where hardware instruments fail to work well. During the past two decades, a variety of data-driven algorithms, such as Principal Component Analysis (PCA), Partial Least Squares (PLS), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have been applied for soft sensor modeling (Ge et al., 2011; Galicia et al., 2011; Fortuna et al., 2007; Yu, 2012). Kadlec et al. (2009) have provided a comprehensive review of this kind of soft sensors. However, soft sensors' estimation performance often deteriorates due to the time varying nature of process characteristics. Recursive methods have been developed for handling this problem by online updating the soft sensor model with newly measured samples (Tang et al., 2012; Qin et al., 1998; Liu et al., 2009). However, there are two main drawbacks of these methods: 1) they may adapt excessively within some narrow period and cannot successfully cope with abrupt changes of process characteristics (Fujiwara et al., 2009); 2) the single global model might not perform well in wide range due to the widely-existing nonlinearities in industrial plants.

Fortunately, local learning based soft sensors can simultaneously overcome these two drawbacks. In the framework of local learning, several sub models are constructed, each of which represents a single process state where the relationship between the primary variable and secondary variables is considered linear. In multi-model strategy, the final model output is the weighted sum of each sub model output. However, the model adaptation such as in the ensemble methods (Kadlec et al., 2011a) is difficult and complex. In this paper, we focus on the other type of local learning-based method, in which only one single local model is chosen to estimate the primary variable. As a result, appropriately selecting local models when model adaptations occur will be of great concern since inaccurate local models may extremely deteriorate soft sensors' performance. In just-in-time (JIT) modeling, the responsible local model is constructed upon samples around the query sample according to the criterion of Euclidean distances and angles between samples (Cheng et al., 2004; Chen et al., 2011; Liu et al., 2012a, Liu et al. 2012b). However, high prediction performance may not be obtained because of the ignorance of correlation between process variables. Therefore, Fujiwara et al. (2009) proposed a correlation-based JIT (CoJIT) method, where the correlation among input-output variables was taken into consideration. Nevertheless, the correlation is not as consistent as the prediction accuracy. Ni et al. (2012) proposed to select local model directly upon the predicted error for the newest sample. However, both CoJIT and the method of Ni et al. (2012) suffer from the issue of over-fitting because they greedily pursue the minimal value of the objective index for only one sample.

Another step in local learning is to define local data sets, upon which local models are constructed. Clustering-based ways such as fuzzy c-means (Fu et al., 2008) are commonly used ones. But it is difficult to determine the number of clusters appropriately and most of these methods are offline, resulting in the failure of extracting new process states. Liu et al (2009) proposed a moving window way, however, it was a global learning method. Fujiwara et al. (2009) proposed to divide process states by a moving window. Normally, high prediction accuracy requires small window moving width, leading to a large online computational load. So the authors used the angle θ between two subspaces to reduce the scale of the local model set. However, the locations of sub data sets are neglected which may lead to prediction bias. Ni et al. (2012) proposed to repeatedly partition the entire data set whenever an adaptation was triggered and thus may be quite time consuming. Kadlec et al. (2011a) presented an adaptive way of partitioning local regions based on t-test. Though it considered both the factors of the modeling function and

process characteristics, the variance in the denominator of T statistic could still negatively affect the performance.

To solve the above two problems, i.e., how to divide process states and how to adapt local models appropriately, this paper proposes an adaptive approach for local learning-based soft sensor modeling. To begin with, the effects of both the variance and mean value of the predicted errors are considered, after which two schemes are proposed for adaptive process states division. On the online operation stage, for addressing the over-fitting issue, we propose a novel criterion for model adaptation, which considers not only the newest sample but also the previous few ones. Note that the weight of each sample will be determined adaptively. The rest of this paper is organized as follows. Section 2 and section 3 provide detailed descriptions of the proposed ways for adaptive process states division and model adaptation, respectively. Two case studies are investigated and discussed in section 4 to demonstrate the effectiveness of the proposed scheme. Finally, the conclusion and future work are put forward in section 5.

2. ADAPTIVE PROCESS STATES DIVISION

In this section, we first analyze the drawback of the way of dividing process states by t-test and then two improved schemes are proposed.

2.1 Dividing Process States by T-test

According to the viewpoint of Kadlec et al. (2011a, 2011b), a local region should be the period where the modeling function remains constant performance and they presented a t-test-based approach to decide whether the local model's performance deteriorated by continuously shifting a window forward. Initially, a local model f_{ini} is built upon the data set $\boldsymbol{Z}_{ini}=[\boldsymbol{X}_{ini},\boldsymbol{Y}_{ini}]$ within the initial window. Then, the window is shifted one step ahead and the data set within the shifted window can be obtained as $\boldsymbol{Z}_{sft}=[\boldsymbol{X}_{sft},\boldsymbol{Y}_{sft}]$. The predicted residuals of $\boldsymbol{X}_{ini}$ and $\boldsymbol{X}_{sft}$ by f_{ini} can be calculated as (1)

$$\boldsymbol{R}_{ini}=f_{ini}(\boldsymbol{X}_{ini})-\boldsymbol{Y}_{ini}\ ,\ \boldsymbol{R}_{sft}=f_{ini}(\boldsymbol{X}_{sft})-\boldsymbol{Y}_{sft} \tag{1}$$

Assume both $\boldsymbol{R}_{ini}$ and $\boldsymbol{R}_{sft}$ follow normal distribution and thus t-test can be applied to judge if the mean value of $\boldsymbol{R}_{sft}$ significantly differs from that of $\boldsymbol{R}_{ini}$. As long as the null hypothesis is valid, $\boldsymbol{Z}_{sft}$ and $\boldsymbol{Z}_{ini}$ are considered to belong to the same process state and the window will continue to be shifted. However, when the performance of f_{ini} begins to deteriorate, the variance of $\boldsymbol{R}_{sft}$ in the denominator of T statistic rises, which slows the increase or sometimes causes a decrease of T statistic. Consequently, the null hypothesis may remain valid even if the process state has already changed.

2.2 Improved Process States Division Schemes

The above analysis shows that the influence of the variance is greater than that of the mean value. Thus we propose to define local regions by considering both the variance and mean value of the predicted residual. To start with, two statistics, namely T_{sft} and χ^2_{sft}, are constructed as (2).

$$T_{sft}=\sqrt{W}\,(\overline{R}_{sft}-\overline{R}_{ini})/S_{sft}\ ,\ \chi^2_{sft}=(W-1)S^2_{sft}/\sigma^2_{ini} \tag{2}$$

where W represents the initial window size; $\overline{R}_{ini}$ and σ_{ini} represent the mean value and standard deviation of the population where $\boldsymbol{R}_{ini}$ comes from; $\overline{R}_{sft}$ and S_{sft} represent the mean value and standard deviation of $\boldsymbol{R}_{sft}$, respectively. Generally, $\overline{R}_{ini}$ is equal to zero and σ_{ini} can be approximated by the sample standard deviation when W is large enough. With the assumption that $\boldsymbol{R}_{ini}$ and $\boldsymbol{R}_{sft}$ follow normal distribution, when the hypothesis $H_v: S_{sft}=\sigma_{ini}$ is valid, χ^2_{sft} follows χ^2 distribution with freedom degree of W-1. Similarly, when the hypothesis $H_m: \overline{R}_{sft}=0$ is valid, T_{sft} follows t distribution with freedom degree of W-1. In this paper, if and only if both of the above two hypotheses are valid which can be judged by (3), $\boldsymbol{Z}_{sft}$ and $\boldsymbol{Z}_{ini}$ are considered to belong to the same process state.

$$|T_{sft}|<\lambda_t\ \ \&\ \ \lambda_{\chi,1}<\chi^2_{sft}<\lambda_{\chi,2} \tag{3}$$

where λ_t and $\lambda_{\chi,1}$, $\lambda_{\chi,2}$ are the threshold values corresponding to the given significance levels α_t and α_χ with $P\{|T_{sft}|<\lambda_t\}$ = 1-α_t and $P\{\lambda_{\chi,1}<\chi^2_{sft}<\lambda_{\chi,2}\}=1-\alpha_\chi$.

On the other hand, with explicit consideration of the variance and mean value of $\boldsymbol{R}_{sft}$, we can directly formulate one index to examine if f_{ini} can describe $\boldsymbol{Z}_{sft}=[\boldsymbol{X}_{sft},\boldsymbol{Y}_{sft}]$ well or not. Set

$$J_f=(1-\gamma)(S_{sft}/\sigma_{ini}-1)^2+\gamma|\overline{R}_{sft}| \tag{4}$$

where γ weights the importance of the mean value. When J_f exceeds the given threshold value $\overline{J}_f$, the performance of f_{ini} is considered deteriorated. The above two schemes for judging if the performance of f_{ini} remains invariant are denoted as M_1 and M_2 for simplicity, respectively.

The procedure of adaptively dividing process states can be summarized as follows.

Step (1): Set initial data set $\boldsymbol{Z}_{ini}$, build the local model f_{ini} and calculate $\boldsymbol{R}_{ini}$ and σ_{ini};

Step (2): Shift the window one step ahead, get $\boldsymbol{Z}_{sft}$ and calculate $\boldsymbol{R}_{sft}$, $\overline{R}_{sft}$ and S_{sft};

Step (3): Calculate T_{sft} and χ^2_{sft} using (2) or J_f using (4), determine whether the performance of f_{ini} deteriorates according to either M_1 or M_2. If not, return to Step (2);

Step (4): A new local region is defined including samples from the first one of the initial window to the penultimate one of the shifted window, construct local model upon the newly extracted local region and reserve it. Then set $\boldsymbol{Z}_{ini} = \boldsymbol{Z}_{sft}$ and return to step (1).

When the above procedures are applied to divide process states adaptively, the number of local models is expected to be greatly reduced.

3. LOCAL MODEL ADAPTATION

The online model adaptation in local learning focused by this paper is of significant concern since the selected local models can extraordinarily influence soft sensors' estimation accuracy. Fujiwara et al. (2009) proposed to choose local model with the minimal weighted sum of the Q and *Hotelling* T^2 statistics computed by PCA. And Q, the distance from the sample to the subspace spanned by PCA model's loading vectors, is the dominant factor. However, Q is not in accordance with the prediction accuracy as shown in Fig. 1(a). For the current sample $(\boldsymbol{x}_0, y_0)$, Q_1 is less than Q_2 and thus model 1 will be selected when adaptation happens. But obviously, e_1, the actual predicted error of model 1 for $(\boldsymbol{x}_0, y_0)$, is much larger than that of model 2, e_2.

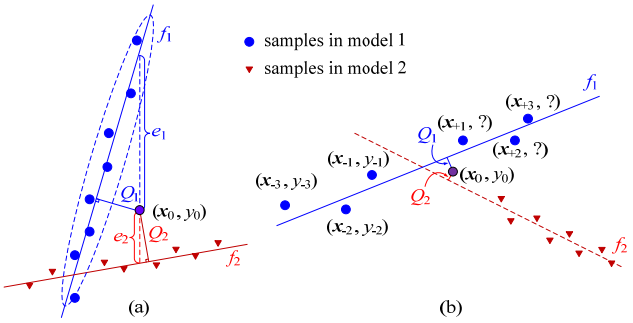

Fig. 1. (a) Selecting local model by Q; (b) Over-fitting issue

Besides, even if the model with the minimal predicted error is selected, over-fitting may be another more serious issue as shown in Fig. 1(b), where model 2 will be selected for the newly measured sample $(\boldsymbol{x}_0, y_0)$ after the adaptation. However, the estimation results of model 2 for the coming unknown samples, i.e. $(\boldsymbol{x}_{+i}, ?)$, are rather disappointing. The adaptation method in Ni et al.'s (2012) work suffered from the same issue because it also excessively pursued to minimize the predicted error for the newest labelled sample only.

To overcome this weakness, in this section, we propose an anti-over-fitting approach for model adaptation, utilizing not only the newest measured sample, but also the previous n ones. Moreover, in order to differentiate the importance of samples at different time instants, we adaptively assign a weight to each of the samples. In the proposed approach, the objective index for selecting local models is formulated as (5)

$$J_e = \sum_{k=0}^{n} a_{-k} e_{-k}^2 \tag{5}$$

where a_{-k} and e_{-k} represent the weight and predicted error of the previous k_{th} sample, respectively. And 0 is the subscript of the newest sample. Meanwhile, the weights are adaptively determined shown as (6)

$$a_{-k} = \exp(-\varphi d[(\boldsymbol{x}_{-k}, y_{-k}), (\boldsymbol{x}_0, y_0)]) \tag{6}$$

where $d[\cdot,\cdot]$ means the Euclidean distance and φ represents the scale parameter whose value is up to the specific data set. The local model that minimizes J_e is chosen to describe the current process dynamics. Note that n is one key parameter in the proposed strategy. The larger n is, the lower the over-fitting probability will be, resulting in the weaker ability of handling process nonlinearity and tracking current process dynamics and vice versa. Fortunately, from our tests and experiences, 2 will be large enough for n to avoid over-fitting and consequently, the default value of n can be set as 2.

In summary, the proposed adaptive local learning-based soft senor modeling method consists of two stages: the offline stage and the online stage.

Offline stage:

Divide process states by one of the adaptive schemes proposed in section 2.2 and reserve the corresponding models.

It is worth to point out that new process states are continuously to be extracted on the online operation stage, which is not available in the work of Kadlec et al. (2011a).

Online stage:

Step (1): Use the current local model f^* to estimate the unknown samples;

Step (2): When a new sample $(\boldsymbol{x}_0, y_0)$ is measured, calculate J_e using (5) with the newest (n+1) samples and f^*. If $J_e < \bar{J}_e$, return to step (1). Here $\bar{J}_e$ is set to regulate the adaptation frequency;

Step (3): Calculate $J_{e,l}$ with the l_{th} local model and the newest (n+1) samples where $l = 1, \cdots, L$ and L is the number of reserved local models on the offline stage;

Step (4): Set $f^* = f_{l^*}$, where $l^* = \arg\min_l \{J_{e,l}\}, l = 1, \cdots, L$, then return to step (1).

4. CASE STUDIES

In this section, two chemical processes are investigated to test the performance of the proposed adaptive soft sensor. For comparison purpose, the predicted results of nonlinear Least Squares SVM (LSSVM) without online adaptation, just-in-time PLS (JIT-PLS), Recursive PLS (RPLS) and CoJIT are also provided. Both CoJIT and the proposed method employ PLS to construct local models. The root mean squares error (RMSE), relative RMSE (RE) and max absolute error (MAE) (Liu et al., 2009) are adopted to measure the estimation accuracy, and the average consumed CPU time (second, denoted as CPT), stored model number (MN) and memory cost (byte, denoted as MC) are employed to measure soft sensors' performance from other perspectives. The computer

configuration is as follows: OS: Win XP; RAM: 2GB; CPU: Pentium Dual E5800 (3.2GHz×2); MATLAB version: 7.1.

4.1 Case Study 1: Continuous Stirred Tank Reactor (CSTR)

CSTR is usually the core equipment in chemical industry. The schematic diagram of an isothermal CSTR is shown in Fig. 2. Irreversible reaction A→B happens and the residual concentration of A, denoted as C_A, is the target variable for soft sensing. Q, Q_c, Q_F, T, h defined in Fig. 2 are selected as the secondary variables with sampling interval of 0.1 min. Detailed descriptions and mathematical model of the CSTR can be found in Fujiwara et al. (2009). Assume the catalyst activity decreases with time linearly and is recovered every 3 months. Meanwhile, we change the setpoint of the reactor temperature T every 10 days between ±2K. Assume C_A is analyzed in laboratory every 12 hours. To capture process dynamics, the input vector is augmented with the sample at the previous sampling instant. The total simulation time is 1 year, among which the data of first 9 months are treated as historical data set and the rest are used as test data set.

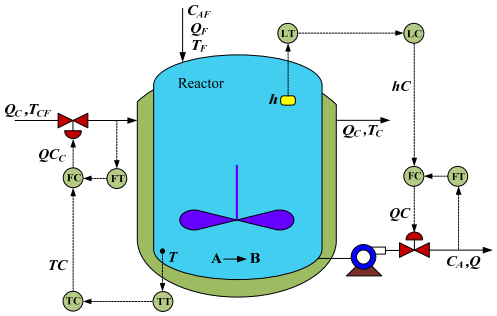

Fig. 2. Schematic diagram and control structure of CSTR

In this section, the process states division method in CoJIT is adopted. The parameters of the proposed method are determined by trial and error: W=14, n=2, PLS' number of latent variables LV_{pls}=8 and φ is set as 2. Parameters of other methods are optimized by particle swarm optimization (PSO) and all variables have been appropriately scaled. Fig. 3 shows the time trend plots of the four adaptive soft sensors' prediction results.

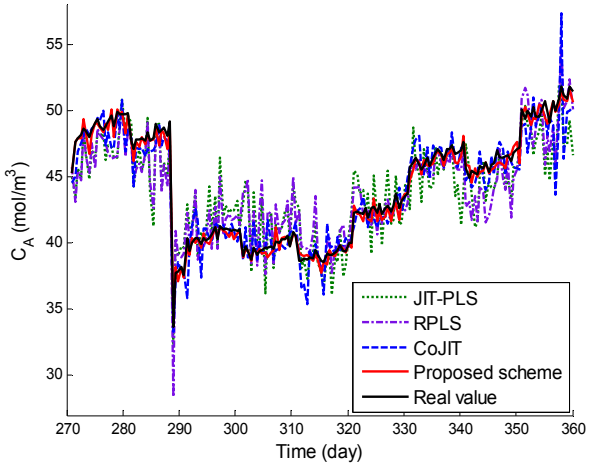

Fig. 3. Time trend plot of the results of adaptive soft sensors

By comparing the ability of tracking the target variable, it is obvious that the proposed method outperforms the other three ones. Furthermore, the simulation results of different soft sensors including LSSVM are listed in Table 1 for quantitative analysis of the estimation performance.

Table 1 Quantitative results of five soft sensors for C_A

	RMSE	RE	MAE	CPT	MN	MC
LSSVM	2.3567	5.23%	6.426	0	1	541
JIT-PLS	2.2834	5.14%	6.616	0.390	0	0
RPLS	2.0847	4.79%	5.462	0.094	1	11
CoJIT	1.3383	2.94%	6.351	1.988	702	58266
Proposed scheme	0.4841	1.08%	1.392	0.562	707	7777

The conclusion drawn from the accuracy indexes in Table 1 is consistent with that from Fig. 3. Moreover, the nonlinear approximator LSSVM could not model the CSTR well due to the strong process nonlinearity. While local learning-based methods have achieved better results. Particularly, the proposed method shows to have great advantage on restraining MAE owing to the anti-over-fitting strategy. On the other hand, if n is set as 0, i.e., only the newest sample is used for model adaptation, RMSE and MAE will rise to 1.297 and 9.845 respectively, indicating over-fitting occurs.

The last three columns show that the proposed method outperforms CoJIT in terms of computational efficiency and memory cost. Actually, in CoJIT, Q and T^2 statistics of all local models need to be calculated in each adaptation and thus the loading matrices and eigenvalues have to be stored. Assume each element of a matrix occupies one byte's space, then a total of $L\times(m+2)\times LV_{pca}$ bytes' extra memory space is indispensable apart from that for storing PLS models' parameters, where L, m and LV_{pca} represent the sample number, input vector's dimensionality and latent variables' number of PCA, respectively. While in the proposed scheme, only $L\times(m+1)$ bytes' memory space is necessary and merely simple vector multiplications are implemented.

If φ is set as 0, the RMSE of our method will rise to 0.8134. That is, all samples have the same weight in (5) is irrational. Furthermore, RMSE will be reduced to 0.5129 when φ is adjusted as 1, which greatly enhances the estimation accuracy. From our experiments, as long as the variables are scaled appropriately, φ is readily to be determined.

The results of our method are obtained with $\bar{J}_e$=0. Generally, smaller $\bar{J}_e$ is corresponding to higher adaptation frequency and better estimation performance. Thus, it is a trade-off between the adaptation frequency and prediction accuracy. For example, when $\bar{J}_e$ is set as 0.1 and 0.2, the adaptation times will be reduced to 133 and 97 from 177 with $\bar{J}_e$=0, and the RMSE will rise to 0.500 and 0.504, respectively.

4.2 Case Study 2: Debutanizer Column Process

In this section, the performance of the above five soft sensors will be further tested through a debutanizer column process.

After that, different strategies for adaptively dividing process states will be investigated in detail, including the two schemes proposed in section 2.2, i.e., M_1 and M_2.

The debutanizer distillation column is a part of the naphtha splitter and desulfuring plant, one of whose main tasks is to minimize butane content at the bottom. Fortuna et al. (2007) provided detailed descriptions of a debutanizer distillation column together with a benchmark data set. In this section, the data set is partitioned into two parts: the first 1650 samples serve as the historical data set and the following 650 ones are used as test data set. Butane content at the bottom is the target variable and 7 easy-to-measure variables are selected as the secondary variables, i.e., u_1~u_7, representing top temperature, top pressure, reflux flow, flow to the next process, 6th tray temperature and two bottom temperatures, respectively. The local model structure is determined using (7) by the analysis of expert knowledge and consideration of the process dynamics (Fortuna et al., 2007).

$$\hat{y}(k) = f[u_1(k), u_2(k), \cdots, u_5(k), u_5(k-1), u_5(k-2), u_5(k-3), \frac{u_6(k)+u_7(k)}{2}, y(k-4), y(k-5), y(k-6)] \quad (7)$$

where $\hat{y}$ represents the estimation value of the target variable. By trial and error, the parameters of the proposed method are set as follows: W=160, LV_{pls}=10, n=2 and $\varphi = 1$. Parameters of other soft sensors are determined by PSO. First, keep the way of dividing process states as the same as that in section 4.1. Table 2 summaries the quantitative performance of the five soft sensors and Fig. 4 shows the scatter plots of butane concentration predictions versus analyzed values.

Table 2 Estimation performance of different soft sensors for butane concentration

	RMSE	RE	MAE	CPT	MN
LSSVM	0.01975	86.3%	0.0922	0	1
JIT-PLS	0.01754	58.3%	0.1007	1.234	0
RPLS	0.01494	28.1%	0.0560	0.766	1
CoJIT	0.01463	28.1%	0.0644	24.278	2089
Proposed scheme	0.01102	12.6%	0.0447	4.723	2144

Table 2 clearly shows that the proposed method has achieved the best estimation performance and is much more computationally efficient than CoJIT. The reasons have already been analyzed before and will not be repeated here. Similarly, as shown in Fig. 4, the scattered points of our method lie closest to the black diagonal line in the whole operation range of the target variable, which further demonstrates the effectiveness of the proposed anti-over-fitting scheme on improving the prediction accuracy.

In this case study, the proposed method has built more than 2000 local models as shown in Table 2 where adjacent models may be redundant to each other because the window moving width d is fixed as 1. Therefore, the proposed two schemes of adaptively partitioning process states, namely M_1 and M_2, can be applied to reduce the model number and the computational load. Besides, for comparison purpose, some other data set partitioning approaches, such as t-test used in Kadlec et al. (2011a)' work (denoted as M_3) and the angle θ between two subspaces used in Fujiwara et al. (2009)' work (denoted as M_4), are also investigated. The initial window size is set as 160.

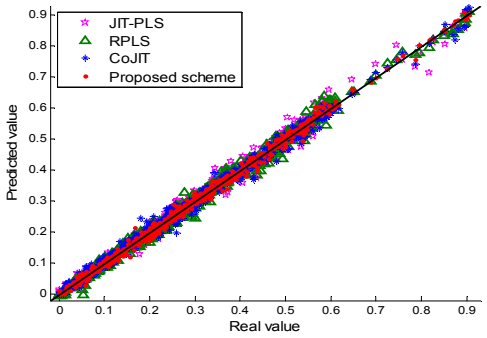

Fig. 4. Scatter plot comparison for butane concentration

In M_1, the confidence level of hypothesis H_v is set as equal as that of H_m. When M_1 and M_3 are applied, the RMSE, MN and CPT under different confidence levels are listed in Table 3. When M_2 with γ=0.002 and M_4 are applied, the results under different threshold values are listed in Table 4.

Table 3 RMSE, MN and CPT versus 1-α in M_1 and M_3

1-α	M_1			M_3		
	RMSE	MN	CPT	RMSE	MN	CPT
1%	0.01104	2128	4.55	0.01108	1964	4.40
5%	0.01110	1464	3.40	0.01117	1362	3.09
10%	0.01130	1045	2.41	0.01134	932	2.19
20%	0.01135	631	1.55	0.01187	494	1.16
30%	0.01141	447	1.09	0.01168	323	0.79
40%	0.01140	345	0.84	0.01174	220	0.54
50%	0.01130	259	0.65	0.01175	162	0.42
70%	0.01153	185	0.51	0.01191	96	0.28
90%	0.01154	120	0.36	0.01231	50	0.17
95%	0.01155	96	0.30	0.01276	37	0.14

Data in Table 3 and Table 4 indicate that as expected, the computational load decreases with the reduction of local model number. Also, the estimation performance deteriorates. However, in M_1 and M_2, the loss of prediction accuracy is quite slight compared with the decrement of model number and computational load. For example, in M_1, with α=0.6, the MN and CPT can be reduced by 84% and 82% respectively, while the RMSE raises only 3%. In addition, it is noted that the MN of M_1 is always more than that of M_3 under the same confidence level and meanwhile the prediction accuracy of M_1 remains higher than that of M_3 with almost the same model number, which implies the effectiveness of taking into consideration the variance in M_1.

Moreover, in order to reflect the advantage of our proposed schemes more clearly, the relationship between model number and prediction accuracy is plotted in Fig. 5, where the curves of M_1 and M_2 lie mainly below those of M_3 and M_4. In particular, RMSE of M_4 rises rapidly when the model

number is less than 1600, indicating the limitation of the angle between the subspaces. On the other hand, the curve of M_1 distinctly exhibits its great advantage on dealing with the trade-off between prediction accuracy and computational load. It should be pointed out that both M_1 and M_3 suffer from the difficulty of interpretability from the statistical perspective, since high estimation performance is with narrow confidence level. In such case, the significance level can be treated merely as a parameter for determining the threshold value.

Table 4 RMSE, MN and CPT versus $\bar{J}_f$ (10^{-3}) and θ (rad)

	M_2				M_4		
$\bar{J}_f$	RMSE	MN	CPT	θ	RMSE	MN	CPT
0.05	0.01106	1580	3.21	0.005	0.01106	1703	3.54
0.1	0.01119	1187	2.56	0.01	0.01133	1254	2.70
0.2	0.01145	735	1.58	0.02	0.01163	809	1.84
0.3	0.01149	516	1.11	0.03	0.01170	578	1.28
0.4	0.01166	373	0.86	0.04	0.01171	454	0.06
0.6	0.01164	266	0.61	0.05	0.011716	353	0.90
0.8	0.01172	190	0.48	0.07	0.01167	244	0.64
1.0	0.01150	154	0.40	0.09	0.01233	172	0.48
2.0	0.01206	84	0.25	0.15	0.01332	75	0.21

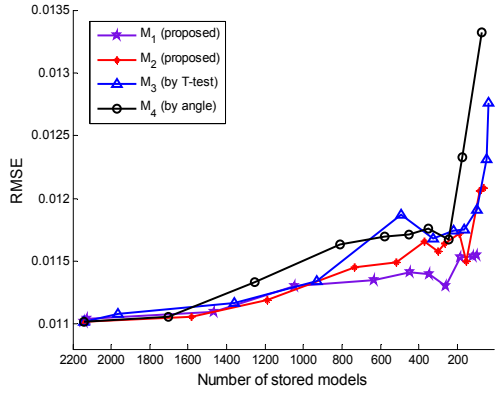

Fig. 5. RMSE versus model number in different methods

5. CONCLUSIONS AND FUTURE WORK

In this paper, approaches are proposed to avoid over-fitting and reduce the online computational load under the framework of local learning, respectively. Two case studies have illustrated the effectiveness of the proposed schemes. However, the harmfulness of outliers is still considerable. Even though numerous outlier detecting methods have been reported, how to distinguish samples of normal but new process states from the real outliers with a high accuracy is still a challenging issue, which will be our future work.

ACKNOWLEDGEMENT

This work is supported by the Natural Science Foundation of China (61273160) and the Fundamental Research Funds for the Central Universities (14CX06067A) and Open Project in Zhejiang Provincial Key Lab of Intelligent Processing Research of Visual Media (2012008).

REFERENCES

Cheng, C. and Chiu, M.S. (2004). A new data-based methodology for nonlinear process modeling. *Chemical Engineering Science*, **59(13)**, 2801-2810.

Chen, K., Ji, J., Wang, H.Q., Liu, Y. and Song, Z.H. (2011). Adaptive local kernel-based learning for soft sensor modeling of nonlinear processes. *Chemical Engineering Science and Design*, **89(10)**, 2117-2124.

Fortuna, L., Graziani S., Rizzo, A. and Xibilia, M.G. (2007). *Soft sensors for monitoring and control of industrial processes*. Springer-Verlag, London.

Fu, Y.F., Su H.Y. and Zhang Y. (2008). Adaptive soft sensor modeling algorithm based on FCMISVM and its application in PX adsorption separation process. *Chinese Journal of Chemical Engineering*, **16(5)**, 746-751.

Fujiwara, K., Kano, M., Hasebe, S. and Takinami, A. (2009). Soft sensor development using correlation-based just-in-time modeling. *AIChE*, **55(7)**, 1754-1764.

Galicia, H.J., He Q.P. and Wang J. (2011). A reduced order soft sensor approach and its application to continuous digester. *Journal of Process Control*, **21(4)**, 489-500.

Ge Z.Q. and Song Z.H. (2011). Semisupervised Bayesian method for soft sensor modeling with unlabeled data samples. *AIChE*, **57(8)**, 2109-2118.

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

Kadlec, P. and Gabrys, B. (2011a). Local learning based adaptive soft sensor for catalyst activation prediction. *AIChE*, **57(5)**, 1288-1301.

Kadlec, P. and Grbić, R. (2011b). Review of adaptation mechanisms for data-driven soft sensors. *Computers and Chemical Engineering*, **35(1)**: 1-24.

Liu, Y., Hu, N.P., Wang, H.Q and Li, P. (2009). Soft chemical analyzer development using adaptive least squares support vector regression with selective pruning and variable moving window size. *Ind. Eng. Chem. Res.*, **48(12)**, 5731-5741.

Liu, Y., Gao, Z.L., Li, P., Wang, H.Q (2012a). Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. *Ind. Eng. Chem. Res.*, **51(11)**, 4313-4327.

Liu, Y.Q., Huang, D.P. and Li, Y. (2012b). Development of interval soft sensors using enhanced just-in-time learning and inductive confidence predictor. *Ind. Eng. Chem. Res.*, **51(8)**, 3356-3367.

Ni W.D., Tan, S.K., Ng, W.J. and Brown, S.D. (2012). Localized, adaptive recursive partial least squares regression for dynamic system modeling. *Ind. Eng. Chem. Res.*, **51(8)**, 8025-8039.

Qin, S.J. (1998). Recursive PLS algorithm for adaptive data modeling. *Computers and Chemical Engineering*, **44(4-5)**, 503-514.

Tang, J., Yu, W. and Chai, T.Y. (2012). Online principal component analysis with application to process modeling. *Neurocomputing*, **87(4)**, 167-178.

Yu, J. (2012). A Bayesian inference based two-state support vector regression framework for soft sensor development in batch bioprocesses. *Computers and Chemical Engineering*, **41(1)**, 134-144.