Just-in-time Kernel Classifier for Online Process Diagnosis

Yi Liu

Institute of Process Equipment and Control Engineering,
Zhejiang University of Technology
Hangzhou, 310032, China
e-mail: yliuzju@zjut.edu.cn

Wenlu Chen

Jiangnan Institute of Computing Technology
Wuxi, 214083, China
e-mail: chenwenlu2000@yahoo.com.cn

Abstract—**A novel just-in-time kernel modeling method is proposed to online fault detection and diagnosis for chemical processes. The model parameters can be suitably selected using a fast cross-validation strategy. For a query sample, an online kernel classifier is constructed adaptively in a just-in-time manner for mode identification, i.e., fault detection and diagnosis, using the most relevant samples around it. The superiority of the proposed kernel classifier is demonstrated through a simulated chemical example, compared with the related method with fixed parameters.**

Keywords-fault detection and diagnosis; mode identification; kernel classifier; just-in-time learning

I. INTRODUCTION

Data-driven fault detection and diagnosis for chemical processes has attracted much attention because they are relatively more suitable for industrial applications. Specially, in the last two decades, various multivariate statistical process monitoring (MSPM) methods have been widely applied to chemical processes [1-6]. However, most of them use only the nominal operating data for modeling, and thus other diagnostic tools are applied for fault diagnosis. Furthermore, MSPM approaches have a strong statistical requirement and linearity assumption, greatly limiting their applications. Other types of data-driven process diagnosis approaches include, but are not limited to, neural networks, support vector machines, Gaussian process regression, pattern recognition, and fuzzy theory [7-10]. These methods can incorporate fault monitoring and diagnosis together because the fault information can generally be used in a supervised manner. However, they are typically problem specified and require that the designer be fairly familiar with the flow sheet of a process [11].

Recently, a new multiclass classification scheme for process fault diagnosis was developed, resulting in a unified solution to both process monitoring and diagnosis. That is, a single kernel learning (KL) algorithm can be applied simultaneously to both the modeling and diagnosis tasks [11]. The obtained results showed that it is a promising method for online fault detection and diagnosis. However, how to suitably select the model parameters, including the kernel parameter and the regularization parameter, is still unsolved. Additionally, only using a single global model, despite of its nonlinear modeling ability, is still difficult to describe the whole complex process. In this paper, a just-in-time kernel modeling method [12] is proposed to fault detection and diagnosis for chemical processes. The model parameters can

be suitably selected using a fast cross-validation (CV) strategy [13, 14]. Consequently, for a query sample, an online kernel classifier is constructed adaptively for fault diagnosis using the most relevant samples around it.

The remainder of this paper is so organized. The kernel classifier in a KL framework and its just-in-time strategy for online diagnosis are described in Section 2. The proposed method is illustrated by a simulated chemical reactor in Section 3. Finally, some conclusions are made in Section 4.

II. JUST-IN-TIME KERNEL CLASSIFIER

A. Kernel Classifier using a KL Framework

The process modeling issue based on the KL framework can be described as such a problem where the aim is to learn a mapping f: $\mathbf{X} \rightarrow \mathbf{Y}$ using a modeling set $\mathbf{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_k, y_k)\} \subset \mathbf{X} \times \mathbf{Y}$. As a temporal step, the KL method first maps the input data $\mathbf{x}_i \in \mathbf{X}$ implicitly into the feature space H by ϕ: $\mathbf{X} \rightarrow H$, where ϕ is a feature map associated with some positive definite kernel, i.e., the so called kernel trick: $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = <\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)>_H$. The function $f \in H$ is then determined in some optimal sense to yield the model [11-13]. A general form of the kernelized nonlinear model for process modeling can be formulated as [11]:

$$y_i = f(\mathbf{w}_k, b_k, \mathbf{x}_i) + e_i = \mathbf{w}_k^T \phi(\mathbf{x}_i) + b_k + e_i \qquad . (1)$$

where y_i and e_i denote the output measurement and the process noise at i instance, $i = 1, \cdots, k$; and $\mathbf{x}_i$ is a general input vector that is usually composed of several measured variables at time i, probably combined with their corresponding delayed forms and with the delayed outputs. The symbols $\mathbf{w}_k$ and b_k are the model parameter vector and the bias term from the k set data, respectively [12].

When applied the philosophy of statistical learning theory and the least squares support vector regression (LSSVR) framework to Eq. (1), the following optimization problem is formulated [13]:

$$\begin{cases} \min J(\mathbf{w}_k, b_k, \mathbf{e}_k) = \dfrac{1}{2}\|\mathbf{w}_k\|^2 + \dfrac{\gamma}{2}\|\mathbf{e}_k\|^2 \\ \text{s.t. } y_i - \mathbf{w}_k^T \phi(\mathbf{x}_i) - b_k - e_i = 0, \ i = 1, \cdots, k \end{cases} . (2)$$

where $\mathbf{e}_k = [e_1, e_2, \cdots, e_k]^T$ is the approximation error. The formulation consists of equality constraints instead of

inequality constraints in the conventional SVR and takes into account a squared error with the regularization term. Therefore, this reformulation greatly simplifies a problem such that the LSSVR solution follows directly from solving a set of linear equations rather than from a convex quadratic program and moreover, makes the algorithm more computationally efficient. The Lagrangian can be constructed to solve the optimization problem in Eq. (2) [13], then the solution can be expressed as

$$
\begin{cases}
\boldsymbol{\alpha}_k = \mathbf{P}_k \left[\mathbf{y}_k - \dfrac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{P}_k \mathbf{y}_k}{\mathbf{1}_k^T \mathbf{y}_k \mathbf{1}_k} \right] \\[2ex]
b_k = \dfrac{\mathbf{1}_k^T \mathbf{P}_k \mathbf{y}_k}{\mathbf{1}_k^T \mathbf{y}_k \mathbf{1}_k}
\end{cases}
\tag{3}
$$

where $\boldsymbol{\alpha}_k = \left[\alpha_{k,1}, \cdots, \alpha_{k,k} \right]^T$ are Lagrange multipliers; $\mathbf{y}_k = \left[y_1, \cdots, y_k \right]^T$; $\mathbf{I}_k \in \mathrm{R}^{k \times k}$ is a unit matrix and $\mathbf{1}_k$ is a vector of ones; $\mathbf{P}_k = \mathbf{H}_k^{-1}$ and $\mathbf{H}_k = \mathbf{K}_k + \mathbf{I}_k / \gamma$ with the kernel matrix $\mathbf{K}_k$, using the "kernel trick" $\mathbf{K}_k(i,j) = \left\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \right\rangle, \forall i, j = 1, \cdots, k$ [13]. Finally, the LSSVR model estimation of the system at time $k+1$, i.e., $\hat{y}_{k+1}$, can be obtained [13]:

$$
\begin{aligned}
\hat{y}_{k+1} &= f(\mathbf{w}_k, b_k, \mathbf{x}_{k+1}) \\
&= \sum_{i=1}^{k} \alpha_{k,i} \left\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_{k+1}) \right\rangle + b_k \\
&= \boldsymbol{\alpha}_k^{\mathrm{T}} \mathbf{k}_{k+1} + b_k
\end{aligned}
\tag{4}
$$

where $\mathbf{k}_{k+1}(i) = \left\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_{k+1}) \right\rangle, \forall i = 1, \cdots, k$ is a kernel vector to describe the similarity.

In fact, as for the classification problem, it can be easily and surprisingly shown that the common classification loss $\left[1 - y_i f(\mathbf{x}_i) \right]^2$ can be rewritten as [11]:

$$
\left[1 - y_i f(\mathbf{x}_i) \right]^2 = \left[y_i y_i - y_i f(\mathbf{x}_i) \right]^2 = \left[y_i - f(\mathbf{x}_i) \right]^2
\tag{5}
$$

where $y_i = \pm 1, i = 1, \cdots, k$. That is, the classification (i.e., pattern recognition) and regression (i.e., function approximation) problems can be described in a single unified KL algorithm [11].

Consequently, when the KL modeling framework is applied to process diagnosis, the corresponding decision function can be obtained in a straight way [11].

$$
\hat{y}_{k+1} = \mathrm{sgn}\left[f(\mathbf{w}_k, b_k, \mathbf{x}_{k+1}) \right] = \mathrm{sgn}\left[\boldsymbol{\alpha}_k^{\mathrm{T}} \mathbf{k}_{k+1} + b_k \right]
\tag{6}
$$

B. Just-in-time Kernel Classifier for Process Diagnosis

The JITL method, which is inspired by the ideas from local modeling and database technology, has been developed as an attractive alternative to nonlinear processes. As for process modeling and control, some JITL-based approaches have recently been proposed for nonlinear chemical processes [12, 16-18]. However, few nonlinear JITL methods are applied for process diagnosis. To account for this, in this work, a just-in-time kernel classifier (JKC) is proposed for process diagnosis.

Generally, for a query sample $\mathbf{x}_q$, there are three main steps to build a JKC model:

Step 1: Select the relevant samples to construct a similar set $\mathbf{S}_{\mathrm{sim}}$ in the database $\mathbf{S}$ based on some defined similarity criterions.

Step 2: Build a JKC model $f_{\mathrm{JKC}}(\mathbf{x}_q)$ using the relevant dataset $\mathbf{S}_{\mathrm{sim}}$.

Step 3: Predict its pattern of the output $\hat{y}_q$ online for the current query sample $\mathbf{x}_q$ and then discard the JKC model $f_{\mathrm{JKC}}(\mathbf{x}_q)$.

With the same three-step procedure, a new JKC model can be built for the next query sample. Without loss of generality, the traditional Euclidean distance-based similarity criterion [16] is utilized to search the similar samples $\mathbf{x}_i$ in the dataset, with the similarity s_{qi} defined below.

$$
s_{qi} = \exp\left(-\left\| \mathbf{x}_i - \mathbf{x}_q \right\| \right)
\tag{7}
$$

The $l_{\max}$ most similar samples can be selected due to the similarity criterion in Eq. (7). Then, rank the $l_{\max}$ similar samples according to the degree of similarity. A cumulative similarity factor (CSF) s_l is defined below [12]:

$$
s_l = \frac{\displaystyle\sum_{i=1}^{l} s_{qi}}{\displaystyle\sum_{i=1}^{l_{\max}} s_{qi}}, \quad l_{\min} \leq l \leq l_{\max}
\tag{8}
$$

which represents the cumulative similarity of l most similar samples compared to the relevant dataset $\mathbf{S}_{\mathrm{sim}}$. The CSF index can be utilized to access the cumulative similarity and then it can determine the l most similar samples in a more reasonable way. As an alternative method, the search range of $[l_{\min}, l_{\max}]$ can be substituted by the choice of s_l, e.g., $0.8 \leq s_l \leq 0.95$. Especially for multi-mode processes, the number of samples for each mode is different. Then, the relevant set of a query sample for different modes are generally different. It is difficult to determine the range of $[l_{\min}, l_{\max}]$ beforehand. Therefore, compared to the method of searching l in the range of $[l_{\min}, l_{\max}]$ directly, the CSF index is more meaningful and its computational burden can be reduced by narrowing the range of s_l suitably [12].

For the JKC model, the user-defined parameters include the regularization parameter γ, the kernel parameter (e.g., the Gaussian kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-\left\| \mathbf{x}_i - \mathbf{x}_j \right\| / \sigma \right]$ with the

width parameter $\sigma > 0$ or the polynomial kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^d$ with the degree parameter d a nature number), and the number of similar samples l. If the number of similar samples l is determined, the relevant dataset $\mathbf{S}_{sim}$ can be obtained. Then, with a pair of parameters $[\gamma, \sigma]$, the JLSSVR model can be built. Based on the fast leave-one-out (FLOO) CV criterion [14], the total error of the JKC model with l samples can be obtained [14]:

$$E_l^{FLOO} = \sum_{i=1}^{l} \left\| e_i^{FLOO} \right\| \\ = \sum_{i=1}^{l} \left\| \frac{\alpha_{l,i}}{P_{l,ii} + s_i^2 / o} \right\| \qquad . (9)$$

where $P_{l,ii}$ is the item at the ith row and ith column of $\mathbf{P}_l$, $\mathbf{s} = \mathbf{P}_l \mathbf{1}_l = [s_1, \cdots, s_l]^T$ and $o = -\mathbf{1}_l^T \mathbf{P}_l \mathbf{1}_l$. The related terms (i.e., $\mathbf{P}_l$ and $\alpha_{l,i}$) are available. Additionally, the computational load of $\mathbf{s}$ and o is small. The complexity of FLOO-CV can be reduced to about $O(l^3)$ operations, compared to the naive LOO-CV with about $O(l^4)$ operations [14]. Consequently, compared to the conventional LOO-CV method, the computation of E_l^{FLOO} is much more efficient and online selection of parameters of JKC is feasible.

III. SIMULATION RESULTS AND DISCUSSION

A nonlinear continuous stirred tank reactor (CSTR) process is investigated here because CSTR is commonly utilized in chemical processes. The CSTR process, in which an exothermic irreversible first-order reaction takes place, is shown in Fig. 1. The concentration C_a of the product left in the reactor is controlled by manipulating the coolant flow q_c through the jacket. Under the mass and energy balance, the dynamics of this CSTR process can be described below [19]: The nominal values of the variables and further information can be referred to Nahas et al [19].

The simulation environment is MatLab V2009b with CPU main frequency 2.3 GHz and 4 GB memory. The misclassification rate is considered to assess the diagnosis performance. In the past research papers, the polynomial kernel is utilized to model many processes with a certain degree in the range of $1 \le d \le 5$ [12, 13, 15]. Without loss of generality, the polynomial kernel is adopted for this case to illustrate the implementation procedure of JKC.

A sequence of 1000 samples is obtained from http://homes.esat.kuleuven.be/~smc/daisy/. With a sampling rate of 5 steps, only 200 samples are used to explore the classification ability of JKC under limited samples. Half of them are for training and the rest are for test, respectively. Here, the input vector is simply chosen as $\mathbf{x}_k = \begin{bmatrix} C_{a,k}, C_{a,k-1}, C_{a,k-2}, q_{c,k}, q_{c,k-1}, q_{c,k-2} \end{bmatrix}$. And the output variable is $C_{a,k+1}$. Additionally, the switch from one mode to another is carried out by changing the operating conditions of the reactor. Each mode has its special operating conditions

and exhibits different characteristics. Consequently, fault detection and diagnosis can be considered as the same problem of mode identification.

First, 4 modes are investigated under the noisefree environment. The obtained results of 100 test samples for their online fault detection and diagnosis using the JKC method can be shown in Fig. 2. All of them can be accurately classified into suitable modes, despite that the process modes have been changing frequently. As can be also shown in Fig. 2, each mode has uneven size. Consequently, the classification results indicate that the JKC approach can be simply applied to practical fault diagnosis problems with uneven sizes.
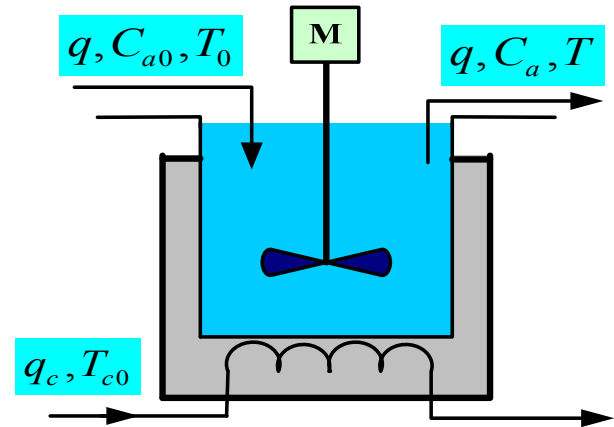

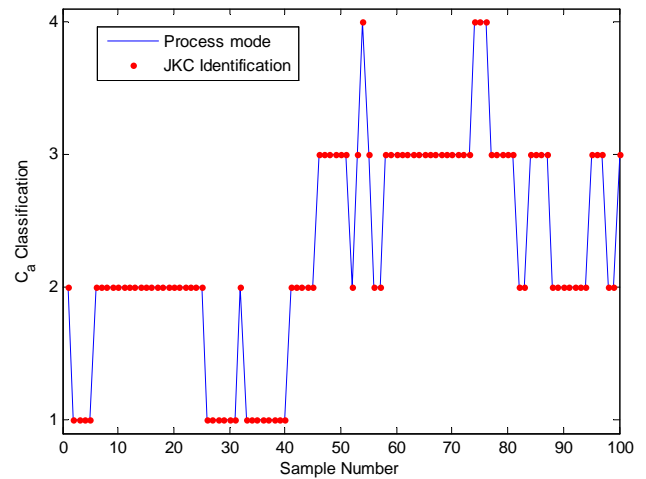Figure 1. A simplified schematic chart of the CSTR process.

Figure 2. Online fault detection and diagnosis (4 modes) using JKC of the CSTR process under the noisefree environment.

Furthermore, to simulate the industrial environment, both the input and output signals are corrupted by independent Gaussian noise with different variances. In such scenarios, some data are overlapping operation modes and exhibit more

complex characteristics. The obtained average results of 20 Monte Carlo simulations under different scenarios can be listed in Table 1. Two kernel classifiers with fixed parameters, including overfitting (larger $[\gamma, d]$) and underfitting (smaller $[\gamma, d]$), are listed for comparison study.

TABLE I. PERFORMANCE COMPARISON USING JKC WITH/WITHOUT PARAMETER SELECTION UNDER DIFFERENT SCENARIOS

Scenarios of Gaussian noise	Misclassification rate (%)		
	JKC with suitable parameters	*Parameters may overfitting*	*Parameters may underfitting*
Noisefree	**0**	1	4
2 %	**18.4**	22.7	19.9
5 %	**38.9**	45.4	40.8

As can be shown in Table 1, the JKC method can achieve better performance, i.e., with a smaller misclassification rate, than the related methods with fixed parameters, e.g., the proposed KL approach in [11]. This is mainly because the suitable parameters can be adjusted query-to-query using the FLOO-CV criterion. Actually, in many practical processes, the quality of modeling data may be poor because they often contain different noise and outliers. Consequently, it is important that the regression/classification model can adjust itself in a relatively suitable manner. Neither overfitting nor underfitting is a good choice if the parameters are user-defined beforehand. The overfitting of the model may lead to more inaccurate predictions when the data contain more uncertainty. Alternatively, due to the online optimization of model parameters, the JKC method can improve the diagnosis performance for these query samples. Therefore, from all the obtained results, the proposed JKC method shows its simplicity, efficiency, and superiority for online fault detection and diagnosis.

IV. CONCLUSION

This paper has addressed the subject of developing a simple and suitable nonlinear method for online process diagnosis. The proposed just-in-time kernel classifier (JKC) can be simply implemented for online fault detection and diagnosis of chemical processes. Using an efficient strategy for parameter selection, the JKC model can obtain reliable performance to meet the practical requirements. The superiority of the proposed method is demonstrated through a simulated chemical example. In our opinion, the JKC can be further applied to complex process due to its lazy learning manner, which will be one of our future directions.

ACKNOWLEDGMENT

REFERENCES

[1] L. H. Chiang, E. L. Russell, and R. D. Braatz. Fault Detection and Diagnosis in Industrial Systems. London, UK: Springer-Verlag; 2001.

[2] V. Venkatasubramanian, R. Rengaswamy, K. W. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Parts I–III", Computers and Chemical Engineering, vol. 27, 2003, pp. 293-346.

[3] C. Undey, and A. Cinar, "Statistical monitoring of multistage, multiphase batch processes," IEEE Control Systems Magazine, vol. 22, 2002, pp. 40-52.

[4] Y. Yao, and F. R. Gao, "A survey on multistage/multiphase statistical modeling methods for batch processes," Annual Review in Control, vol. 33, 2009, pp. 172-183.

[5] C. H. Zhao, F. L. Wang, N. Y. Lu, and M. X. Jia, "Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes," Journal of Process Control, vol. 17, 2007, pp. 728-741.

[6] J. H. Chen, C. M. Song, and T. Y. Hsu, "On-line monitoring of batch processes using IOHMM based MPLS," Industrial & Engineering Chemistry Research, vol. 49, 2010, pp. 2800-2811.

[7] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," Computers and Chemical Engineering 2004;28:1389-1401.

[8] S. Dash, R. Rengaswamy, and V. Venkatasubramanian, "Fuzzy-logic based trend classification for fault diagnosis of chemical processes," Computers and Chemical Engineering, vol. 27, 2003, pp. 347-362.

[9] T. Chen, and J. Zhang, "On-line multivariate statistical monitoring of batch processes using Gaussian mixture model," Computers and Chemical Engineering, vol. 34, 2010, pp. 500-507.

[10] T. Chen, and Y. Sun, "Probabilistic contribution analysis for statistical process monitoring: A missing variable approach," Control Engineering Practice, vol. 17, 2009, pp. 469-477.

[11] H. Q. Wang, P. Li, F. R. Gao, Z. H. Song, and S. X. Ding, "Kernel classifier with adaptive structure and fixed memory for process diagnosis," AIChE Journal, vol. 52, 2006, pp. 3515-3531.

[12] Y. Liu, Z. L. Gao, P. Li, H. Q. Wang, "Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes," Industrial & Engineering Chemistry Research, vol. 51, 2012, pp. 4313-4327.

[13] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, Least Squares Support Vector Machines, Singapore: World Scientific; 2002.

[14] G. C. Cawley, and N. L. C. Talbot, "Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters," Journal of Machine Learning Research, vol. 8, 2007, pp. 841-861.

[15] Y. Liu, H. Q. Wang, J. Yu, and P. Li, "Selective recursive kernel learning for online identification of nonlinear systems with NARX form," Journal of Process Control, vol. 20, 2010, pp. 181-194.

[16] G. Bontempi, M. Birattari, and H. Bersini, "Lazy learning for local modeling and control design," International Journal of Control, vol. 72, 1999, pp. 643-658.

[17] C. Cheng, and M. S. Chiu, "Nonlinear process monitoring using JITL-PCA," Chemometrics and Intelligent Laboratory Systems, vol. 76, 2005, pp. 1-13.

[18] K. Fujiwara, M. Kano, S. Hasebe, and A. Takinami, "Soft-sensor development using correlation-based just-in-time modeling," AIChE Journal, vol. 55, 2009, pp. 1754-1765.

[19] E. P. Nahas, M. A. Henson, and D. E. Seborg, "Nonlinear internal model control strategy for neural network models," Computers and Chemical Engineering, vol. 16, 1992, pp. 1039-1057.