

Research Article

Soft Sensor Development Based on Quality-Relevant Slow Feature Analysis and Bayesian Regression with Application to Propylene Polymerization

Miao Zhang , Le Zhou, Jing Jie , and Xiaoli Wu

School of Automation and Electrical Engineering, Zhejiang University of Science & Technology, Hangzhou 310023, China

Correspondence should be addressed to Miao Zhang; zhangmiao@zust.edu.cn

Received 22 March 2021; Revised 14 August 2021; Accepted 21 August 2021; Published 30 December 2021

Academic Editor: Chengkuo Lee

Copyright © 2021 Miao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data-driven soft sensors are widely used to predict quality indices in propylene polymerization processes to improve the availability of measurements and efficiency. To deal with the nonlinearity and dynamics in propylene polymerization processes, a novel soft sensor based on quality-relevant slow feature analysis and Bayesian regression is proposed in this paper. The proposed method can handle the dynamics of the process better by extracting quality-relevant slow features, which present both the slowly varying characteristic and the correlations with quality indices. Meanwhile, a Bayesian inference model is developed to predict the quality indices, which takes advantages of a probability framework with iterative maximum likelihood techniques for parameter estimation and a sparse constraint for avoiding overfitting. Finally, a case study is conducted with data sampled from a practical industrial propylene polymerization process to demonstrate the effectiveness and superiority of the proposed method.

1. Introduction

As modern process industries become larger scale and more integrated, pivotal key performance indices about product quality, process safety, and pollution reduction should be closely monitored [1–4]. Besides, real-time measuring of quality indices is important for process monitoring, control, and optimizing products [5–7]. Thermoplastic polymers such as polypropylene are important materials and have been used in various sectors. The melt index (MI) which determines the grade of the polymer product is considered one of the most crucial indicators in quality control for industrial propylene polymerization (PP) processes. Up to date, an accurate first-principle MI model is still not available. The MI is usually evaluated offline with an analytical procedure in the laboratory which takes 1.5–2 h to complete. This can result in a delay in the quality control system as there is no available quality indicator at this time. In the absence of an economical or effective online measurement, soft sensors could serve as an alternative solution [8–14]. Additionally, with the wide availability of process data in

PP processes, increasing data-driven soft sensors have been adopted to predict the MI.

Data-driven approaches are used as an alternative to the mathematical model. The data reading from sensors can often contain a large number of variables that describe the same process phenomena. When encountering high-dimensional data, it is useful to project the data into the latent space, which is more compact than the original space. For MI prediction, the principal component analysis (PCA) [15] and partial least squares (PLS) [16] have been applied. Unfortunately, as the model is linear, it is inadequate when the soft sensing variables are nonlinear. To this end, nonlinear modeling methods are used in MI prediction, such as artificial neural networks (ANNs) [17], support vector machines (SVMs) [18], Gaussian process regression (GPR) [19, 20], and relevance vector machine (RVM) [21] [22–25]. Recently, Liu et al. proposed an adversarial transfer learning- (ATL-) based soft sensor [26] and a domain adaptation transfer learning soft sensor for product quality prediction [7]. As classical methods, PCA and PLS have achieved great successes with respect to quality prediction

by extracting latent variables (LVs) [27–29]. However, they can only be used to extract static LVs, which are unsuitable for modeling dynamic processes because they are limited by containing temporally related process information.

In practical chemical processes, frequent fluctuations of equipment characteristics with time always result in dynamic processes. For soft sensing of quality indices, dynamic data characteristic could play an important role in regression, the result of which may highly influence the performance for the quality control system. A new merged unsupervised algorithm, slow feature analysis (SFA) [30], is applied in our work for its remarkable ability to extract slowly varying and temporally related features for modeling dynamical processes. SFA has been extensively used for various reasons, such as blind source separation [31], signal processing [32], and process monitoring [33–36]. Unlike the static models generated by PCA or PLS, SFA-based models can better describe process temporal behaviors through the extraction of slowness of LVs. When used for modeling, the extracted slowly varying LVs will be termed as process intrinsic properties, and the fast-varying LVs are seen as process noise inversely. Up to now, some SFA-based methods have been applied in prediction modeling. Shang et al. [37] proposed probabilistic slow feature analysis (PSFA) based a soft sensor model for quality prediction. Fan et al. [38] proposed a robust PSFA- (RPSFA-) based regression model that models outliers in the observation data using Student's t -distribution. Zhong et al. [39] put forward an online quality prediction method based on modified regularized SFA (ReSFA). Jiang et al. [40] proposed a real-time semisupervised predictive modeling strategy for industrial continuous catalytic reforming process using SFA. However, traditional slow features are calculated only considering slowness of process variations, while their correlations with quality indices are neglected for feature extraction. Slow features are extracted and then used for quality prediction by performing regression using the ordinary least square, which means that they may not describe nonlinear relationship among variables well. Considering the nonlinearity of the propylene polymerization process, using nonlinear regression modeling method is quite necessary.

Hence, to deal with the nonlinearity and dynamics in PP processes, a soft sensor based on quality-relevant slow feature analysis and Bayesian regression is proposed in this paper. First, the quality-relevant slow feature analysis (QSFA) is adopted to extract slow features which present both the slowly varying characteristic and the correlations with quality indices. Then, the extracted quality-relevant slow features (QSFs) are sorted according to the correlation index and selected for regression modeling by the cross-validation method. Based on the selected QSFs, a Bayesian inference framework named relevance vector regression (RVR) is developed to predict the quality variable, i.e., MI for the polypropylene products. It takes advantages of a probability framework with iterative maximum likelihood techniques for parameter estimation and a sparse constraint for avoiding overfitting. Finally, the effectiveness of the proposed method (QSFA-RVR) is confirmed through real data obtained from industrial propylene polymerization plants.

The rest of the paper is organized as follows. Section 2 presents the methods for the development of the proposed soft sensor. First, a brief introduction of the traditional SFA is presented. Then, the formulation for the QSFA-RVR prediction model is presented, which contains the QSFA method, the relevance vector regression, and the quality prediction method based on QSFA-RVR. Experiment results on real data from practical industry processes are analyzed and discussed in Section 3. Finally, the conclusion is drawn in Section 4.

2. Methods

2.1. Overview of Slow Feature Analysis. SFA is an unsupervised algorithm proposed by Wiskott and Sejnowski [30], which extracts invariant features from quickly varying signals. The slow variations are always termed as intrinsic features of signals. Meanwhile, rapid variations often allude to measurement noise. The detailed methodology of SFA is illustrated as follows [41].

Given a J -dimensional input signal \mathbf{X} , SFA is aimed at finding mapping functions from input to slow features $\mathbf{s}_j = g_j(\mathbf{X})$, $j \in [1, 2, \dots, J]$ that minimize

$$\min_{g_j(\cdot)} \Delta(\mathbf{s}_j) := \langle \dot{\mathbf{s}}_j^2 \rangle, \quad (1)$$

under the constraints

$$\begin{aligned} \langle \mathbf{s}_j \rangle &= 0 \text{ (zero mean),} \\ \dot{\mathbf{s}}_j \left\{ g_j(\cdot) \right\}_{j=1}^J & \langle \mathbf{s}_j^2 \rangle = 1 \text{ (unit variance),} \\ \forall i \neq j, \langle \mathbf{s}_i \mathbf{s}_j \rangle &= 0 \text{ (decorrelation and order),} \end{aligned} \quad (2)$$

where $\langle \cdot \rangle$ denotes time averaging and denotes the first-order derivative of \mathbf{s}_j with respect to time.

To simplify this optimization problem, it is assumed that the SFs are a linear combination of input variables. The linear mapping is

$$\mathbf{s}_j = \mathbf{X} \mathbf{w}_j. \quad (3)$$

The optimization problem in (1) can be solved by two steps of singular value decomposition (SVD) [42], which is tantamount to a numerical solution to the generalized eigenvalue decomposition (GED) problem in

$$\mathbf{A} \mathbf{W} = \mathbf{B} \mathbf{W} \mathbf{\Omega}, \quad (4)$$

where $\mathbf{A} = \langle \dot{\mathbf{X}} \dot{\mathbf{X}}^T \rangle$ denotes the covariance matrix of the first-order derivative of \mathbf{X} and $\mathbf{B} = \langle \mathbf{X} \mathbf{X}^T \rangle$ denotes the covariance matrix of \mathbf{X} ; $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j]^T$ is the matrix of generalized eigenvectors, which are coefficient vectors of linear mappings; and $\mathbf{\Omega} = \text{diag} \{ \lambda_1, \dots, \lambda_j \}$ is a diagonal matrix with λ_j

$= \langle \dot{\mathbf{s}}_j^2 \rangle$ being generalized eigenvalues, which are arranged in an ascending order.

By performing a first SVD on \mathbf{B} with equation (5), the original input \mathbf{X} is whitened as Eq. (6).

$$\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (5)$$

$$\mathbf{Z} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T\mathbf{X}, \quad (6)$$

where \mathbf{U} is an orthogonal matrix, $\mathbf{\Lambda}$ is a diagonal matrix, and $\text{cov}(\mathbf{Z}) = \langle \mathbf{Z}\mathbf{Z}^T \rangle = \mathbf{I}_J$. Then, a second SVD on the covariance of the first-order derivative of \mathbf{Z} yields

$$\langle \dot{\mathbf{Z}}\dot{\mathbf{Z}}^T \rangle = \mathbf{P}^T\mathbf{\Omega}\mathbf{P}, \quad (7)$$

where \mathbf{P} is an orthogonal matrix.

Thus, the weighting matrix \mathbf{W} can be calculated, and the slow features \mathbf{s} can be obtained using equation (3).

2.2. Quality-Relevant Slow Feature Analysis. Conventional SFA only considers slowness of process variations when extracting slow features. SFA's competence to retain process information containing quality indices is limited. An intuition is presented to integrate the relationships between slow features and quality indices into SFA's optimization problem. In this way, an improved SFA algorithm, named quality-relevant slow feature analysis (QSFA), can extract quality-relevant slow features (QSF) from a process, and these slow features can better reveal the essence of the process. Moreover, a new objective function for SFA should be designed to enable the QSFA that considers both slowness of process variations and their correlations with quality indices. Given the normalized process variables \mathbf{X} and the quality index \mathbf{y} , a new objective function $J(\mathbf{w})$ composed of three subobjective functions is designed. The first subobjective function J_1 which considers that the QSF should change slowly is given as follows:

$$\min J_1(\mathbf{w}) = \Delta(\dot{\mathbf{s}}) = (\dot{\mathbf{X}}\mathbf{w})^T(\dot{\mathbf{X}}\mathbf{w}), \quad (8)$$

where \mathbf{s} is the QSF and \mathbf{w} is the corresponding weighting vector. Considering that the QSF should be highly correlated with the quality index, the second subobjective function J_2 is aimed at maximizing the correlation between \mathbf{s} and \mathbf{y} , which is given as follows:

$$\max J_2(\mathbf{w}) = \frac{\mathbf{s}^T\mathbf{y}\mathbf{y}^T\mathbf{s}}{\sqrt{\text{var}(\mathbf{s})}} = \frac{(\mathbf{X}\mathbf{w})^T\mathbf{y}\mathbf{y}^T(\mathbf{X}\mathbf{w})}{\sqrt{(\mathbf{X}\mathbf{w})^T(\mathbf{X}\mathbf{w})}}. \quad (9)$$

The third subobjective function J_3 is used to maximize the variance of \mathbf{s} which represents the variation information of the process.

$$\max J_3(\mathbf{w}) = \sqrt{\text{var}(\mathbf{s})} = \sqrt{(\mathbf{X}\mathbf{w})^T(\mathbf{X}\mathbf{w})}. \quad (10)$$

Thus, a new optimization problem of QSFA can be described as

$$\begin{aligned} \max J(\mathbf{w}) &= \frac{J_2 J_3}{J_1} = (\mathbf{X}\mathbf{w})^T\mathbf{y}\mathbf{y}^T(\mathbf{X}\mathbf{w}), \\ \text{s.t. } &(\dot{\mathbf{X}}\mathbf{w})^T(\dot{\mathbf{X}}\mathbf{w}) = 1 \\ &\mathbf{w}^T\mathbf{w} = 1. \end{aligned} \quad (11)$$

By introducing Lagrange multipliers λ and δ , equation (14) can be easily obtained. Obviously, the value of the objective function $J(\mathbf{w})$ is equal to $(\lambda + \delta)$.

$$L(\mathbf{w}) = \mathbf{w}^T\mathbf{X}^T\mathbf{y}\mathbf{y}^T\mathbf{X}\mathbf{w} - \lambda(\mathbf{w}^T\dot{\mathbf{X}}^T\dot{\mathbf{X}}\mathbf{w} - 1) - \delta(\mathbf{w}^T\mathbf{w} - 1), \quad (12)$$

$$\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}} = 2\mathbf{X}^T\mathbf{y}\mathbf{y}^T\mathbf{X}\mathbf{w} - 2\lambda\dot{\mathbf{X}}^T\dot{\mathbf{X}}\mathbf{w} - 2\delta\mathbf{w} = 0, \quad (13)$$

$$\mathbf{w}^T\mathbf{X}^T\mathbf{y}\mathbf{y}^T\mathbf{X}\mathbf{w} = \lambda + \delta, \quad (14)$$

Therefore, we can rewrite the solution of equation (13) as the following GED problem [43]:

$$\mathbf{X}^T\mathbf{y}\mathbf{y}^T\mathbf{X}\mathbf{w} = (\lambda\dot{\mathbf{X}}^T\dot{\mathbf{X}} + \delta\mathbf{I})\mathbf{w}, \quad (15)$$

where \mathbf{I} is a J -dimensional identity matrix.

By performing SVD decomposition on the covariance matrix $\dot{\mathbf{X}}^T\dot{\mathbf{X}}$, we have

$$\dot{\mathbf{X}}^T\dot{\mathbf{X}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \quad (16)$$

The whitening transformation with the whitening matrix \mathbf{H} is described as follows:

$$\mathbf{Z} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T\dot{\mathbf{X}}^T = \mathbf{H}^T\dot{\mathbf{X}}^T. \quad (17)$$

Obviously, it holds that $\mathbf{Z}\mathbf{Z}^T = \mathbf{H}^T\dot{\mathbf{X}}^T\dot{\mathbf{X}}\mathbf{H} = \mathbf{I}$. Thus, finding a weighting vector \mathbf{w} is identical to finding a vector $\mathbf{H}\boldsymbol{\phi}$ that satisfies $\mathbf{s} = \mathbf{X}\mathbf{H}\boldsymbol{\phi}$. The GED problem in equation (15) can be reformulated as

$$\mathbf{X}^T\mathbf{y}\mathbf{y}^T\mathbf{X}\mathbf{H}\boldsymbol{\phi} = (\lambda\dot{\mathbf{X}}^T\dot{\mathbf{X}} + \delta\mathbf{I})\mathbf{H}\boldsymbol{\phi}. \quad (18)$$

Left multiplying both sides of equation (18) by \mathbf{H}^T and substituting $\mathbf{H}^T\dot{\mathbf{X}}^T\dot{\mathbf{X}}\mathbf{H} = \mathbf{I}$ and $\mathbf{H}^T\mathbf{H} = \mathbf{\Lambda}^{-1}$, we obtain

$$\mathbf{H}^T\mathbf{X}^T\mathbf{y}\mathbf{y}^T\mathbf{X}\mathbf{H}\boldsymbol{\phi} = (\lambda\mathbf{I} + \delta\mathbf{\Lambda}^{-1})\boldsymbol{\phi}. \quad (19)$$

It is an eigenvalue decomposition problem, and the eigenvector $\boldsymbol{\phi}$ is calculated. Finally, the QSF \mathbf{s} can thus be computed as

$$\mathbf{s} = \mathbf{X}\mathbf{w} = \mathbf{X}\mathbf{H}\boldsymbol{\phi} = \mathbf{X}\mathbf{U}\mathbf{\Lambda}^{-1/2}\boldsymbol{\phi}. \quad (20)$$

2.3. Relevance Vector Regression. After the QSFs have been extracted and selected, a sparse Bayesian framework named relevance vector regression (RVR) is developed for the prediction model. Given a database $\{x_n, t_n\}_{n=1}^N$ with the input data x_n and the corresponding target, the target can be expressed as follows [44, 45]:

$$t_n = y(x_n) + \varepsilon_n, \quad (21)$$

where $\varepsilon_n \sim N(0, \sigma^2)$. The function $y(x)$ is defined as follows:

$$y(x) = \sum_{i=1}^N \omega_i K(x, x_i) + \omega_0 = \sum_{i=1}^N \omega_i \phi_i(x), \quad (22)$$

where $\phi_i(x) = K(x, x_i)$ is the kernel function and $\boldsymbol{\omega} = [\omega_0, \omega_1, \dots, \omega_N]$ is the weighted parameter vector of the kernel function.

The targets is given as $p(t_n | x_n) = N(t_n | y(x_n), \sigma^2)$. The likelihood function of target values $\mathbf{t} = [t_1, t_2, \dots, t_N]$ is written as

$$p(\mathbf{t} | \mathbf{x}, \boldsymbol{\omega}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{t} - \Phi\boldsymbol{\omega}\|^2 \right\}, \quad (23)$$

where Φ is a design matrix consisting of kernel functions. A zero-mean Gaussian distribution is used to constrain the weight parameters $\boldsymbol{\omega}$:

$$p(\boldsymbol{\omega} | \alpha) = \prod_{i=0}^N N(\omega_i | 0, \alpha_i^{-1}), \quad (24)$$

where α is a vector of hyperparameters ($\alpha = [\alpha_0, \alpha_1, \dots, \alpha_N]^T$).

Then, the posterior distribution overweights can be calculated through the Bayesian rule:

$$\begin{aligned} p(\boldsymbol{\omega} | \mathbf{t}, \alpha, \sigma^2) &= \frac{p(\mathbf{t} | \boldsymbol{\omega}, \sigma^2) p(\boldsymbol{\omega} | \alpha)}{p(\mathbf{t} | \alpha, \sigma^2)} \\ &= (2\pi)^{-(N+1)/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\omega} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\omega} - \boldsymbol{\mu}) \right\}, \end{aligned} \quad (25)$$

where the posterior mean and covariance are given by

$$\boldsymbol{\mu} = \sigma^{-2} \Sigma \Phi^T \mathbf{t}, \quad (26)$$

$$\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}, \quad (27)$$

with $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$.

For a new test sample x_* and the corresponding target t_* , the predictive distribution is given by

$$p(t_* | x_*, \mathbf{x}, \mathbf{t}, \alpha_{\text{MP}}, \sigma_{\text{MP}}^2) = \int p(t_* | x_*, \boldsymbol{\omega}, \sigma_{\text{MP}}^2) p(\boldsymbol{\omega} | \mathbf{x}, \mathbf{t}, \alpha_{\text{MP}}, \sigma_{\text{MP}}^2) d\boldsymbol{\omega}, \quad (28)$$

where α_{MP} and σ_{MP}^2 are most probable values for hyper-

parameters and variance, respectively. Using equations (23) and (25), the predictive model is presented as

$$p(t_* | \mathbf{t}) = N(t_* | \boldsymbol{\mu}^T \boldsymbol{\phi}(x_*), \sigma_*^2), \quad (29)$$

where the variance is $\sigma_*^2 = \sigma_{\text{MP}}^2 + \boldsymbol{\phi}(x_*)^T \Sigma \boldsymbol{\phi}(x_*)$.

Thus, a Bayesian prediction model is developed. The sparseness of the model is guaranteed by the constrained weights that follow zero-mean prior distribution. The α and σ^2 are estimated by maximizing $p(\mathbf{t} | \alpha, \sigma^2)$ given by tipping [44]:

$$\begin{aligned} p(\mathbf{t} | \alpha, \sigma^2) &= \int p(\mathbf{t} | \boldsymbol{\omega}, \sigma^2) \cdot p(\boldsymbol{\omega} | \alpha) d\boldsymbol{\omega} \\ &= (2\pi)^{-N/2} |C|^{-1/2} \exp \left\{ \frac{1}{2} \mathbf{t}^T C^{-1} \mathbf{t} \right\}, \end{aligned} \quad (30)$$

where the covariance is given by $C = \sigma^2 I + \Phi A^{-1} \Phi^T$.

Values of α and σ^2 can be calculated by the iterative method:

$$\alpha_i^{\text{new}} = \frac{1 - \alpha_i \Sigma_{ii}}{\mu_i^2}, \quad (31)$$

$$(\sigma^2)^{\text{new}} = \frac{\|\mathbf{t} - \Phi\boldsymbol{\mu}\|^2}{N - \sum_{i=1}^N (1 - \alpha_i \Sigma_{ii})}. \quad (32)$$

2.4. Quality Prediction Method Based on QSFA-RVR. As plenty of data have been sampled from the propylene polymerization process, a soft sensor model based on quality-relevant slow feature analysis and Bayesian regression is constructed to predict the MI value for quality monitoring. The model extracts quality-relevant slow features from the sampled data and built a conditional probability distribution over sparse weight parameters; then, MI values are predicted according to the Bayesian inference. The parameters of the model are estimated by a recursive expectation maximum technique. It is recognized that different QSFs may have different significances and only some of them are critical to quality prediction. The noncritical QSFs should be removed from model development to avoid undesirable disturbance. Here, the cross-validation method is adopted to determine the retained QSFs for Bayesian regression. The flowchart of the QSFA-RVR soft sensing approach for quality prediction is shown in Figure 1. The step-by-step procedures for constructing the quality prediction model are described as follows:

Step 1. Classify the modeling data into two groups, namely, the training dataset $\{\tilde{\mathbf{X}}_{\text{tr}}, \mathbf{y}_{\text{tr}}\}$ and testing dataset $\{\tilde{\mathbf{X}}_{\text{te}}, \mathbf{y}_{\text{te}}\}$.

Step 2. Normalize the training dataset $\{\tilde{\mathbf{X}}_{\text{tr}}, \mathbf{y}_{\text{tr}}\}$ with $\tilde{\mathbf{X}}_{\text{tr}} - u/\sigma$ to obtain $\{\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}}\}$, where u and σ are the mean and standard deviation of $\tilde{\mathbf{X}}_{\text{tr}}$.

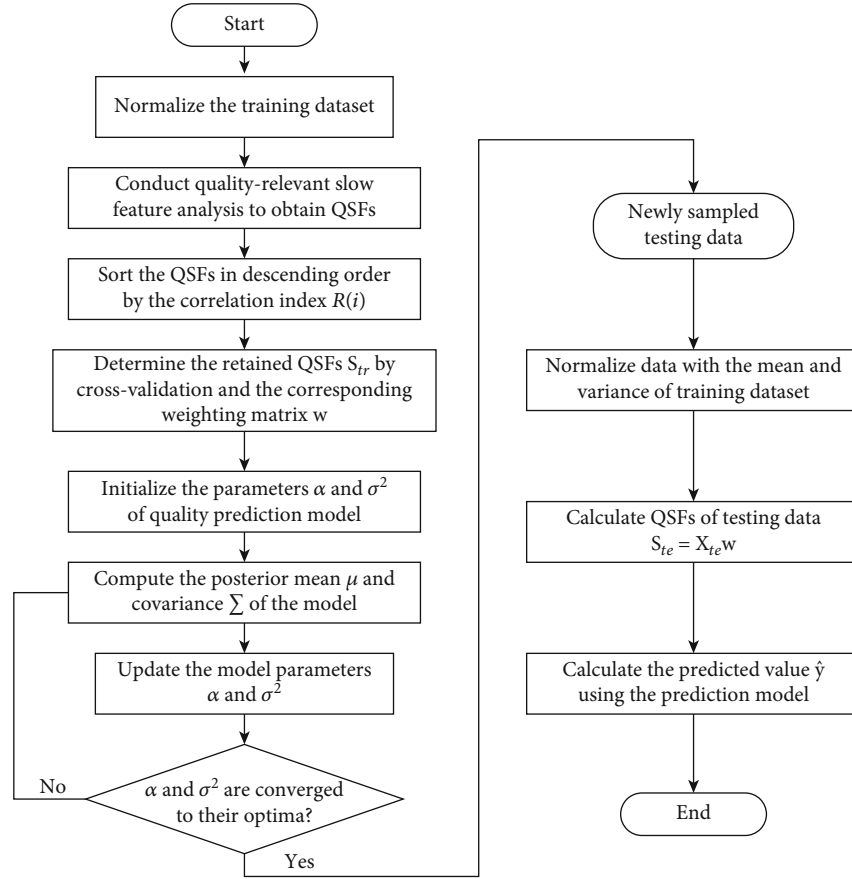


FIGURE 1: Flowchart of the QSFA-RVR soft sensing approach for quality prediction.

Step 3. Conduct quality-relevant slow feature analysis with $\{\mathbf{X}_{tr}, \mathbf{y}_{tr}\}$ and obtain a series of QSFs $\mathbf{s} = \{\mathbf{s}_{tr_1}, \mathbf{s}_{tr_2}, \dots, \mathbf{s}_{tr_j}\}$ according to equations (15)–(20).

Step 4. Sort the QSFs $\mathbf{s} = \{\mathbf{s}_{tr_1}, \mathbf{s}_{tr_2}, \dots, \mathbf{s}_{tr_j}\}$ in descending order evaluated by the correlation index $R(i)$:

$$R(i) = \frac{\mathbf{s}_{tr_i}^T \mathbf{y}}{\sqrt{\text{var}(\mathbf{s}_{tr_i})} \sqrt{\text{var}(\mathbf{y})}}. \quad (33)$$

For brevity, the sorted QSFs are still noted $\mathbf{s} = \{\mathbf{s}_{tr_1}, \mathbf{s}_{tr_2}, \dots, \mathbf{s}_{tr_j}\}$, and the corresponding weighting matrix is saved as $\tilde{\mathbf{w}}$.

Step 5. Determine the retained QSFs $\mathbf{s}_{tr} = \{\mathbf{s}_{tr_1}, \mathbf{s}_{tr_2}, \dots, \mathbf{s}_{tr_M}\}$ for regression modeling using the cross-validation method, where M is the number of retained features. The corresponding weighting matrix is saved as \mathbf{w} .

Step 6. Initialize the parameters α and σ^2 of the quality prediction model.

Step 7. Given the values of parameters α and σ^2 and the training set $\{\mathbf{s}_{tr}, \mathbf{y}_{tr}\}$, compute the posterior mean

μ and covariance Σ of the model according to equations (26) and (27).

Step 8. Given the values of posterior statistics μ and Σ , determine the model parameters α and σ^2 by equations (31) and (32).

Step 9. If values of parameters α and σ^2 are converged to their optima, continue to Step 10. Otherwise, go back to Step 7.

Step 10. For the testing data $\tilde{\mathbf{X}}_{te}$, it is normalized as \mathbf{X}_{te} according to the mean and variance of the training dataset.

Step 11. Calculate the corresponding QSFs $\mathbf{s}_{te} = \{\mathbf{s}_{te_1}, \mathbf{s}_{te_2}, \dots, \mathbf{s}_{te_M}\}$ by projecting the normalized \mathbf{X}_{te} onto the weighting matrix \mathbf{w} , that is, $\mathbf{s}_{te} = \mathbf{X}_{te} \mathbf{w}$.

Step 12. A quality prediction model is formulated by equation (29) with input \mathbf{s}_{te} . An estimated value of quality index \hat{y} could be obtained by the prediction model.

To quantify the prediction accuracy, five performance indices are used for comparisons, namely, the mean absolute error (MAE), the mean relative error (MRE), the root mean square error (RMSE), Theil's inequality coefficient (TIC), and standard deviation of absolute error (STD). The error indicators are defined as follows:

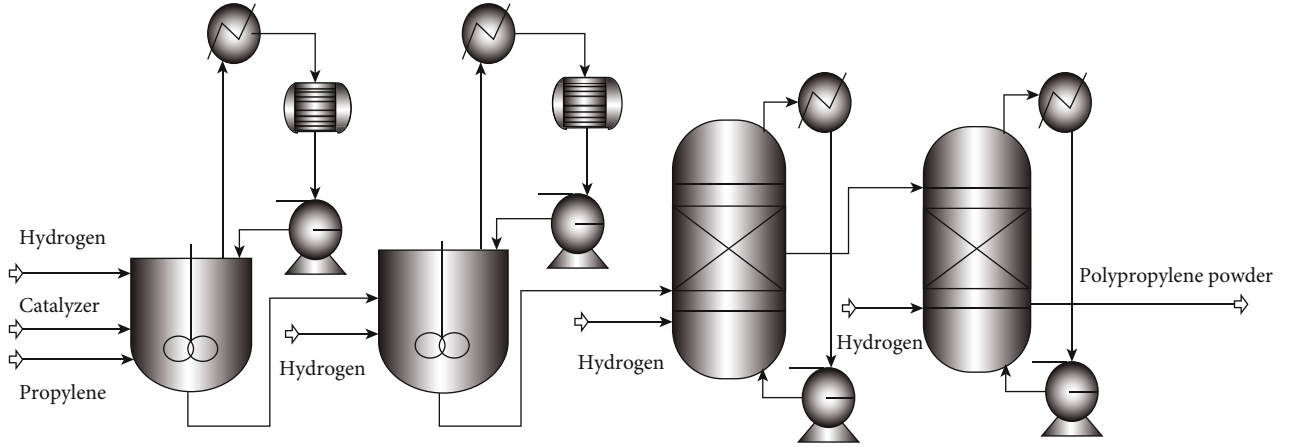


FIGURE 2: Schematic diagram of the industrial propylene polymerization process.

TABLE 1: Process variables in the propylene polymerization process used for modeling.

No.	Measured variables	No.	Measured variables
1	Temperature in the first CSTR	6	Flow rate of feedstock in the first FBR
2	Pressure in the first CSTR	7	Concentration of hydrogen in the first CSTR
3	Liquid level in the first CSTR	8	Flow rate of the main catalyst
4	Flow rate of propylene in the first CSTR	9	Flow rate of the second catalyst
5	Flow rate of feedstock in the second CSTR		

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (34)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \quad (35)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (36)$$

$$TIC = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^n y_i^2} + \sqrt{\sum_{i=1}^n \hat{y}_i^2}}, \quad (37)$$

$$STD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (e_i - \bar{e})^2}, \quad (38)$$

where $e_i = y_i - \hat{y}_i$, $\bar{e} = 1/n \sum_{i=1}^n e_i$, y_i is the real MI value, \hat{y}_i is the predicted MI value, and \bar{y} is the mean value of the output. The MAE, MRE, and RMSE confirm the prediction accuracy of the prediction models. The smaller the value of these indicators is, the higher the accuracy of the prediction model is. The STD indicates the stability of the prediction model. The smaller the value is, the more stable the prediction model is. The TIC indicates a good level of agreement between the proposed model and the studied process.

3. Results and Discussion

The proposed QSFA-RVR model is applied to predict the MI in an industrial propylene polymerization process located in

China. Figure 2 shows the schematic diagram of the industrial process. The process consists of four reactors in series: the first two continuous stirred tank reactors (CSTR) and the last two fluidized-bed reactors (FBR). The feed to the reactor is comprised of propylene, hydrogen, and the Ziegler-Natta catalyst. In the first two reactors, the polymerization reaction takes place in a liquid phase, and in the third and fourth reactors, the reaction is completed in the vapor phase to produce the powdered polymer products. The MI, which depends on the catalyst properties, reactant composition, reactor temperature, and so on, can determine different brands of the products. To develop a prediction model to estimate the MI, a total of nine process variables are chosen as input variables according to the workers' experience and reaction mechanism analysis. The selected variables are listed in Table 1, including the flow rates, temperature, and pressure. The MI is analyzed at a sampling time of 2 hours in the laboratory, and the dataset of nine process variables is acquired from the distributed control system of the polypropylene process. There are 170 datasets, and they are divided into training and testing datasets. The first 119 datasets (about 70%) are used for training, and the remaining 51 datasets (about 30%) are used for testing. It should be noted that the current MI data are from the same grade of polypropylene production, which is a slowly varying dynamic process. The data are filtered to discard abnormal situations and to improve the quality of the prediction results. The variables are normalized with the method of statistical normalization.

First, QSFs are extracted based on the proposed method. To provide a visual picture of extracted QSFs, Figure 3 visualizes all QSFs extracted on the training set for the PP

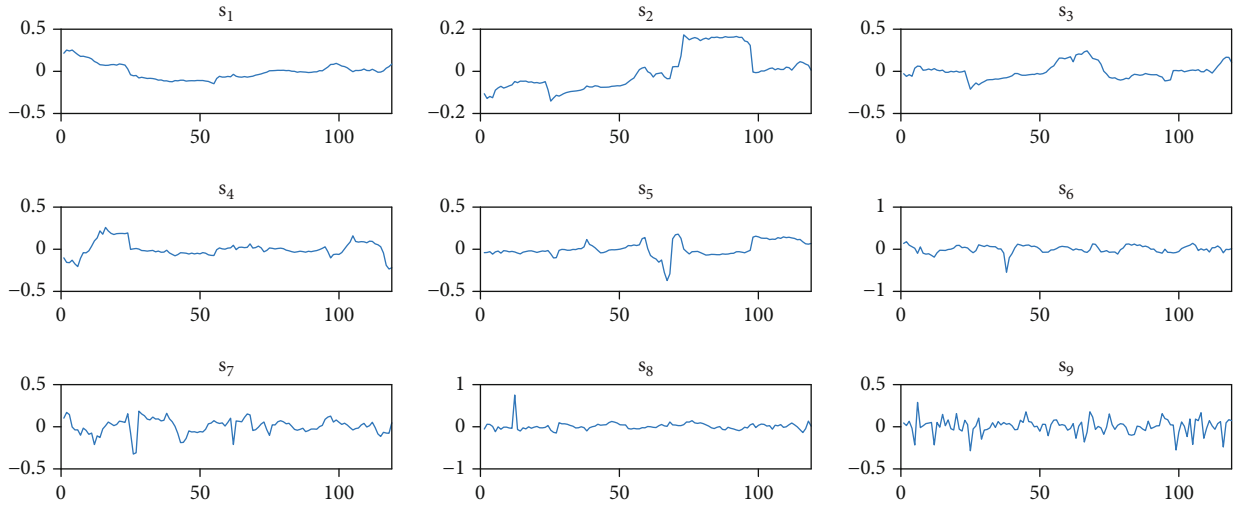


FIGURE 3: QSFs extracted on the training set for PP process.

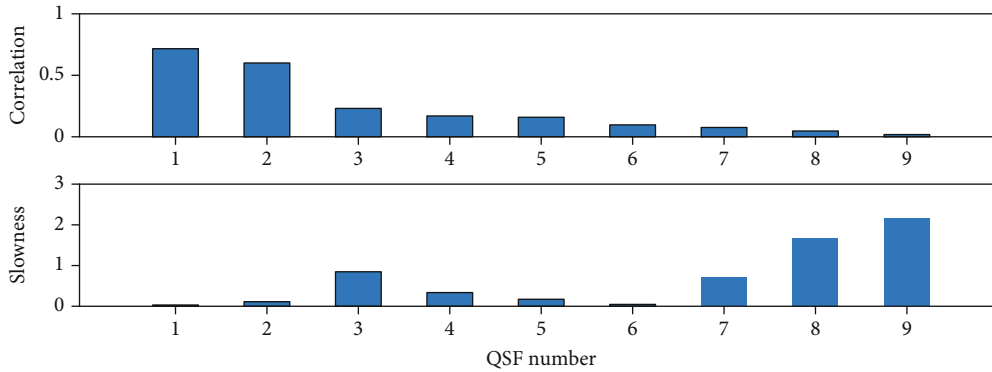


FIGURE 4: Correlation and slowness of QSFs using training data with respect to MI.

process. It is observed that the first six QSFs change more slowly and the slowness decreases. To investigate these QSFs in detail, two quantitative indices are defined, one of which measures the slowness of each QSF and the other measures the correlation with quality. The slowness index is calculated as $S(i) = \Delta(\hat{s}_i)$, which has been defined in equation (10). The correlation index $R(i)$ defined in equation (38) is used to evaluate the correlation between a QSF and a quality index. Then, the slowness of QSFs and correlation coefficients with respect to the quality index, i.e., MI, are calculated and shown in Figure 4. These QSFs are sorted in a descending order and their values of slowness are correspondingly arranged. As can be seen from the figure, the slowness of QSFs is consistent with their quality interpretation. In particular for the first three QSFs, as the correlation weakened, the slowness increased. Compared to the first seven QSFs, the 8th and the 9th QSFs are fast time-varying features and almost irrelevant to the quality index. It reveals that the MI is almost all determined by slowly time-varying QSFs. Thus, a good prediction performance can be achieved based on the extracted QSFs which present both the slowly varying characteristic and the correlations with MI.

For prediction purpose, the cross-validation method is adopted to determine the retained QSFs for Bayesian regression. The number of retained features is $M = 5$. Then, the proposed QSFA-RVR model is built based on the selected QSFs. At the same time, the RVR model, the quality-relevant slow feature regression (QSFR) model, and the PCA-based RVR (PCA-RVR) model have also been developed in comparison with the proposed model. Performances of different models for predicting MI values on the testing dataset are shown in Figure 5. The analytic values obtained from the laboratory are marked with points, and prediction values of RVR, QSFR, PCA-RVR, and QSFA-RVR are marked with triangle, asterisks, squares, and pentagrams, respectively. As can be seen from the figure, the proposed QSFA-RVR model performance is the best among all models. It not only can achieve more satisfactory expected results but also can trace process variations well.

Moreover, another qualitative comparison is displayed in Figure 6. The scattered plots give exhibition on how the predicted values gather around the real values. If the plots are on the black diagonal line, the predicted results are more accurate and vice versa. As illustrated in Figure 6, the

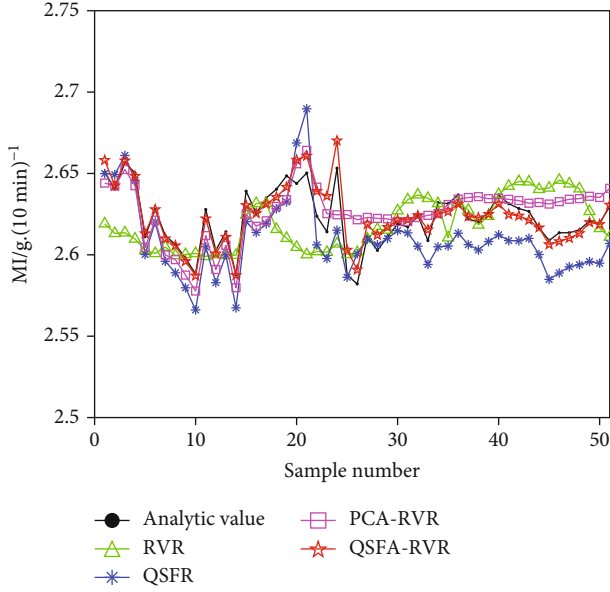


FIGURE 5: Prediction performance of different models on the testing dataset.

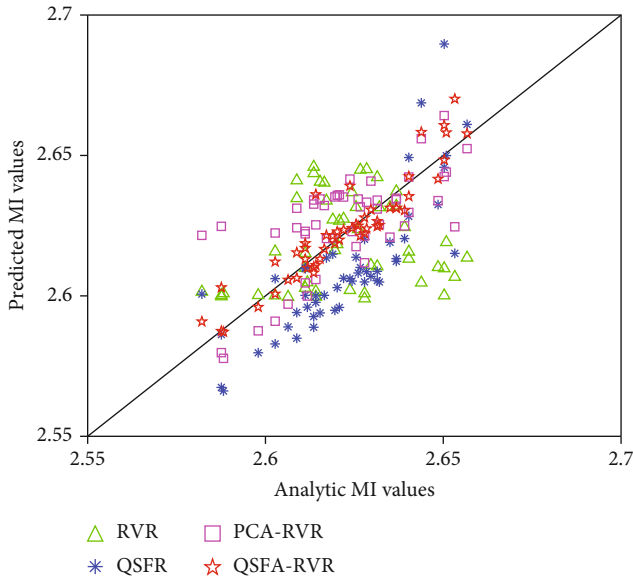


FIGURE 6: Prediction performance of different models on the testing dataset.

proposed QSFA-RVR model has less dispersion of the prediction than the other three models. It generates more precise estimated values as these estimated values almost stay around the black diagonal line.

For comparison, the traditional feature selection method LASSO is considered for soft sensing. The prediction results of different models on the testing dataset are listed in Table 2. It is obvious that the QSFA-RVR model has the best performance over all. In detail, the QSF model obtains an MAE of 0.0168, an MRE of 0.64%, a RMSE of 0.0189, a TIC of 0.0036, and a STD of 0.0139. The PCA-RVR model

TABLE 2: Quality prediction results of different models on the testing dataset.

Model	MAE	MRE (%)	RMSE	TIC	STD
LASSO	0.0187	0.71	0.0233	0.0045	0.0212
RVR	0.0174	0.66	0.0219	0.0042	0.0217
QSF	0.0168	0.64	0.0189	0.0036	0.0139
PCA-RVR	0.0119	0.45	0.0144	0.0027	0.0126
QSFA-RVR	0.0051	0.19	0.0069	0.0013	0.0069

gives an MAE of 0.0119, an MRE of 0.45%, a RMSE of 0.0144, a TIC of 0.0027, and a STD of 0.0126. The QSFA-RVR model shows even better results. The MAE, MRE, RMSE, TIC, and STD are 0.0051, 0.19%, 0.0069, 0.0013, and 0.0069, respectively. The error measurement of the proposed model has percentage decreases of 69.6%, 70.3%, 63.5%, 63.9%, and 50.4% compared to the QSF model and percentage decreases of 57.1%, 57.8%, 52.1%, 51.9%, and 45.2% compared to the PCA-RVR model. The comparison results prove that the soft sensor based on the QSFA-RVR model can achieve a good performance in the MI prediction for the propylene polymerization process.

The above experiments are performed on a personal computer with the configuration shown as follows: operating system: Windows 10 (64-bit); CPU: Intel Core i5-7200U (2.70 GHz); RAM: 8.00 GB; and MATLAB 2017b software. The computation time of the QSF, PCA-RVR, and QSFA-RVR models is 0.96 s, 1.65 s, and 1.94 s, respectively. The prediction time for testing data only needs less than 2 s. Since the sampling time of the industrial MI prediction is about 2 h, the proposed method qualifies the online soft sensor for the MI prediction. The time lags of MI in the form of average residence time do exist and have been considered in our research work.

4. Conclusions

In this paper, a novel soft sensor based on quality-relevant slow feature analysis and Bayesian regression is proposed for quality prediction in industrial PP processes. The proposed method can handle the dynamics of the process better by extracting quality-relevant slow features and deal with the nonlinearity by introducing the Bayesian inference model. A case study about the MI prediction in a real industrial PP plant is carried out to evaluate the performance of the proposed QSFA-RVR method. For comparison, the RVR model, the QSF model, and the PCA-RVR model are also developed and evaluated. The application of the proposed model to the testing dataset demonstrates its superiority. The QSFA-RVR model predicts the MI with an MRE of 0.19%, which is much more accurate than the PCA-RVR model with an MRE of 0.45%, while much better than the QSF model with an MRE of 0.64%. The research results reveal the prediction accuracy and validity of the proposed model, which indicate that the QSFA-RVR modeling approach can be a promising and efficient methodology for industrial MI prediction.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (62003306), the Natural Science Foundation of Zhejiang Province (LQ21F030009), the General Scientific Research Project of Education Department of Zhejiang (Y201941367), and the Zhejiang Public Welfare Technology Application Research Project (LGG19F030005).

References

- [1] L. Shen, Z. Chen, Z. Jiang, and W. Gui, "Soft sensor modeling of blast furnace wall temperature based on temporal-spatial dimensional finite element extrapolation," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, p. 2500314, 2021.
- [2] Z. Ge, "Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes," *Journal of Process Control*, vol. 65, pp. 107–117, 2018.
- [3] X. Yuan, Y. Gu, Y. Wang, C. Yang, and W. Gui, "A deep supervised learning framework for data-driven soft sensor modeling of industrial processes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4737–4746, 2020.
- [4] L. Zhou, J. Zheng, Z. Ge, Z. Song, and S. Shan, "Multimode process monitoring based on switching autoregressive dynamic latent variable model," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 10, pp. 8184–8194, 2018.
- [5] L. Yao and Z. Ge, "Scalable semisupervised GMM for big data quality prediction in multimode processes," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3681–3692, 2019.
- [6] J. Liu, T. Liu, and J. Chen, "Quality prediction for multi-grade processes by just-in-time latent variable modeling with integration of common and special features," *Chemical Engineering Science*, vol. 191, pp. 31–41, 2018.
- [7] Y. Liu, C. Yang, K. Liu, B. Chen, and Y. Yao, "Domain adaptation transfer learning soft sensor for product quality prediction," *Chemometrics and Intelligent Laboratory Systems*, vol. 192, p. 103813, 2019.
- [8] I. A. Udugama, M. Alvarez Camps, M. A. Taube et al., "Novel soft sensor for measuring and controlling product recovery in a high-purity, multicomponent, side-draw distillation column," *Industrial & Engineering Chemistry Research*, vol. 58, no. 43, pp. 20026–20035, 2019.
- [9] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: a novel deep learning strategy," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3721–3730, 2020.
- [10] X. Yuan, L. Li, Y. Shardt, Y. Wang, and C. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Transactions on Industrial Electronics*, vol. 99, pp. 1–1, 2020.
- [11] X. Yuan, Y. Gu, Y. Wang, C. Yang, and W. Gui, "A deep supervised learning framework for data-driven soft sensor modeling of industrial processes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1–10, 2019.
- [12] Q. Jiang, X. Fu, S. Yan et al., "Neural network aided approximation and parameter inference of non-Markovian models of gene expression," *Nature Communications*, vol. 12, no. 1, p. 2618, 2021.
- [13] W. Xie, J. Wang, C. Xing, S. Guo, and L. Zhu, "Variational autoencoder bidirectional long and short-term memory neural network soft-sensor model based on batch training strategy," *IEEE Transactions on Industrial Informatics*, vol. 99, pp. 1–1, 2020.
- [14] J. Wang, W. Shao, and Z. Song, "Semi-supervised variational Bayesian Student's t mixture regression and robust inferential sensor application," *Control Engineering Practice*, vol. 92, pp. 104155.1–104155.15, 2019.
- [15] W. Lou and X. Liu, "Melt index prediction of polypropylene by neural networks model based on PCA-GA-RBF," *Journal of Petrochemical Universities*, vol. 20, no. 3, p. 82, 2007.
- [16] F. Ahmed, L. H. Kim, and Y. K. Yeo, "Statistical data modeling based on partial least squares: application to melt index predictions in high density polyethylene processes to achieve energy-saving operation," *Korean Journal of Chemical Engineering*, vol. 30, no. 1, pp. 11–19, 2013.
- [17] J. Gonzaga, L. Meleiro, C. Kiang, and R. Maciel Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Computers & Chemical Engineering*, vol. 33, no. 1, pp. 43–49, 2009.
- [18] M. Zhang and X. Liu, "A real-time model based on optimized least squares support vector machine for industrial polypropylene melt index prediction," *Journal of Chemometrics*, vol. 30, no. 6, pp. 324–331, 2016.
- [19] L. L. T. Chan and J. Chen, "Melt index prediction with a mixture of Gaussian process regression with embedded clustering and variable selections," *Journal of Applied Polymer Science*, vol. 134, no. 40, p. 45237, 2017.
- [20] Y. Liu and Z. Gao, "Industrial melt index prediction with the ensemble anti-outlier just-in-time Gaussian process regression modeling method," *Journal of Applied Polymer Science*, vol. 132, no. 22, p. 41958, 2015.
- [21] M. Zhang, L. Zhou, J. Jie, and X. Liu, "A multi-scale prediction model based on empirical mode decomposition and chaos theory for industrial melt index prediction," *Chemometrics and Intelligent Laboratory Systems*, vol. 186, pp. 23–32, 2019.
- [22] Y. Liu, Y. Liang, and Z. Gao, "Industrial polyethylene melt index prediction using ensemble manifold learning-based local model," *Journal of Applied Polymer Science*, vol. 134, no. 29, p. 45094, 2017.
- [23] X. Liu and C. Zhao, "Melt index prediction based on fuzzy neural networks and PSO algorithm with online correction strategy," *AIChE Journal*, vol. 58, no. 4, pp. 1194–1202, 2012.
- [24] J. Zhu, M. Jiang, G. Peng, L. Yao, and Z. Ge, "Scalable soft sensor for nonlinear industrial big data via bagging stochastic variational Gaussian processes," *IEEE Transactions on Industrial Electronics*, vol. 99, pp. 1–1, 2020.
- [25] W. Shao, Z. Ge, and Z. Song, "Semi-supervised mixture of latent factor analysis models with application to online key

- variable estimation,” *Control Engineering Practice*, vol. 84, pp. 32–47, 2019.
- [26] Y. Liu, C. Yang, M. Zhang, Y. Dai, and Y. Yao, “Development of adversarial transfer learning soft sensor for multi-grade processes,” *Industrial & Engineering Chemistry Research*, vol. 59, no. 37, pp. 16330–16345, 2020.
- [27] J. Y. Wu, S. Sfarra, and Y. Yao, “Sparse principal component thermography for subsurface defect detection in composite products,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5594–5600, 2018.
- [28] H. Liu, C. Yang, B. Carlsson, S. J. Qin, and C. Yoo, “Dynamic nonlinear partial least squares modeling using Gaussian process regression,” *Industrial & Engineering Chemistry Research*, vol. 58, no. 36, pp. 16676–16686, 2019.
- [29] A. Miranian and M. Abdollahzade, “Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series prediction,” *IEEE Transactions on Neural Networks & Learning Systems*, vol. 24, no. 2, pp. 207–218, 2013.
- [30] L. Wiskott and T. J. Sejnowski, “Slow feature analysis: unsupervised learning of invariances,” *Neural Computation*, vol. 14, no. 4, pp. 715–770, 2002.
- [31] Ha Quang Minh and L. Wiskott, “Multivariate slow feature analysis and decorrelation filtering for blind source separation,” *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2737–2750, 2013.
- [32] W. Nie, A. Liu, Y. Su, and S. Wei, “Multi-view feature extraction based on slow feature analysis,” *Neurocomputing*, vol. 252, pp. 49–57, 2017.
- [33] C. Zhao and B. Huang, “A full-condition monitoring method for nonstationary dynamic chemical processes with cointegration and slow feature analysis,” *AICHE Journal*, vol. 64, no. 5, pp. 1662–1681, 2018.
- [34] S. Zhang and C. Zhao, “Slow-feature-analysis-based batch process monitoring with comprehensive interpretation of operation condition deviation and dynamic anomaly,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3773–3783, 2019.
- [35] C. Shang, F. Yang, B. Huang, and D. Huang, “Recursive slow feature analysis for adaptive monitoring of industrial processes,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 11, pp. 8895–8905, 2018.
- [36] C. Shang, F. Yang, X. Gao, X. Huang, J. A. K. Suykens, and D. Huang, “Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis,” *AICHE Journal*, vol. 61, no. 11, pp. 3666–3682, 2015.
- [37] C. Shang, B. Huang, F. Yang, and D. Huang, “Probabilistic slow feature analysis-based representation learning from massive process data for soft sensor modeling,” *AICHE Journal*, vol. 61, no. 12, pp. 4126–4139, 2015.
- [38] L. Fan, H. Kodamana, and B. Huang, “Identification of robust probabilistic slow feature regression model for process data contaminated with outliers,” *Chemometrics and Intelligent Laboratory Systems*, vol. 173, pp. 1–13, 2018.
- [39] W. Zhong, C. Jiang, X. Peng, Z. Li, and F. Qian, “Online quality prediction of industrial terephthalic acid hydropurification process using modified regularized slow-feature analysis,” *Industrial & Engineering Chemistry Research*, vol. 57, no. 29, pp. 9604–9614, 2018.
- [40] C. Jiang, W. Zhong, Z. Li, X. Peng, and M. Yang, “Real-time semisupervised predictive modeling strategy for industrial continuous catalytic reforming process with incomplete data using slow feature analysis,” *Industrial & Engineering Chemistry Research*, vol. 58, no. 37, pp. 17406–17423, 2019.
- [41] L. Shang, Y. Wang, X. Deng, Y. Cao, P. Wang, and Y. Wang, “A model predictive control performance monitoring and grading strategy based on improved slow feature analysis,” *IEEE Access*, vol. 7, pp. 50897–50911, 2019.
- [42] C. Shang, B. Huang, F. Yang, and D. Huang, “Slow feature analysis for monitoring and diagnosis of control performance,” *Journal of Process Control*, vol. 39, pp. 21–34, 2016.
- [43] G. Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, USA, 2003.
- [44] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [45] S. Kaltwang, S. Todorovic, and M. Pantic, “Doubly sparse relevance vector machine for continuous facial behavior estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1748–1761, 2016.