Control Engineering Practice 61 (2017) 72–80

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Moving window adaptive soft sensor for state shifting process based on weighted supervised latent factor analysis

Le Yao, Zhiqiang Ge*

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control, Science and Engineering, Zhejiang University, Hangzhou 310027, P.R. China

ARTICLE INFO

Keywords:
Adaptive soft sensor
State shifting
Moving window
Weighted supervised latent factor analysis
Expectation Maximization algorithm

ABSTRACT

Process nonlinearity and state shifting are two of the main factors that cause poor performance of online soft sensors. Adaptive soft sensor is a common practice to ensure high predictive accuracy. In this paper, the moving window method is introduced to the supervised latent factor analysis model to capture the state shifting feature of the process. To make the moving window strategy more efficient, the weighted form of the supervised latent factor analysis approach is applied. In this method, contributions of training samples are expressed through incorporating the similarity index into the noise variance of the process variable, which renders strong adaptability of the method for describing nonlinear relationships and abrupt changes of the process. A numerical example and a real industrial process are provided to demonstrate the effectiveness of the proposed adaptive soft sensor.

1. Introduction

As the modern plants become more and more complicated, large amounts of process data are stored in the database through field instruments (Kano & Nakagawa, 2008). The large scale data contains a lot of information about the process, which can be utilized for process monitoring, control, optimization, etc. (Kadlec, Gabrys, & Strandt, 2009; Khatibisepehr, Huang, & Khare, 2013; Zhang, Fan, & Du, 2015). Data-driven soft sensor is a significant application of the data analysis technique, which is built to estimate the values of difficult-to-measure variables by using values of easy-to-measure variables. Compared to traditional first principle modeling methods which typically synthesized prior knowledge or experiences, data-driven soft sensors are more flexible and can be easily deployed in real industrial plants (Kaneko & Funatsu, 2014a; Zhou, Lu, & Chai, 2014).

Generally, principal component regression (PCR) (Barshan, Ghodsi, & Azimifar, 2011; Yuan, Ge, & Song, 2014) and partial least squares (PLS) (Janik, Skjemstad, & Shepherd, 2007; Qin, 1998) are two of the most widely used linear methods for data-driven soft sensor modeling as they are simply constructed and easy to implement. Besides, the nonlinear modeling methods like artificial neural network (ANN) (Hoskins & Himmelblau, 1988; Tu, 1996) and support vector machine (SVM) (Kaneko & Funatsu, 2014b; Yu, 2012a) are also popular for soft sensing of nonlinear processes. In recent years, the probabilistic modeling methods have become a tendency. Among the numerous probabilistic methods, the probabilistic principal component regression (PPCR) is the most widely used method, which preferably gives the prediction results in a probabilistic manner (Ge, Gao, & Song, 2011; Lawrence, 2005). However, the traditional PPCR method assumes that the process variables varies in a homogeneous noise level. To relax this limitation, the latent factor analysis (LFA) model has been artfully extended to the supervised form and used for soft sensing with full labeled data, which takes different noise levels of process variables into account (Ge, 2015).

However, in real industrial plants, the process state often varies in different levels according to the different manufacture conditions (Grbić, Slišković, & Kadlec, 2013; Kaneko & Funatsu, 2014c; Khatibisepehr, Huang, & Xu, 2012; Zhang & Li, 2013). Generally, model degradation is one of the critical problems that traditional soft sensors have to face, which can lead to a deterioration in prediction performance (Kadlec, Grbić, & Gabrys, 2011; Kaneko & Funatsu, 2013). This often occurs after a state shifting in a plant, such as changes in raw materials, catalyst performance deterioration, fouling on pipes, and equipment degradation. However, conventional soft sensor is generally developed on a steady process state, the model only learns information of one mode. Even if the model is developed successfully, its performance deteriorates when process state shifts, since the model lacks information of the new state. Such a situation would affect the evaluation of the product quality, and if an incident is not detected in-time, the safety of plant would not be ensured. In some

* Corresponding author.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

http://dx.doi.org/10.1016/j.conengprac.2017.02.002
Received 28 May 2016; Received in revised form 10 January 2017; Accepted 2 February 2017
0967-0661/

way, the degradation of model limits the practical use of soft sensors in chemical industrial processes (Yu, 2012b; Yu, Chen, & Rashid, 2013).

In order to solve this problem, adaptive soft sensors have been proposed by using new measurement data. Typical adaptive soft sensors are constructed based on methods such as the just-in-time (JIT) (Fujiwara, Kano, Hasebe, & Takinami, 2009; Jin, Chen, & Yang, 2014; Liu & Chen, 2013), time difference (TD) (Kaneko & Funatsu, 2011a, 2011b) and moving window (MW) (Du, Liang, & Jiang, 2004; Liu, Chen, & Shen, 2010) approaches. The JIT models have been widely utilized in soft sensor modeling, which are constructed with selected training samples. It means that training samples similar to the query sample are selected to build the local predictive model (Ge & Song, 2010; Yao & Ge, 2017). JIT method needs a selected dataset from the database of Distributed Control System (DCS) which contains information of the whole process. Generally, the selected dataset is very large and searching for similar samples from it is always time consuming. The TD model could take effect when the drifts and gradual changes in the state of a plant without reconstruction of the model. Generally, soft sensor models are regarded as a mathematic functions, whose input variables and output variable are commonly defined as the x-variables and y-variable, respectively. In TD-based soft sensor models, the differential of x-variables is taken as the input of the model to estimate the difference of y-variable, and the difference is added to the predicted value of y-variable. The shifting value of y-variable can be taken as a bias and eliminated while generating the output of the model (Kaneko & Funatsu, 2011c). Nevertheless, if the relationship between x-variables and the y-variable is changed, TD method cannot adapt to the value variation any more. For MW method, the dataset in the window is updated through incorporating the newly measured samples and discarding the oldest samples. In most cases, the dataset selected in the window as these are assumed to be the most relevant to the current process character. Every time the window updates, the latest information of the process is obtained, on which the model built can effectively describe the current state even the process starts to shift gradually (Ni, Tan, & Ng, 2012). Comparing with the whole-dataset modeling method, learning model parameters with a short-fixed dataset can greatly cut down the modeling time. Nevertheless, it is difficult for MW model to handle strong nonlinear process with a rapid or instant change because the model is affected by old data before the change. Aiming at this problem, a weighted method is proposed in this paper as the basic modeling approach to reduce the influence of abruption.

In this paper, an adaptive soft sensor based on supervised latent factor analysis (SLFA) method is proposed for the state shifting process. Two datasets should be declared before describing the modeling method: the training samples and the query sample. The training samples contain both process variables and quality variable, which are utilized to train a regression model. In the regression model, the process variables are taken as the inputs and the quality variable is the output. However, the query sample is merely composed by process variables and its quality variable value is needed to be estimated by the regression model. After a period of time, the real value of the query sample can be obtained from the laboratory analysis. At the beginning of the approach, the basic SLFA model is trained by the original training samples collected from the database. While there is a query sample comes, the value of quality variable can be predicted by the trained model. To introduce the latest process information into the soft sensor model without increasing the computation burden, the moving window method is utilized to update the training samples through adding the newly acquired samples and removing the oldest samples. The model trained by the updated training samples is called the local model which completely represents the current process state. As the window slides, the local model is updated constantly, which renders the soft sensor a strong adaptability to the state shifting of the process. Therefore, through composing a set of local linear models, a nonlinear process can be approximately described. To further increase the prediction accuracy of the soft sensor, the similarities between the training samples and the query sample are considered while building the local model. Empirically, training data samples have different levels of similarity to the query sample, thus the weights are assigned differently to the training samples. Through adding the weighted method, the prediction accuracy of the moving window based SLFA model can be greatly improved. Unlike the JIT method, this kind of similarity calculating method is conducted in a window, which greatly reduces the computation time of searching samples in a huge database.

The layout of this paper is given as follows. In Section 2, the supervised latent factor analysis (SLFA) model is briefly introduced. Then, the moving window based weighted supervised latent factor analysis (MW-WSLFA) method is proposed and the detailed description of parameter identification under the probabilistic framework is presented in Section 3. Next to that, the adaptive soft sensor is constructed based on MW-WSLFA. Two case studies are utilized for performance evaluation in Section 5. Finally, conclusions are made.

2. Supervised latent factor analysis

The supervised latent factor analysis (SLFA) model is established under a probabilistic framework, which tries to build a relationship between full labeled input and output dataset (Ge, 2015). The input and output datasets are defined as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in R^{m\times n}$ and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n] \in R^{r\times n}$, respectively, where m is the variable number of input dataset $\mathbf{X}$, r is the variable number of output dataset $\mathbf{Y}$, and n is the sample number of each variable. The SLFA model structure can be expressed as follows

$$\mathbf{X} = \mathbf{Pt} + \mathbf{e} \tag{1}$$

$$\mathbf{Y} = \mathbf{Ct} + \mathbf{f} \tag{2}$$

where $\mathbf{P} \in R^{m\times k}$, $\mathbf{C} \in R^{r\times k}$ are factor loading matrices of dataset $\mathbf{X}$ and $\mathbf{Y}$, respectively. A common variable $\mathbf{t} \in R^{k\times n}$ called latent factor is introduced to interpret the relationship between input and output variables, where k is the number of latent factors. For SLFA model, the variances of measured noise $\mathbf{e}$ are different for sampled variables, which are represented by $\mathbf{\Sigma} = diag\{\sigma_q^2\}_{q=1,2,\ldots,m}$. Specially, if all σ_q^2 equal the same value, the SLFA model is equivalent to probabilistic principal component regression (PPCR) model.

In SLFA model, the probability distribution of the latent factor and the measured noise are Gaussian. Then $p(\mathbf{t}) = N(0, \mathbf{I})$, $p(\mathbf{e}) = N(0, \mathbf{\Sigma}_\mathbf{x})$ and $p(\mathbf{f}) = N(0, \mathbf{\Sigma}_\mathbf{y})$, where $\mathbf{\Sigma}_\mathbf{x} = diag\{\sigma_{q,\mathbf{x}}^2\}_{q=1,2,\ldots,m}$ and $\mathbf{\Sigma}_\mathbf{y} = diag\{\sigma_{q,\mathbf{y}}^2\}_{q=1,2,\ldots,m}$. For measured variables, the marginal probability $p(\mathbf{x}, \mathbf{y})$ can be formulated as follows

$$\begin{aligned} & p(\mathbf{x}, \mathbf{y}|\mathbf{P}, \mathbf{C}, \mathbf{\Sigma}_\mathbf{x}, \mathbf{\Sigma}_\mathbf{y}) \\ &= \int p(\mathbf{x}, \mathbf{y}|\mathbf{t}, \mathbf{P}, \mathbf{C}, \mathbf{\Sigma}_\mathbf{x}, \mathbf{\Sigma}_\mathbf{y})p(\mathbf{t})d\mathbf{t} \\ &= \int p(\mathbf{x}|\mathbf{t}, \mathbf{P}, \mathbf{\Sigma}_\mathbf{x})p(\mathbf{y}|\mathbf{t}, \mathbf{C}, \mathbf{\Sigma}_\mathbf{y})p(\mathbf{t})d\mathbf{t} \end{aligned} \tag{3}$$

where the variables $\mathbf{x}$ and $\mathbf{y}$ are conditionally independent given the latent factor. Based on the noise probability function, the conditional probability of $\mathbf{x}$ and $\mathbf{y}$ can be calculated as $p(\mathbf{x}|\mathbf{t}, \mathbf{P}, \mathbf{\Sigma}_\mathbf{x}) = N(\mathbf{Pt}, \mathbf{\Sigma}_\mathbf{x})$ and $p(\mathbf{y}|\mathbf{t}, \mathbf{C}, \mathbf{\Sigma}_\mathbf{y}) = N(\mathbf{Ct}, \mathbf{\Sigma}_\mathbf{y})$, respectively. For the measured dataset $\mathbf{X}$ and $\mathbf{Y}$, the parameter set of latent factor analysis model $\mathbf{\Theta}=\{\mathbf{P}, \mathbf{C}, \mathbf{\Sigma}_\mathbf{x}, \mathbf{\Sigma}_\mathbf{y}\}$ is commonly identified through an efficient EM algorithm.

In the E-step of the EM algorithm, the posterior probability of the latent factor is determined based on the parameters $\mathbf{\Theta}^s$ obtained in the previous M-step. According to the Bayes' rule, the posterior probability of the latent factor can be calculated as follows:

$$p(\mathbf{t}|\mathbf{x}, \mathbf{y}, \mathbf{\Theta}^s) = p(\mathbf{x}|\mathbf{t}, \mathbf{P}, \mathbf{\Sigma}_\mathbf{x})p(\mathbf{y}|\mathbf{t}, \mathbf{C}, \mathbf{\Sigma}_\mathbf{y})p(\mathbf{t}) \tag{4}$$

Since the terms on the right side of Eq. (4) are Gaussian distributed, the posterior probability of the latent factor is also Gaussian, and the sufficient statistics can be calculated as follows: