Computers and Chemical Engineering 104 (2017) 164–171

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Locally weighted kernel partial least squares regression based on sparse nonlinear features for virtual sensing of nonlinear time-varying processes

Xinmin Zhang[a], Manabu Kano[a,*], Yuan Li[b]

[a] Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan
[b] Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, PR China

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form 3 April 2017
Accepted 12 April 2017
Available online 21 April 2017

Keywords:
Soft-sensor
Virtual sensing
Nonlinear time-varying processes
SKFCFs
Instance-wise kernelized elastic net learning
Locally weighted kernel partial least squares

ABSTRACT

Virtual sensing technology is crucial for monitoring product quality when real-time measurement is not available. To deal with both strong nonlinearity and time-varying dynamics of industrial processes, we propose a novel locally weighted kernel PLS (LW-KPLS) based on sparse nonlinear features in this research. Unlike the conventional locally weighted PLS (LW-PLS), the proposed method weights the training samples by using sparse kernel feature characterization factors (SKFCFs), which take account of the strength of nonlinear dependency between samples in the Hilbert feature space. By integrating the nonlinear features into the locally weighted regression framework, LW-KPLS not only can cope with the time-varying characteristics but also is more suitable for highly nonlinear processes. The proposed method was validated through a numerical example, a penicillin fermentation process, and a real industrial cleaning process for residual drug substances. The results have demonstrated that the proposed LW-KPLS outperforms the conventional PLS, KPLS, LW-PLS, and eLW-KPLS in the prediction performance.

1. Introduction

In modern industrial processes, it is important to monitor the product quality and other key variables in order to assure both the product quality and the process safety and reduce energy and material consumption. Hardware analyzers have practical problems such as time-consuming maintenance, need for calibration, aged deterioration, insufficient accuracy, long dead-time, and slow dynamics as clarified through the questionnaire survey in Japan (Kano and Fujiwara, 2013). Therefore soft-sensors have been widely used over the past decades for quality control and process monitoring. Among them, data-driven soft-sensors have attracted increasing attention because a massive amount of process data is becoming available in industry. Soft-sensors have been successfully applied to various industrial processes. Multiple linear regression (MLR), principal component regression (PCR), and partial least squares (PLS) regression are the most popular modeling approaches (Kadlec et al., 2009; Kano and Ogawa, 2010). However, these linear regression methods are based on a linearity assumption, which limits their applications to nonlinear industrial processes.

To handle the nonlinearity, a series of nonlinear regression methods such as nonlinear PLS (NLPLS) (Wilson et al., 1997), artificial neural networks (Zupan and Gasteiger, 1991), and kernel regression methods such as kernel PLS (KPLS) (Rosipal and Trejo, 2001; Zhang et al., 2010) have been developed. Among these methods, kernel methods avoid nonlinear optimization through the introduction of the nonlinear transformation kernel function. KPLS maps data points from the original space to the Hilbert feature space, where a linear PLS model is developed. According to the Covers theorem (Rosipal and Trejo, 2001), the nonlinear relationship among variables in the original space is most likely to be linear after high-dimensional nonlinear mapping. Thus, KPLS can efficiently capture the nonlinearity and improve prediction performance. KPLS has been successfully applied to many industrial processes (Zhang et al., 2010).

Generally, industrial processes are time-varying due to changes of process characteristics and operating conditions. For example, equipment characteristics are changed by catalyst deactivation, scale adhesion, and equipment aging in chemical processes (Kano and Fujiwara, 2013). Such changes in process characteristics cause

* Corresponding author.
E-mail addresses: zhang.xinmin.58s@st.kyoto-u.ac.jp (X. Zhang), manabu@human.sys.i.kyoto-u.ac.jp (M. Kano), li-yuan@mail.tsinghua.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.compchemeng.2017.04.014
0098-1354/

accuracy deterioration of soft-sensors, and the model maintenance is the most important issue concerning soft-sensors in the process industry (Kano and Ogawa, 2010). Hence, it is necessary to update soft-sensors automatically. Recursive methods such as recursive PLS and recursive support vector regression (SVR) have been developed to adapt a prediction model to a new operating condition recursively (Helland et al., 1992). The recursive methods update the covariance matrices when a new data becomes available, but they cannot deal with abrupt changes (Kano and Fujiwara, 2013). Alternatively, just-in-time (JIT) modeling methods have been developed to cope with changes in process characteristics as well as nonlinearity, and they have been widely used for virtual sensing and process monitoring (Fujiwara et al., 2009; Ge and Song, 2010; Hirai and Kano, 2015; Kim et al., 2356; Xie et al., 2014; Zhang et al., 2015). In JIT modeling, a local model is built from a historical dataset using the most relevant samples around one query sample, which is a target sample, when an estimated output value corresponding to the query is required. Thus, it can trace operating conditions and cope with the process nonlinearity. Locally weighted regression (LWR) (Shigemori et al., 2011) and locally weighted PLS (LW-PLS) are JIT modeling methods that have successful industrial applications (Hirai and Kano, 2015; Kim et al., 2011; Nakagawa et al., 2012). In LWR and LW-PLS, the prediction performance is mainly dependent on the definition of similarity or weights of samples. To design prediction models, a variety of weights have been investigated, which generally can be classified into distance-based weights (Leung et al., 2004), distance-and-angle-based weights (Ge and Song, 2008), correlation-based weights (Fujiwara et al., 2009), covariance-based weights (Hazama and Kano, 2015), and regression-coefficient-based weights (Kim et al., 2011; Shigemori et al., 2011). To build good local models, it is crucial to appropriately set the weights according to the strength of nonlinearity. However, nonlinearity may change as a function of operating condition or time. An adaptive version of LW-PLS was proposed, which adaptively determine the weights according to the strength of nonlinearity between each input variable and an output variable around a query (Kim et al., 2013b). However, most local regression methods cannot cope with changes in nonlinearity. Consequently, they do not always function well. In addition, input features used for the conventional local regression methods are still original process variables, and the nonlinearity of regression models only rests in the different treatment of samples, i.e. local weighting. Hence, the conventional methods do not necessarily meet the demand for the prediction accuracy especially for highly nonlinear industrial processes.

To solve these issues, a novel locally weighted kernel partial least squares (LW-KPLS) regression method based on sparse nonlinear features is proposed in this work. Unlike the conventional LW-PLS, the proposed LW-KPLS method uses sparse kernel feature characterization factors (SKFCFs) to weight the training samples. SKFCFs are derived from a instance-wise kernelized elastic net learning (IW-KENL) model and describe nonlinear dependency between the query and training samples in the Hilbert feature space. SKFCFs are essential to understand which samples are important to construct a locally weighted model for the current query, and which samples provide irrelevant or redundant information and can be eliminated. Thus SKFCFs are useful to weight the training samples and construct a LW-KPLS model. Furthermore, by integrating the nonlinear features into the locally weighted regression framework, the proposed LW-KPLS not only is more suitable for highly nonlinear processes than the conventional LW-PLS but also can cope with time-varying characteristics.

The rest of this paper is organized as follows. Section 2 gives a brief introduction of LW-PLS and kernelized elastic net. Then, locally weighted kernel PLS based on sparse nonlinear features is presented in Section 3. In Section 4, the proposed method is applied to a numerical example, a penicillin fermentation process, and a real industrial cleaning process for residual drug substances, and its application results are compared with PLS, KPLS, LW-PLS, and eLW-KPLS. The conclusions of this work are presented in Section 5.

2. Existing methods

2.1. Locally weighted PLS (LW-PLS)

LW-PLS (Kim et al., 2011) is one kind of JIT modeling methods, in which PLS is used to construct local regression models based on the similarity between the query and historical samples. Here the basic algorithm of LW-PLS is briefly explained.

The nth samples ($n = 1, 2, \ldots, N$) of input and output variables are denoted by

$$\boldsymbol{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]^T \tag{1}$$

$$\boldsymbol{y}_n = [y_{n1}, y_{n2}, \ldots, y_{nL}]^T \tag{2}$$

where M and L denote the numbers of input and output variables, respectively. $\boldsymbol{X} \in \Re^{N\times M}$ and $\boldsymbol{Y} \in \Re^{N\times L}$ are the input and output variable matrices whose nth row are $\boldsymbol{x}_n^T$ and $\boldsymbol{y}_n^T$. N is the number of the samples.

In LW-PLS, data matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are stored in a database. When an output estimation is required for a query sample $\boldsymbol{x}_q$, the similarity ω_n between $\boldsymbol{x}_q$ and $\boldsymbol{x}_n$ is calculated, then a local PLS model is constructed by weighting samples with a similarity matrix $\boldsymbol{\Omega} \in \Re^{N\times N}$ defined by

$$\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_N) \tag{3}$$

where diag(•) represents a diagonal matrix.

The output estimate $\hat{\boldsymbol{y}}_q$ corresponding to the query sample $\boldsymbol{x}_q$ is calculated as follows.

(1) Determine the number of latent variables R, and set $r = 1$.
(2) Calculate the weight matrix $\boldsymbol{\Omega}$.
(3) Calculate $\boldsymbol{X}_r$, $\boldsymbol{Y}_r$, and $\boldsymbol{x}_{q,r}$

$$\boldsymbol{X}_r = \boldsymbol{X} - \mathbf{1}_N[\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M] \tag{4}$$

$$\boldsymbol{Y}_r = \boldsymbol{Y} - \mathbf{1}_N[\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_L] \tag{5}$$

$$\boldsymbol{x}_{q,r} = \boldsymbol{x}_q - [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M]^T \tag{6}$$

$$\bar{x}_m = \frac{\sum_{n=1}^{N} \omega_n x_{nm}}{\sum_{n=1}^{N} \omega_n} \tag{7}$$

$$\bar{y}_l = \frac{\sum_{n=1}^{N} \omega_n y_{nl}}{\sum_{n=1}^{N} \omega_n} \tag{8}$$

where $\mathbf{1}_N \in \Re^N$ is a vector of ones.
(4) Derive the rth latent variable of $\boldsymbol{X}$

$$\boldsymbol{t}_r = \boldsymbol{X}_r \boldsymbol{w}_r \tag{9}$$

where $\boldsymbol{w}_r$ is the eigenvector of $\boldsymbol{X}_r^T \boldsymbol{\Omega} \boldsymbol{Y}_r \boldsymbol{Y}_r^T \boldsymbol{\Omega} \boldsymbol{X}_r$, which corresponds to the maximum eigenvalue.
(5) Derive the rth loading vector of $\boldsymbol{X}$ and the regression coefficient vector.

$$\boldsymbol{p}_r = \frac{\boldsymbol{X}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r}{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r} \tag{10}$$

$$\boldsymbol{q}_r = \frac{\boldsymbol{Y}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r}{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r} \tag{11}$$

(6) Derive the rth latent variable of $\boldsymbol{x}_q$.

$$t_{q,r} = \boldsymbol{x}_{q,r}^T \boldsymbol{w}_r \tag{12}$$