Chemometrics and Intelligent Laboratory Systems 137 (2014) 57–66

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants

Hiromasa Kaneko, Kimito Funatsu *

Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:
Received 8 May 2014
Received in revised form 11 June 2014
Accepted 12 June 2014
Available online 21 June 2014

Keywords:
Process control
Soft sensor
Degradation
Online support vector regression
Ensemble learning
Bayes' rule

ABSTRACT

A soft sensor predicts the values of some process variable **y** that is difficult to measure. To maintain the predictive ability of a soft sensor model, adaptation mechanisms are applied to soft sensors. However, even these adaptive soft sensors cannot predict the **y**-values of various process states in chemical plants, and it is difficult to ensure the predictive ability of such models on a long-term basis. Therefore, we propose a method that combines online support vector regression (OSVR) with an ensemble learning system to adapt to nonlinear and time-varying changes in process characteristics and various process states in a plant. Several OSVR models, each of which has an adaptation mechanism and is updated with new data, predict **y**-values. A final predicted **y**-value is calculated based on those predicted **y**-values and Bayes' rule. We analyze a numerical dataset and two real industrial datasets, and demonstrate the superiority of the proposed method.

1. Introduction

Soft sensors can predict process variables that are difficult to measure online, and have been widely used in industrial modeling [1–3]. The relationship between a difficult-to-measure variable (**y**-variable) and easy-to-measure variables (**X**-variables) is constructed theoretically or statistically. In practice, data-driven soft sensors, in which models are built using statistical methods and databases, are mainly employed. By inputting values of the **X**-variables into the constructed model, the values of a **y**-variable can be predicted online, because the **X**-variables are measured in real time and the prediction is very fast.

However, soft sensors have some practical problems. One crucial issue is the degradation of the soft sensor models [4]. The predictive accuracy of soft sensors decreases gradually, a result of the changing state of a chemical plant due to factors such as catalyzing performance loss, and sensor and process drift.

To reduce this degradation, adaptation mechanisms can be applied to soft sensors [5]. For example, new **X**-variable data are measured in a chemical plant and used to reconstruct the soft sensor models and predict the **y**-variables. Kaneko and Funatsu categorized the degradation of soft sensor models and adaptive soft sensor models, such as moving window (MW), just-in-time (JIT), and time difference (TD) models [4]. MW models [6,7] are constructed with a recently measured dataset; JIT models [8–10] are constructed by assigning larger weights to the data that are most similar to the prediction data; and TD models [11,12] are constructed by considering the time difference of a **y**-variable and that of the **X**-variables. Ensemble learning can be applied to adaptive models [13,14]. In addition, when data distributions are multimodal, multiple modeling approaches [15,16] can be combined with adaptive models.

The characteristics of MW, JIT, and TD models have been discussed and analyzed using numerical and real industrial data [4]. These models are not entirely sufficient when rapid temporal changes occur in a process, and thus, novel techniques are required.

We previously combined an online support vector regression (OSVR) method [17], in which a nonlinear SVR [18] model is efficiently updated (see Appendix A), with a time variable [19], and discussed the optimization of the SVR hyperparameters for the OSVR model. The SVR hyperparameters can be selected using cross-validation on training data of reasonable length, and the OSVR model with this time variable can adapt to abrupt changes in process characteristics, even when the relationship between **X** and **y** is nonlinear and time varying [20].

However, one soft sensor model cannot accurately predict **y**-values for all process states in a chemical plant. An OSVR model must first set the SVR hyperparameters C, ε, and γ using a training dataset (see Appendix A), and the resulting model will be able to predict **y**-values for process states that are similar to that of the training data. However, when the process state changes significantly, the predictive ability of the OSVR model decreases.

We therefore propose a combination of more than one OSVR model, and use ensemble learning to allow the soft sensors to adapt to various process states in a chemical plant. Multiple OSVR models with different hyperparameters predict multiple **y**-values. We combine the predicted

* Corresponding author. Tel.: +81 3 5841 7751; fax: +81 3 5841 7771.
E-mail address: funatsu@chemsys.t.u-tokyo.ac.jp (K. Funatsu).

http://dx.doi.org/10.1016/j.chemolab.2014.06.008
0169-7439/

y-values based on the current predictive ability of each OSVR model and Bayes' rule, which has been actively applied to soft sensor analyses [16, 21–23], to produce a final predicted **y**-value. The proposed method is referred to as ensemble online support vector regression (EOSVR). The predictive ability of multiple OSVR models updated with new data, and weighted appropriately, allows the accurate prediction of **y**-values in each process state. In addition, the standard deviation of the predicted **y**-values enables us to estimate the prediction error in the final predicted **y**-value for each process state.

To verify the effectiveness of the proposed method, we analyze numerical simulation data in which the relationships between **X** and **y** are strongly nonlinear and change from moment to moment. The performance of the proposed model is compared with that of other traditional adaptive models. Then, the proposed method is applied to two real industrial processes, an alkylaluminum production process and an exhaust gas denitration process.

2. Method: EOSVR

2.1. Offline analysis

In the proposed EOSVR method, multiple combinations of the SVR hyperparameters C, ε, and γ (see Appendix A) are optimized offline for various states in a plant. The basic concept of the optimization of multiple combinations of SVR hyperparameters is shown in Fig. 1. First, the window size *ws* is set, and then the SVR hyperparameters are optimized by moving the window. The value of *ws* is relatively unimportant, as OSVR models are insensitive to this parameter [20]. A certain size of *ws* suffices for OSVR models. The data in each window are then as follows:

$$(\mathbf{X}_1, \mathbf{y}_1), (\mathbf{X}_2, \mathbf{y}_2), \ldots, (\mathbf{X}_i, \mathbf{y}_i), \ldots, (\mathbf{X}_n, \mathbf{y}_n) \tag{1}$$

where $\mathbf{X}_i \in R^{ws \times v}$ and $\mathbf{y}_i \in R^{ws \times 1}$ are the ith data set of the **X**-variables and that of a **y**-variable, respectively (v is the number of **X**-variables). When the window is moved by h data points, the ith dataset is from the $h(i-1)+1$th data point to the hith data point. For each dataset, the SVR hyperparameters are optimized. The optimized hyperparameters are represented as follows:

$$(C_1, \varepsilon_1, \gamma_1), (C_2, \varepsilon_2, \gamma_2), \ldots, (C_i, \varepsilon_i, \gamma_i), \ldots, (C_n, \varepsilon_n, \gamma_n). \tag{2}$$

To optimize the SVR hyperparameters, we use an exhaustive grid search through the candidates of each hyperparameter, and the performance is measured by cross-validation on the training data. Duplicate combinations of the SVR parameters are removed, leaving the following m combinations of parameters:

$$(C_1, \varepsilon_1, \gamma_1), (C_2, \varepsilon_2, \gamma_2), \ \ldots, (C_m, \varepsilon_m, \gamma_m). \tag{3}$$

2.2. Online analysis

Fig. 2 shows the basic concept of prediction using the proposed method. New **X**-variable data $\mathbf{x}(t)$ at time t are input into the m OSVR models; the models predict **y**-values as follows:

$$\begin{aligned} y_{\mathrm{p},1}(t) &= f_1(\mathbf{x}(t)) \\ y_{\mathrm{p},2}(t) &= f_2(\mathbf{x}(t)) \\ &\vdots \\ y_{\mathrm{p},m}(t) &= f_m(\mathbf{x}(t)), \end{aligned} \tag{4}$$

where $f_1, f_2, \ldots, f_m$ are the OSVR models. We then combine $y_{\mathrm{p},1}(t), y_{\mathrm{p},2}(t), \ldots, y_{\mathrm{p},m}(t)$ to obtain a final predicted **y**-value $y_{\mathrm{p}}(t)$ using Bayes' rule [22]. When S is the current (unobserved) state in a plant and M_i is the ith OSVR model, the probability of M_i given S, $P(\mathrm{M}_i|\mathrm{S})$, is required to combine the prediction results of the m OSVR models. Given $P(\mathrm{M}_i|\mathrm{S})$, the final predicted **y**-value $y_{\mathrm{p}}(t)$ is obtained as follows:

$$y_{\mathrm{p}}(t) = \sum_{i=1}^{m} P(\mathrm{M}_i|\mathrm{S}) y_{\mathrm{p},i}(t). \tag{5}$$

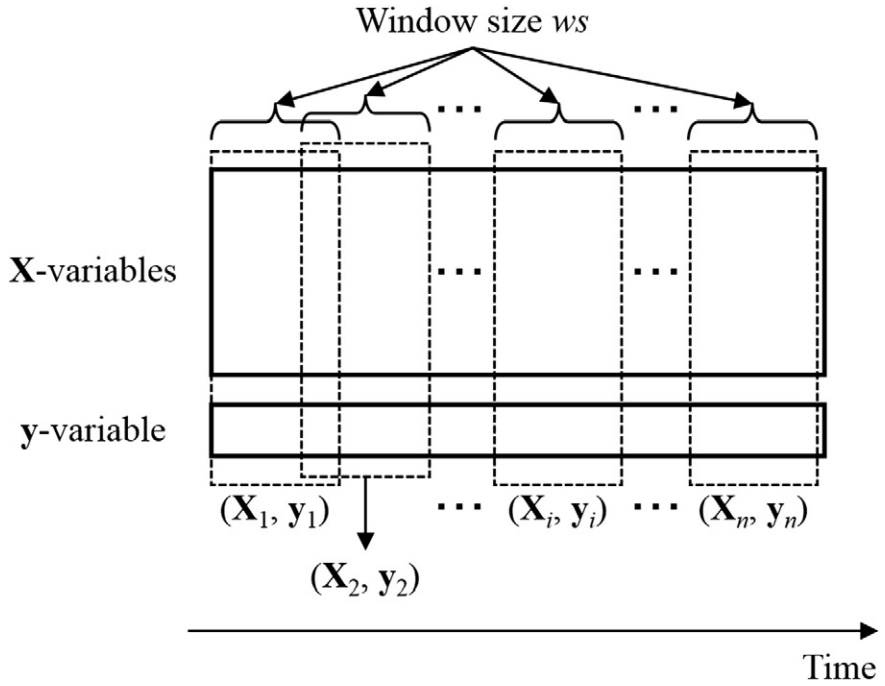

Fig. 1. Basic concept of optimizing the multiple combinations of SVR hyperparameters. Datasets $(\mathbf{X}_1, \mathbf{y}_1)$, $(\mathbf{X}_2, \mathbf{y}_2)$, …, $(\mathbf{X}_i, \mathbf{y}_i)$, …, and $(\mathbf{X}_n, \mathbf{y}_n)$ are used to optimize the SVR hyperparameters $(C_1, \varepsilon_1, \gamma_1)$, $(C_2, \varepsilon_2, \gamma_2)$, …, $(C_i, \varepsilon_i, \gamma_i)$, …, and $(C_n, \varepsilon_n, \gamma_n)$, respectively.