Practical Use of Savitzky-Golay Filtering-Based Ensemble Online SVR

Hiromasa Kaneko*, Takuya Matsumoto**, Shigeki Ootakara***, Kimito Funatsu*

* Department of Chemical System Engineering, The University of Tokyo, Tokyo, Japan (e-mail: hkaneko@chemsys.t.u-tokyo.ac.jp, funatsu@chemsys.t.u-tokyo.ac.jp).
** Process Technology Center, Mitsui Chemicals Inc., Chiba, Japan (e-mail: Takuya.Matsumoto@mitsuichemicals.com)
*** Osaka Works, Mitsui Chemicals, Inc., Osaka, Japan (e-mail: Shigeki.Ootakara@mitsuichemicals.com)

Abstract: As a result of collaboration between Mitsui Chemicals, Inc. and the University of Tokyo, a soft sensor tool was developed and implemented in several plants in Mitsui Chemicals, Inc. A soft sensor is an inferential model constructed between process variables that are easy to measure (X) and process variables that are difficult to measure (y). y-values can be estimated in real time by inputting X-values into a soft sensor. To maintain predictive ability of a soft sensor to be high, we employ ensemble online support vector regression (EOSVR) model as an adaptive soft sensor model, which can adapt to both nonlinear changes and time-varying changes. Additionally, to reduce noise in estimated y-values, Savitzky-Golay (SG) filtering is used for estimated y-values. Our proposed method is called EOSVR-SG and implemented as a soft sensor tool. In this paper, we show our soft sensor tool used in real chemical plants and its execution results in which the EOSVR-SG model could estimate y-values accurately and smoothly.

Keywords: Process control, Soft sensor, Support vector regression, Adaptive model, Ensemble, Noise reduction, Smoothing, Savitzky-Golay filtering.

1. INTRODUCTION

Soft sensors are widely used to predict process variables that are difficult to measure online (Kano and Nakagawa, 2008; Kadlec et al., 2009). An inferential model is constructed between process variables that are easy to measure online, which are called X-variables, and process variables that are difficult to measure online, which are called y-variables. The values of y can then be predicted using that model. Through the use of soft sensors, values of y can be predicted with a high degree of accuracy. Both lab samples and measurements of online analyzers are examples of y-variables.

One of the crucial difficulties of soft sensors is that predictive accuracy drops because of changes in the state of chemical plants. This is called soft sensor model degradation (Kaneko and Funatsu, 2013a). To reduce degradation, the model is reconstructed with the newest data. For instance, a moving window (MW) model (Qin, 1998; Kadlec and Gabrys, 2010) is constructed with the data that were measured most recently, and a just-in-time (JIT) model (Schaal et al., 2002; Cheng and Chiu, 2004) is constructed with data that are more similar to a query than other data. Models such as MW, JIT, and time difference (Kaneko and Funatsu, 2011) models that have adaptive mechanisms are called adaptive models (Kadlec et al., 2011). Each adaptive model has strengths and weaknesses for each type of degradation of soft sensor models (Kaneko and Funatsu, 2013a).

As an MW approach, ensemble online support vector regression (EOSVR) has been proposed recently to adapt to nonlinear relationships between X-variables and an y-variable and to time-varying changes in process states (Kaneko and Funatsu, 2014). Through several case studies including the use of real industrial data set, it was confirmed that EOSVR had higher predictive ability than the other adaptive soft sensors did.

However, operation data include not only important variations caused by process changes but also noise such as measurement noise and sensor noise, which leads to decrease predictive ability of soft sensors in both model construction and prediction. In addition, noisy estimated y-values are inappropriate for process control. For instance, it is difficult to judge the end of transition when estimated y-values include noise. We have to handle noise appropriately for soft sensors to estimate important variations in y-values accurately.

Chemometric methods can handle noise in preprocessing data and in regression analysis. Principal component analysis (PCA) (Wold, 1987) and partial least squares (PLS) regression (Wold et al., 2001) are statistical methods projecting data from an original space of m process variables to a new space of n uncorrelated variables (components) while reducing the dimension (m > n). By using only the first some principal components that include main variations in a data set, we can reduce effect of noise to further analysis. In support vector regression (SVR) (Bishop, 2006), robust models for noise can be constructed by using the ε-insensitive loss function and handling y-errors whose absolute values are lower than ε are set as zero. PCA, PLS and SVR can be...
combined with kernel functions and modified to nonlinear methods (Muller et al., 2001).

However, the above chemometric methods do not consider characteristics of operating data or time-series data (Lütkepohl, 2005). For example, temporally close data have strong relationships and correlations. Kaneko and Funatsu (2015a) proposed smoothing-combined soft sensors for noise reduction and improvement of predictive ability. Before both model construction and prediction in soft sensor analysis, a smoothing method such as a simple moving average, a linearly weighted moving average, an exponentially weighted moving average and Savitzky-Golay (SG) filtering (Savitzky and Golay, 1964) is used for time-series data of each X-variable, which is measured frequently, and noise can be reduced while considering relationships between temporally close data. Case studies using simulated and industrial datasets confirm that the use of the proposed methods enables soft sensors to predict y-values smoothly and accurately and SG filtering had higher performance or lower prediction errors than the other smoothing methods.

In this study, we developed a method combining EORSV and SG filtering, which is called EOSVR-SG and implemented a MATLAB® tool performing the method in real plants in Mitsui Chemicals, Inc. After a y-value of a query is estimated by the EOSVR model, SG filtering is employed to an estimated y-variable. The smoothed y-value is the final estimation result.

As case studies, we show the prediction results of oursoft sensor tool in real plants of Osaka Works, Mitsui Chemicals, Inc. The EOSVR-SG model produces smooth estimated y-values and achieves highly predictive ability.

2. METHOD

2.1 Ensemble Online Support Vector Regression (EOSVR)

The SVR method (Bishop, 2006) applies a support vector machine (SVM) to regression analyses, and can be used to construct nonlinear models by applying a kernel trick along with the SVM. The primal form of the SVR can be given as the following optimization problem.

Minimize

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \max\left(0, |y_i - f(x_i)| - \varepsilon\right).$$

where \(y_i\) and \(x_i\) are training data, \(f\) is the SVR model, \(w\) is a vector of weights of \(X\) to \(y\), \(\varepsilon\) is a threshold, \(N\) is the number of training data, and \(C\) is a penalty factor that controls the trade-off between model complexity and training errors. The second term of Eq. (1) is the \(\varepsilon\)-insensitive loss function, which is given as follows:

$$|y_i - f(x_i)|_\varepsilon = \max\left(0, |y_i - f(x_i)| - \varepsilon\right).$$

By minimizing Eq. (1), we can construct a regression model with a good balance between its generalization capabilities and its ability to adapt to the training data. A y-value estimated by inputting data \(x\) is represented as follows:

$$f(x) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i^*) K(x_i, x) + b,$$

where \(b\) is a constant and \(K\) is a kernel function. The kernel function in our application is a radial basis function given as follows:

$$K(x_i, x) = \exp\left(-\gamma \|x_i - x\|^2\right),$$

where \(\gamma\) is a tuning parameter that controls the width of the kernel function.

Basic concept of EOSVR is shown in Fig. 1. Different SVR models mean different sets of the SVR hyperparameters \(C, \varepsilon, \gamma\). First in EOSVR, to obtain multiple combinations of \(C, \varepsilon, \gamma\) offline for various states in a plant, the window size \(ws\) is set, and then, the SVR hyperparameters are optimized by moving the window. The data in each window are then as follows:

$$(X_1, y_1), (X_2, y_2), ..., (X_n, y_n),$$

where \(X_i\) and \(y_i\) are the \(i^{th}\) data set of the X-variables and that of a y-variable, respectively. When the window is moved by \(h\) data points, the \(ij^{th}\) dataset is from the \(h(i-1)+1\)th data point to the \(h(i)\)th data point. For each dataset, the SVR hyperparameters are optimized based on cross-validation, which can be performed very quickly (Kaneko and Funatsu, 2014).
Duplicate combinations of the SVR parameters are removed, leaving the following $m$ combinations of parameters:

$$(C_i, \varepsilon_i, \gamma_i), (C_2, \varepsilon_2, \gamma_2), \ldots, (C_n, \varepsilon_n, \gamma_n).$$

The value of $\varepsilon$ is relatively unimportant, as SVR models are insensitive to this parameter (Kaneko and Funatsu, 2013b).

Then, for each combination of SVR hyperparameters, an SVR model is constructed and is updated using a new measurement of a $y$-variable and corresponding data of $X$-variables, which is online SVR (Kaneko and Funatsu, 2013b). When a new combination of SVR parameters is optimized and it is not a duplication of the old parameter combinations, a new SVR model is constructed and $m$ is set as $(m+1)$.

In prediction, a query datum $x(t)$ at time $t$ are input into the $m$ SVR models; the models predict $y$-values as follows:

$$y_{p,1}(t) = f_1(x(t))$$
$$y_{p,2}(t) = f_2(x(t)),$$
$$\vdots$$
$$y_{p,m}(t) = f_m(x(t))$$

where $f_1, f_2, \ldots, f_{m-1}$ and $f_m$ are the SVR models. We then combine $y_{p,1}(t), y_{p,2}(t), \ldots, y_{p,m-1}(t)$ and $y_{p,m}(t)$ to obtain a final predicted $y$-value $y_p(t)$ using Bayes’ rule (Khatibi sepehr et al., 2013). When $S$ is the current (unobserved) state in a plant and $M_i$ is the $i$th SVR model, the probability of $M_i$ given $S$, $P(M_i|S)$, is required to combine the prediction results of the $m$ SVR models. Given $P(M_i|S)$, the final predicted $y$-value $y_p(t)$ is obtained as follows:

$$y_p(t) = \sum_{i=1}^{m} P(M_i|S) y_{p,i}(t).$$

Using Bayes’ rule, $P(M_i|S)$ is calculated as follows:

$$P(M_i|S) = \frac{P(S|M_i)P(M_i)}{\sum_{j=1}^{m} P(S|M_j)P(M_j)}.$$

In this study, the prior probability of each $M_i$ is assumed to be equal, that is:

$$P(M_i) = \frac{1}{m}.$$

In practice, by changing this prior probability, we can make the strategy of model combination more flexible. Then, we require $P(S|M_i)$, which represents the predictive ability of the $i$th SVR model in the current state $S$. In this paper, we assume that the predictive ability is inversely proportional to the root-mean-square error for the midpoints between the $k$-nearest-neighbor data points (RMSE$_{\text{midknn}}$) (Kaneko and Funatsu, 2013c). Hence, $P(S|M_i)$ is given as follows:

$$P(S|M_i) = \frac{z_i}{\sum_{j=1}^{n} z_j},$$

where

$$z_i = \frac{1}{\text{RMSE}^2_{\text{midknn},i}}.$$

RMSE$_{\text{midknn},i}$ is the RMSE$_{\text{midknn}}$ of the $i$th SVR model with the latest $w$ data. Then, the final predicted $y$-value $y_p(t)$ can be calculated using Eq. (8), considering the predictive ability of each SVR model in the current state of a plant.

Prediction errors are estimated using the standard deviation (STD) of the multiple predicted $y$-values in Eq. (7) (see Fig. 1). When the predicted $y$-values are close together and the STD is small, the prediction error will also be small. The STD can be used as an index of the prediction errors.

### 2.2 Savitzky-Golay (SG) Filtering

The SG method, which is mainly used in spectral analysis, is a method performing smoothing (noise-processing) and numerical differentiation simultaneously. Noisy spectra can be well smoothed while data characteristics such as peaks in spectra remain. Also in soft sensor analysis in this paper, it could be used as a smoothing method.

It is assumed that a process variable $x$ has a value $x_i$ at time $t_i$ and a measurement time interval is constant $k$ (for example, 1 min). In the SG filtering, $x$ is approximated by a high order expression of time $t$, which leads to a smooth curve of $t$, and the high order expression is differentiated with respect to $t$, which leads to numerical differentiation.

When we calculate the smoothed value and the value of $p$th derivation for $(t_i, x_i)$, an $M_i$ function of $t$ is given as follows:

$$f(t) = \sum_{j=1}^{m} b_j t^j.$$

$b_j$ is estimated by the method of least squares by using $(2N+1)$ points, i.e. $(t_{i-N}, x_{i-N}), (t_{i-N+1}, x_{i-N+1}), (t_{i-N+2}, x_{i-N+2}), \ldots, (t_{i}, x_{i}), (t_{i+N-2}, x_{i+N-2}), (t_{i+N-1}, x_{i+N-1})$ and $(t_{i+N}, x_{i+N})$. In other words, multiple regression analysis is conducted between $x$ as an objective variable and $t$, $t^2$, ..., and $t^4$ as explanatory variables. The smoothed value at $t_i$ is given by inputting $t_i$ into $f(t)$ in eq. (13). In addition, the value of $p$th derivation at $t_i$ is given by inputting $t_i$ into $f(t)$ after differentiating $p$ times. This is repeated for overall $x$. In fact, $b_j$ in eq. (13) is calculated more effectively than the method of least squares, the details of which are shown in the reference (Yoshimura and Takayanagi, 2012).

### 2.3 EOSVR+SG

In spectral analysis, the SG filter is used for each spectrum, i.e. a datum. On the other hand, an SG filter is used not for each datum (a spectrum), but for each process variable in soft sensor analysis. Because process variables have time-series data that are continuous temporally, the SG filter will reduce noise in each process variable as well as in spectra.
In soft sensor analysis, particularly in prediction, not all $2N+1$ points can be used since future values cannot be obtained when the current time is $t_i$. In this case, $(2N+1)$ points: $(t_i-2N, x_i-2N)$, $(t_i-2N+1, x_i-2N+1)$, ..., $(t_i, x_i)$ and $(t_i+1, x_{i+1})$ are used and $b_j$ is estimated by the method of least squares.

An actual $y$-variable is not available at the same frequency as X-variables, but an estimated $y$-variable can be available at the same frequency as X-variables. Therefore an SG filter is used for an estimated $y$-variable, which has plenty of estimated data, for example, an estimated $y$-variable is obtained every minute. Data of an estimated $y$-variable are replaced with data smoothed by SG filters in prediction. $y$-values are estimated by the EOSVR model.

2.4 Overview of Our Developed Soft Sensor Tool

When the soft sensor tool starts, the EOSVR model is automatically constructed using existing data. The flow in prediction using EOSVR+SG is shown in Fig. 2. When a query is input into the soft sensor tool, the EOSVR model estimates a $y$-value of the query and the estimated value is saved. Then recent estimated $y$-values are loaded and SG filtering is conducted. Then, the smoothed $y$-value datum is treated as output of the soft sensor tool.

When new measurement of $y$ is obtained, the EOSVR model is updated with corresponding X-values.

The hyperparameters such as $ws$ and $N$ are loaded through a csv file. Our soft sensor tool is implemented using MATLAB®.

![Diagram of prediction using EOSVR+SG](image)

Fig. 2. The flow in prediction using EOSVR+SG. After the prediction with the EOSVR model, the predicted $y$-value is smoothed with SG filtering.

3. Case Studies

As case studies, we show prediction results of our soft sensor tool in an exhaust gas denitration plant and in a formalin production plant, both of which are operated in Osaka Works, Mitsui Chemicals, Inc. $ws$ was set as 50 and $N$ was determined through trial and error processes. The superiority of the EOSVR model over other models such as MWPPLS, MWSVR, JITPLS and LWPLS was confirmed in the reference (Kaneko and Funatsu, 2014).

3.1 Exhaust Gas Denitration Plant

Our soft sensor was applied to an exhaust gas denitration plant. NH$_3$ is injected into a denitration reactor, where exhaust gas passes through a catalytic layer, and NOx is decomposed into harmless N$_2$ and water vapor. The reactions are as follows:

\[
4\text{NO}+4\text{NH}_3+2\text{O}_2 \rightarrow 4\text{N}_2+6\text{H}_2\text{O}, \quad (14)
\]

\[
\text{NO}+\text{NO}_2+2\text{NH}_3 \rightarrow 2\text{N}_2+3\text{H}_2\text{O}. \quad (15)
\]

Soft sensors can be used to control the remaining NH$_3$ concentration at the outlet of the denitration reactor (denitration outlet NH$_3$), NOx concentration at the outlet of the denitration reactor (denitration outlet NOx), and NOx concentration at the outlet of the outlet gas duct (outlet gas duct NOx).

In this case study, the $y$-variable is outlet gas duct NOx, and the X-variables consist of 23 process variables without time delays such as temperature, pressure, and flow rates at the gas mixer and denitration reactor. Although these concentrations can be measured every minute, soft sensors are required because hardware sensors are damaged, which need maintenance as a result. Dynamics in the process does not have to be considered.

Fig. 3 shows execution results of our soft sensor tool, which are prediction results for outlet gas duct NOx in several hours. Although the EOSVR model could estimate gas duct NOx accurately, estimated $y$-values sometimes had meaningless big spikes that were undesirable for process monitoring and process control because plant operators would misunderstand the spikes as process disturbances and failure of hardware sensors.

In contrast, as is shown in Fig. 3(b), estimated $y$-values had good agreement with measured $y$-values and there were no big spikes in estimated $y$-values. Our soft sensor tool could produce very accurate prediction results in a real exhaust gas denitration plant in Mitsui Chemicals, Ins. successfully. Since the predictive ability of our soft sensor is sufficient, the soft sensor can be used as alternative of the hard sensor when failure of the hardware sensor happens.

3.2 Formalin Production Plant
We applied our soft sensor tool to a formalin production plant. The plant consists of a reactor, absorption tower, ion exchange tower, and demethanol tower. Methanol is a raw material used to produce formaldehyde through the following dehydrogenation reaction:

\[
\text{CH}_3\text{OH} \rightarrow \text{H}_2\text{CO}+\text{H}_2. \tag{14}
\]

Formic acid can be produced by the further oxidation of formaldehyde. To monitor a reaction situation, specific weight and the concentrations of methanol, formaldehyde and formic acid are measured at the bottom of the absorption tower. However, they are measured every several hours with a given delay.

In this case study, the y-variable is concentration of methanol and the X-variables are 28 process variables without time delays such as temperature, pressure, and flow rate in each tower, which are measured online every minute. The y-measurement is much less frequent than the X-measurement and takes time, which is considered in further analysis.

Fig. 3 shows execution results of our soft sensor tool in an exhaust gas denitrification plant in Mitsui Chemicals, Inc. x-axes and y-axes are time and NO\textsubscript{X} concentration at the outlet of the outlet gas duct, respectively. Red lines and blue lines represent measured y-values and estimated y-values, respectively.

Fig. 4 shows execution results of our soft sensor tool in a formalin production plant in Mitsui Chemicals, Inc. x-axes and y-axes are time and concentration of methanol, respectively. Red lines and blue lines represent measured y-values and estimated y-values, respectively.
days. Long horizontal lines in measured y-values mean little measurement frequency of y. As is shown in Fig. 4(a), although EOSVR model may estimate concentration of methanol well, variation in estimated y-values is too much to acquire information on important variation in concentration of methanol from estimated y-values. Noisy estimated y-values cannot be applied to process control even when estimated y-values have good agreement with measured y-values.

On the other hand, EOSVR-SG achieved both accurate estimation results and smooth estimated y-values (see Fig. 4(b)), which is very useful information for plant operators to monitor concentration of methanol. It should be noted that EOSVR+SG could predict important variations in concentration of methanol beforehand and this leads plant operators to achieve effective process control. In addition, since the predictive ability of EOSVR-SG is very high and our soft sensor can be used as alternative of offline analysis of methanol concentration, the measurement frequency of methanol concentration can be reduced more, which relieves the burden of sampling and measurement for plant operators.

4. CONCLUSIONS

We have been developing a soft sensor tool that can construct a soft sensor model and can estimate y-values using the constructed model. Our soft sensor tool includes EOSVR-SG that can adapt nonlinear changes and time-varying changes in process states and can reduce noise in estimated y-values. Through the case studies in real chemical plants in Mitsui Chemicals, Inc., it was confirmed that our soft sensor tool could produce accurate estimation results and smooth estimated y-values, which are very useful for process monitoring and process control.

Because our soft sensor tool is not specialized in specific plants, it can be easily applied to other plants widely. We believe that the use of our soft sensor tool will enable chemical plants to be operated more effectively and stably.

ACKNOWLEDGMENTS

H. Kaneko and F. Funatsu acknowledge the Core Research for Evolutionary Science and Technology (CREST) project “Development of a knowledge-generating platform driven by big data in drug discovery through production processes” of the Japan Science and Technology Agency (JST).

REFERENCES


