Development of soft-sensor using locally weighted PLS with adaptive similarity measure

Sanghong Kim [a], Ryota Okajima [a,1], Manabu Kano [b,*], Shinji Hasebe [a]

[a] *Department of Chemical Engineering, Kyoto University, Kyoto 6158510, Japan*
[b] *Department of Systems Science, Kyoto University, Kyoto 6068501, Japan*

ARTICLE INFO

ABSTRACT

Recently, just-in-time (JIT) modeling, such as locally weighted partial least squares (LW-PLS), has attracted much attention because it can cope with changes in process characteristics as well as nonlinearity. Since JIT modeling derives a local model from past samples similar to a query sample, it is crucial to appropriately define the similarity between samples. In this work, a new similarity measure based on the weighted Euclidean distance is proposed in order to cope with nonlinearity and to enhance estimation accuracy of LW-PLS. The proposed method can adaptively determine the similarity according to the strength of the nonlinearity between each input variable and an output variable around a query sample. The usefulness of the proposed method is demonstrated through numerical examples and a case study of a real cracked gasoline fractionator of an ethylene production process.

1. Introduction

In various industrial processes, it is necessary to measure and control product quality to produce high-quality, competitive products. However, online measurement is not always available due to unacceptable expenses of analytical instruments or long measurement/analysis delay. To solve this problem, inferential models using online measured variables as predictor variables have been adopted in many fields such as chemical, bioprocess, steel, and pharmaceutical [1,2]. According to the recent questionnaire survey of process control in the chemical industry in Japan [3], 90% of the inferential models are constructed by using linear regression methods such as multiple regression analysis (MRA) and partial least squares (PLS). This fact shows that linear models are practically useful. In some cases, nonlinear models are required to achieve high estimation accuracy for processes having strong nonlinearity. Thus, nonlinear modeling methods such as neural networks [4–7], support vector regression [8–10] and polynomial functions [11–13] have been used to construct nonlinear inferential models.

The above-mentioned questionnaire survey revealed that the most important problem of current inferential models is how to cope with changes in process characteristics and keep high estimation accuracy for a long period of time, i.e., model maintenance [3]. The importance of this problem was also pointed out in [1,14]. To cope with changes in process characteristics, many kinds of recursive modeling methods, which update models by prioritizing newer samples, have been

developed [15]. When process characteristics change gradually, the prioritized samples are supposed to be similar to a query sample, for which an output estimation is required. For such a case, recursive methods can cope with gradual changes in process characteristics. However, they cannot cope with an abrupt change in process characteristics caused by replacement of a catalyst, cleaning of equipment, etc., because a query sampled just after an abrupt change becomes significantly different from the prioritized samples.

Locally weighted regression (LWR) [16], which is also called just-in-time learning, lazy learning or model-on-demand, constructs a local model by prioritizing samples in a database according to the similarity between them and a query sample. Hence, LWR can cope with abrupt changes as well as gradual ones in contrast to the recursive methods introduced in [15]. Furthermore, it can cope with nonlinearity since it builds a local model repeatedly. To build an accurate model with LWR, the similarity needs to be properly defined. In general, similarity is defined on the basis of the Euclidean distance or the Mahalanobis distance [10,17–22]. Other similarity measures proposed so far include the angle [14,23,24], the distance between an output estimate for a query sample derived by a global model and output measurements for samples in a database [25], the correlation [26,27] and the weighted Euclidean distance [28–30]. In addition to define the similarity properly, it is crucial to update a database when new data become available in order to cope with changes in process characteristics. More detailed explanation and review of the problem of the changes in process characteristics and LWR can be found in [31].

This study focuses on the problem of nonlinearity and the definition of the similarity, and does not deal with the problem of changes in process characteristics. The similarity based on the weighted Euclidean

* Corresponding author. Tel.: +81 75 753 3367; fax: +81 75 753 3371.
  E-mail address: manabu@human.sys.i.kyoto-u.ac.jp (M. Kano).
[1] Present address: Nippon Steel Corporation, Tokyo 1008071, Japan.

distance is further investigated for its simplicity. PLS is adopted for local modeling since it can cope with collinearity and has been widely accepted in various fields. The main contribution of this paper is to discuss how the weight of each input should be determined and to propose a method for deriving appropriate weights from operation data stored in a database.

The rest of this paper is organized as follows. In Section 2, the algorithm of locally weighted PLS (LW-PLS) is explained. Section 3 discusses how to determine the weight of each input, and a method for deriving the appropriate weights from operation data is proposed. Section 4 shows the effectiveness of the proposed method through numerical examples. In Section 5, an application result of the proposed method to an industrial distillation process is reported. Finally, this research is concluded in Section 6.

2. Locally weighted partial least squares

The nth sample ($n = 1,2,\cdots,N$) of input and output variables is denoted by

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

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

where M is the number of input variables, L is the number of output variables and the superscript T denotes the transpose of a vector or matrix. $\boldsymbol{X} \in \Re^{N \times M}$ and $\boldsymbol{Y} \in \Re^{N \times L}$ are input and output variable matrices whose nth rows are $\boldsymbol{x}_n^\mathrm{T}$ and $\boldsymbol{y}_n^\mathrm{T}$, respectively. N is the number of samples.

LW-PLS is a just-in-time (JIT) modeling method. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are stored in a database in order to construct a local PLS model. 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, and a local PLS model is constructed by weighting samples with a similarity matrix $\boldsymbol{\Omega} \in \Re^{N \times N}$ defined by

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

In general, the output estimate $\hat{\boldsymbol{y}}_q \in \Re^L$ is calculated through the following procedure.

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

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

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

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

$$\overline{x}_m = \sum_{n=1}^{N} \omega_n x_{nm} \Big/ \sum_{n=1}^{N} \omega_n \tag{7}$$

$$\overline{y}_l = \sum_{n=1}^{N} \omega_n y_{nl} \Big/ \sum_{n=1}^{N} \omega_n \tag{8}$$

where $\boldsymbol{1}_N \in \Re^N$ 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 $\boldsymbol{X}_r^\mathrm{T} \boldsymbol{\Omega} \boldsymbol{Y}_r \boldsymbol{Y}_r^\mathrm{T} \boldsymbol{\Omega} \boldsymbol{X}_r$ which corresponds to the maximum eigen value.

5. Derive the rth loading vector of $\boldsymbol{X}$

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

and the regression coefficient vector

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

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

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

7. If $r = R$, calculate the output estimate

$$\hat{y}_q = [\overline{y}_1, \overline{y}_2, \cdots, \overline{y}_L]^\mathrm{T} + \sum_{r=1}^{R} t_{q,r} \boldsymbol{q}_r \tag{13}$$

and finish estimation. Otherwise, set

$$\boldsymbol{X}_{r+1} = \boldsymbol{X}_r - \boldsymbol{t}_r \boldsymbol{p}_r^\mathrm{T} \tag{14}$$

$$\boldsymbol{Y}_{r+1} = \boldsymbol{Y}_r - \boldsymbol{t}_r \boldsymbol{q}_r^\mathrm{T} \tag{15}$$

$$\boldsymbol{x}_{q,r+1} = \boldsymbol{x}_{q,r} - t_{q,r} \boldsymbol{p}_r. \tag{16}$$

8. Set $r = r + 1$ and go to step 4.

When the similarity matrix $\boldsymbol{\Omega}$ is an identity matrix, LW-PLS becomes the same as linear PLS. At step 3, the weighted mean of each variable is subtracted from each column of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{x}_q^\mathrm{T}$ to make the query sample near to the origin of the multidimensional space. At steps 4–8, the latent variable $\boldsymbol{t}$, the loading vector $\boldsymbol{p}$ and the regression coefficient vector $\boldsymbol{q}$ are derived iteratively, and the output estimate $\hat{\boldsymbol{y}}_q$ is calculated when $r = R$.

The definition of the similarity affects the estimation performance of LW-PLS significantly. In the original algorithm of LW-PLS [32], the similarity ω_n is defined on the basis of the Euclidean distance. The similarity based on the Euclidean distance or the Mahalanobis distance is used most frequently to construct a local regression model [17–21,10,32]. In addition, the estimation accuracy can be improved by using the similarity based on the weighted Euclidean distance [28–30]. In these researches, weights were defined on the basis of regression coefficients of a global MRA model or an LW-PLS model where all the weights were one. However, the conventional weighted Euclidean distance may deteriorate the estimation performance as described below. Hence, this research discusses how the weight of each input should be determined in Section 3.1 and proposes a method for deriving appropriate weights from operation data in Section 3.2.

3. New similarity measure

3.1. How should weights be determined?

In the present work, it is assumed that the number of output variables is one, and the following form of the similarity ω is investigated:

$$\omega_n = \exp\left(-\frac{d_n}{\sigma_d \varphi}\right) \tag{17}$$

$$d_n = \sqrt{\left(\boldsymbol{x}_n - \boldsymbol{x}_q\right)^\mathrm{T} \boldsymbol{\Theta} \left(\boldsymbol{x}_n - \boldsymbol{x}_q\right)} \tag{18}$$

$$\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \theta_2, \cdots, \theta_M) \tag{19}$$

where σ_d is a standard deviation of d_n ($n = 1,2,\cdots,N$) and φ is a localization parameter; the similarity decreases steeply when φ is small and gradually when φ is large. In addition, $\boldsymbol{\Theta} \in \mathfrak{R}^{M \times M}$ is a weighting matrix and θ_m is a weight of the mth input variable.

Fig. 1 shows a simple example, in which a relationship between a local linear model and weights $\theta_m (m = 1,2,3)$ is illustrated by using very small number of samples. In each figure, it is assumed that the values and the weights of the other inputs, and φ are constant. Relationship between input 1 and the output variable is linear as shown in Fig. 1 (top). Large θ_1 causes overfitting by prioritizing samples 4 and 5; therefore, θ_1 should be small. On the contrary, the relationship between input 2 and the output variable is nonlinear as shown in Fig. 1 (middle). Thus, θ_2 should be large to cope with nonlinearity between input 2 and the output variable. In addition, the strength of nonlinearity around a query sample may change depending on the
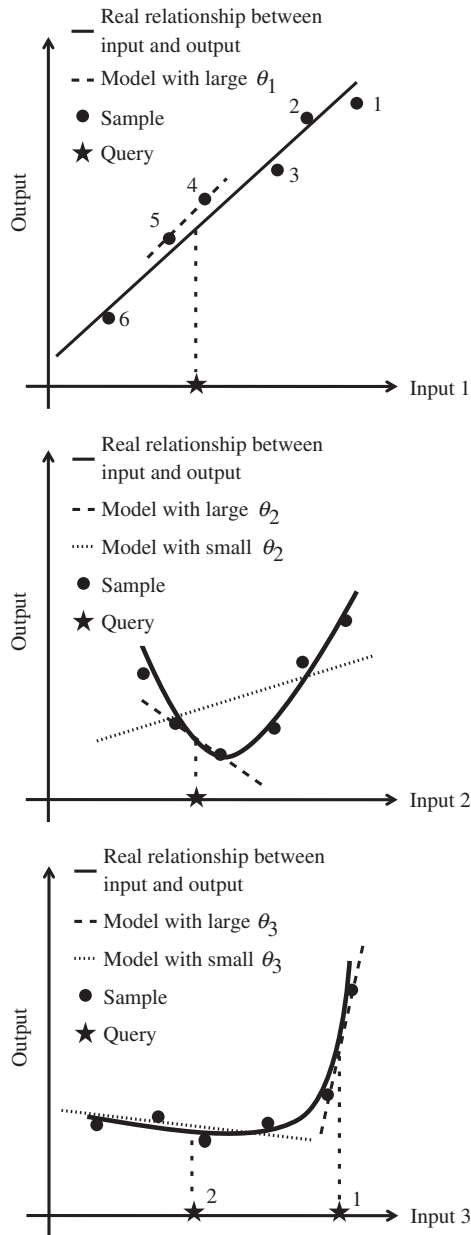
value of the input variable as shown in Fig. 1 (bottom). In this case, θ_3 should be large for query 1 and small for query 2.

The weights proposed in [28–30] do not necessarily correspond to the strength of nonlinearity around a query sample. For example, a regression coefficient of an input can be large even when the input–output relationship is linear. In such a case, the large weight might cause a deterioration of the estimation performance as shown in Fig. 1 (top).

3.2. Proposed procedure for calculating similarity

In Section 3.1, it was revealed that weights of inputs should correspond to strength of nonlinearity between the inputs and an output around a query sample. In addition, a regression coefficient, i.e. slope in Fig. 1, significantly changes around a query sample when the nonlinearity around it is strong. Although the regression coefficient is constant when input–output relationship is linear, it changes depending on the strength of nonlinearity. To evaluate the change of the regression coefficient of an input around a query sample and to determine the weights, the weighted variance of each input's regression coefficients of LW-PLS models is utilized. The similarities between a query sample and samples in a database are utilized as the weights when the weighted variance is calculated. Since similarity depends on the weight θ, iterative calculation is conducted to derive similarity and θ.

Offline and online calculation procedures of the weights are as follows.

· Offline part

1. Determine the number of latent variables R, the localization parameter φ and the maximum iteration number I.
2. Set $i = 1$ and $\theta_{m,i-1} = 1$ for all m.
3. Regard each of N samples in the database as a query sample and construct N LW-PLS models by using $\theta_{m,i-1}$.
4. Calculate the variance $V_{m,i}$ of N regression coefficients of the mth input variable $a_{nm,i}$, then set $\theta_{m,i} = (V_{m,i})^{\alpha}$. Here, α is a tuning parameter.
5. If $i = I$ or the following equation is satisfied for all m, finish the offline calculation. Here, ε_1 is a tolerance.

$$\left| \frac{\theta_{m,i} - \theta_{m,i-1}}{\theta_{m,i-1}} \right| \leq \varepsilon_1 \tag{20}$$

6. Set $i = i + 1$ and go to step 3.

In the offline part, θ_m is first set to 1, then θ_m is updated to $(V_{m,i})^{\alpha}$; a_{nm} and V_m are calculated repeatedly until θ_m converges.

· Online part

1. Determine the maximum iteration number J, and set $j = 1$ and $\theta_{m,j-1} = \theta_m$ obtained in the offline part.
2. Calculate the similarity $\omega_{n,j-1}$ by using $\theta_{m,j-1}$.
3. Calculate the weighted variance $V_{m,j}$ of a_{nm} obtained in the offline part.

$$V_{m,j} = \sum_{n=1}^{N} \omega_{n,j-1} (a_{nm} - \bar{a}_m)^2 \Big/ \sum_{n=1}^{N} \omega_{n,j-1} \tag{21}$$

$$\bar{a}_m = \sum_{n=1}^{N} \omega_{n,j-1} a_{nm} \Big/ \sum_{n=1}^{N} \omega_{n,j-1} \tag{22}$$

4. Set $\theta_{m,j} = [(V_{m,j})^{\alpha} + \theta_{m,j-1}]/2$.

Fig. 1. The relationship between a local model and weights $\theta_m (m = 1,2,3)$. (Top) a case where relationship between an input and an output is linear. (Middle) a case where relationship between an input and an output is nonlinear. (Bottom) a case where the strength of nonlinearity changes depending on the value of an input variable.

5. If $j = J$ or the following equation is satisfied for all m, finish the online calculation. Here, ε_2 is a tolerance.

$$\left| \frac{\theta_{m,j} - \theta_{m,j-1}}{\theta_{m,j-1}} \right| \leq \varepsilon_2 \tag{23}$$

6. Set $j = j + 1$ and go to step 2.

In the online part, θ_m is updated by using the weighted variance V_m of a_{nm} obtained in the offline part and the similarity ω_n between a query sample to evaluate the strength of nonlinearity around a query sample.

This procedure contains seven parameters to be determined: the number of latent variables R, the localization parameter φ, the tuning parameter α, the maximum iteration number in the offline part I and in the online part J, and the tolerance in the offline part ε_1 and in the online part ε_2. R, φ and α can be determined by applying cross validation to all data or by building and validating models with different datasets, i.e., model construction data and parameter tuning data. The proposed method includes the conventional LW-PLS, which uses normal Euclidean distance since the proposed method becomes the same as the conventional one when $\alpha = 0$. Thus, the estimation accuracy of the proposed LW-PLS model is the same as or better than that of the conventional LW-PLS model when α is tuned properly.

4. Numerical example

In this section, the proposed method is compared with the conventional methods in two numerical examples. The following four methods are compared.

LW-PLS 1) LW-PLS with $\theta_m = 1$.
LW-PLS 2) LW-PLS with θ_m defined as the absolute value of the mth variable's regression coefficient of a global MRA model [28].
LW-PLS 3) LW-PLS with θ_m defined as the absolute value of the mth variable's regression coefficient of an LW-PLS model constructed by LW-PLS 1 [29].
LW-PLS 4) LW-PLS with θ_m defined by the proposed method.

4.1. Problem settings

The following two cases are investigated; in each case, x_m and y are inputs and an output, respectively.

· Case 1

$$w_m \sim N\left(0, 0.02^2\right) (m = 0, 1, 2, 3) \tag{24}$$

$$s_m \sim \mathrm{rand}(-5, 5) (m = 1, 2, 3) \tag{25}$$

$$x_m = s_m + w_m (m = 1, 2, 3) \tag{26}$$

$$y = 10s_1 + 5s_2^2 + \exp(s_3) + w_0 \tag{27}$$

· Case 2

$$w_m \sim N\left(0, 0.02^2\right) (m = 0, 1, \cdots, 6) \tag{28}$$

$$s_m \sim \mathrm{rand}(-5, 5) (m = 1, 2, \cdots, 6) \tag{29}$$

$$x_m = s_m + w_m (m = 1, 2, \cdots, 6) \tag{30}$$

$$x_7 = s_6 + w_6 \tag{31}$$

$$y = s_2^3 + 3s_3 + s_4^2 + \exp(s_5) + 3s_6 + w_0 \tag{32}$$

Here, rand(a, b) denotes the uniform random distribution in closed interval [a, b], and $N(\mu, \sigma^2)$ denotes the normal distribution whose mean is μ and standard deviation is σ. In both cases, 3000 samples were generated and divided into three groups: samples for model construction (1000 samples), parameter tuning (1000 samples) and model validation (1000 samples). Models were constructed with different values of localization parameter φ, the number of latent variables R, and α, by using samples for model construction. Then, the estimation errors were calculated by using samples for parameter tuning, and the set of parameters that minimized the estimation error was selected. The search range of φ, R and α is [0.01,0.03, $\cdots$,0.09], [1,2,3] and [0.01,0.03,$\cdots$,0.09], respectively. The appropriate search range of the parameters depends on the situation; therefore, it is recommended to make the search range wide enough in order to get the optimal parameters. In LW-PLS 4, tolerances ε_1 and ε_2 are 0.01. Both of the maximum iteration numbers I and J are 30.

4.2. Results and discussions

Table 1 shows the selected parameters and root mean square error for validation samples (RMSE 1). The proposed method achieved the minimum RMSE 1 in both cases and was considerably superior to the conventional methods. Fig. 2 shows the relationship between RMSE for parameter tuning samples (RMSE 2) and φ when the proposed method is applied to case 1 ($R = 3$). RMSE 2 was large when φ was too small or too large. Overfitting occurred when φ was too small, and models were unable to cope with nonlinearity between input and output variables when φ was too large. Tables 2 and 3 show θ_m when $\boldsymbol{x}_q = [0,0,3]^T$ and $\boldsymbol{x}_q = [0,0,-3]^T$ in case 1, respectively. Fig. 3 shows the transition of θ_m in the online part of the weights calculation procedure. Here, θ_m is normalized so that the sum of θ_m be 1 in LW-PLS 2, 3 and 4. In case 1, where the relationship between x_1 and the output is linear, therefore, θ_1 should be 0. When $\boldsymbol{x}_q = [0,0,3]^T$, θ_3 should be larger than θ_2 because x_3 has stronger nonlinearity around the query sample than x_2, i.e.

$$\mathrm{abs}\left(\frac{\partial^2 y}{\partial x_2^2}\bigg|_{\boldsymbol{x}=[0,0,3]^T} \right) < \mathrm{abs}\left(\frac{\partial^2 y}{\partial x_3^2}\bigg|_{\boldsymbol{x}=[0,0,3]^T} \right) \tag{33}$$

where abs(a) denotes the absolute value of a. On the other hand, when $\boldsymbol{x}_q = [0,0,-3]^T$, θ_2 should be larger than θ_3 because

$$\mathrm{abs}\left(\frac{\partial^2 y}{\partial x_2^2}\bigg|_{\boldsymbol{x}=[0,0,-3]^T} \right) > \mathrm{abs}\left(\frac{\partial^2 y}{\partial x_3^2}\bigg|_{\boldsymbol{x}=[0,0,-3]^T} \right) \tag{34}$$

The proposed method derived appropriate θ for both query samples while the other methods could not. This is the reason why the proposed method could achieve the best performance in the four methods.

Table 1
Selected parameters and RMSE for validation samples in numerical examples.

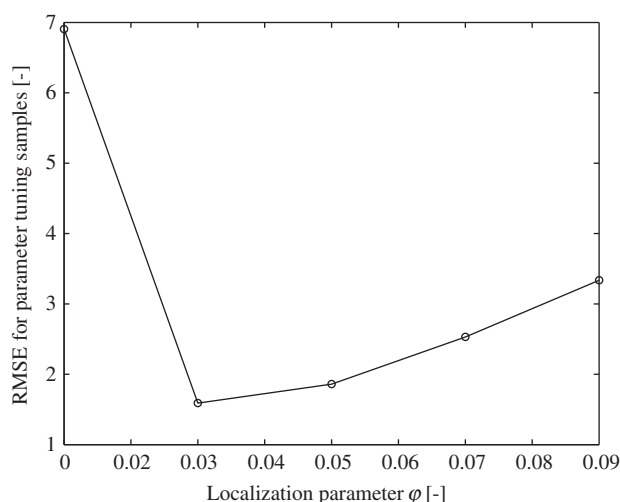| Case	Method	R	φ	α	RMSE
1	LW-PLS 1	3	0.05	–	3.84
	LW-PLS 2	3	0.05	–	5.29
	LW-PLS 3	3	0.05	–	4.69
	LW-PLS 4	3	0.03	0.8	1.59
2	LW-PLS 1	6	0.21	–	18.93
	LW-PLS 2	6	0.09	–	16.53
	LW-PLS 3	5	0.21	–	21.17
	LW-PLS 4	6	0.06	0.8	5.31

Fig. 2. The relationship between RMSE for parameter tuning samples and the localization parameter φ when the proposed method is applied in case 1 ($R = 3$).

5. Application to an industrial distillation process

In this section, an application result of the proposed method to an industrial distillation process is reported. A soft-sensor for estimating the aroma concentration was constructed in order to realize highly efficient operation of a cracked gasoline (CGL) fractionator of an ethylene production process at the Showa Denko K.K. (SDK) Oita plant in Japan. Aroma denotes the generic name for benzene, toluene, xylene and styrene, etc. In this case study, linear PLS, LW-PLS 1, 2, 3 and 4 were compared. The search range of φ, R and α is [0.2,0.4,···,2, 2.5,3.0,···,10], [1,2,···,9], [0.2,0.4,···,2.0], respectively.

5.1. CGL fractionator

A schematic diagram of the CGL fractionator of the ethylene production process is shown in Fig. 4. The CGL fractionator is controlled with multivariable model predictive control (MPC) with an optimizer, and the aroma concentration in the CGL is used as one of the constraints in the optimizer. Although the operation data of the CGL fractionator are stored in the database every hour, the aroma concentration is analyzed in a laboratory usually once a day because of its long analysis time. For safety, the process was operated at a condition that has a wide margin and is far from the constraints. Thus, real-time accurate estimation of the aroma concentration is crucial in order to make the operating condition closer to the constraint and reduce the energy consumption.

5.2. Operation data

Although 19 variables are measured in the CGL fractionator, only eight variables were selected as the input variables of the soft-sensor on the basis of the process knowledge. In addition, the coil outlet temperature of a cracking furnace measured four hours before was used together with the selected input variables, since the product composition is affected by the operating condition of the cracking furnace

Table 2
Derived weights of $\boldsymbol{x}_q = [0,0,3]^{\mathrm{T}}$ in case 1.

Method	θ_1	θ_2	θ_3
LW-PLS 1	1.00	1.00	1.00
LW-PLS 2	0.53	0.08	0.39
LW-PLS 3	0.32	0.02	0.66
LW-PLS 4	0.00	0.32	0.68

Table 3
Derived weights of $\boldsymbol{x}_q = [0,0,-3]^{\mathrm{T}}$ in case 1.

Method	θ_1	θ_2	θ_3
LW-PLS 1	1.00	1.00	1.00
LW-PLS 2	0.53	0.08	0.39
LW-PLS 3	0.57	0.34	0.09
LW-PLS 4	0.01	0.81	0.18

which is located in the upstream of the CGL fractionator, and it takes about four hours for materials to reach the CGL fractionator from the cracking furnace. Hence, the total number of input variables is nine. The selected input variables of the soft-sensor are listed in Table 4 and Fig. 4. The operation data obtained from January 1, 2010 to August 4, 2011 were stored in the database. Then, the tuning parameters were determined using these data, and the aroma concentration was estimated for the operation data obtained from August 6, 2011 to December 31, 2011. Here, all variables were mean-centered and scaled in order to make each variable's standard deviation one.

5.3. Results and discussions

Table 5 shows the selected parameters and RMSE 1. In LW-PLS 4, tolerances ε_1 and ε_2 are 0.01. The maximum iteration numbers I and J are 20 and 30, respectively. The average calculation time of output estimation for each query was 4.8 ms when Intel® Core™ i7-2620M (2.7 GHz × 2) and 8 GB RAM were used.

In this process, the output variable (aroma concentration) is measured to one place of decimal, thus, the differences of RMSEs between Linear PLS and LW-PLS 1, and between LW-PLS 2, 3 and 4 are not significant. The reason why LW-PLS 2, 3 and 4 derived the better result than the other methods might be that the strength of nonlinear effect of each input on the output is different. Table 6 shows the maximum, mean and minimum values, and standard deviation of the mth weight θ_m when LW-PLS 4 is applied. Here, $\theta_1,\theta_2,···,\theta_8$ for each query are normalized so that their sum becomes 1. θ_1 is the largest and the nonlinear effect of input 1 on the output is expected to be strong. In addition, the strength of nonlinear effect of each input on the output does not seem to depend on the value since the standard deviations of the weights are small. This could be the reason why RMSEs of LW-PLS 2, 3 and 4 are similar.
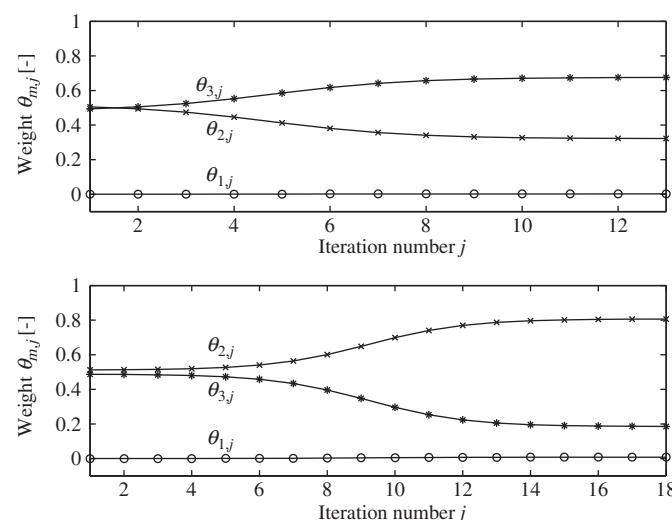

Fig. 3. The transition of θ_m in the online part of the weights calculation procedure in case 1. (Top) $\boldsymbol{x}_q = [0,0,3]^{\mathrm{T}}$. (Bottom) $\boldsymbol{x}_q = [0,0,-3]^{\mathrm{T}}$.
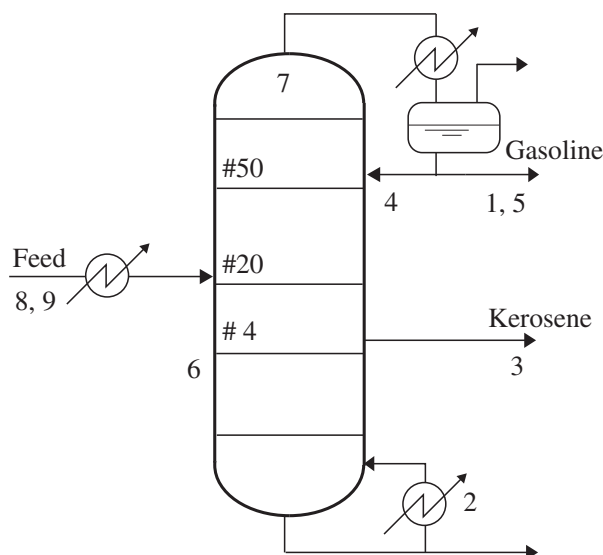
Fig. 4. Schematic diagram of the CGL fractionator of the ethylene production process at the Showa Denko K.K. (SDK) Oita plant.

Table 4
Input variables of the soft-sensor for the CGL fractionator.

No.	Variable name
1	Outlet cracked gasoline density
2	Reboiler flow rate
3	Outlet cracked kerosene flow rate
4	Reflux volume
5	Outlet cracked gasoline flow rate
6	Tray #4 differential pressure
7	Tower top temperature
8	Feed flow rate
9	Cracked furnace coil outlet temperature

6. Conclusion

To construct highly accurate locally weighted partial least squares (LW-PLS) models, an adaptive similarity measure was proposed. In the proposed method, weights of input variables are determined through iterative calculation by using the weighted variance of the regression coefficients. The results of the case studies showed that the proposed method could adaptively derive the appropriate weights and more accurate models than the conventional methods in numerical examples. Furthermore, root mean square error was improved by 11.3% by using the proposed method compared to LW-PLS in which conventional similarity based on the Euclidean distance without weights is used. These results clearly demonstrate the usefulness of the proposed method, which uses newly defined similarity based on the weighted Euclidean distance.

Table 5
Selected parameters and RMSE for validation samples in a case study of the CGL fractionator.

Method	R	φ	α	RMSE
Linear PLS	2	–	–	1.20
LW-PLS 1	2	6.5	–	1.15
LW-PLS 2	2	1.0	–	0.99
LW-PLS 3	2	1.0	–	0.98
LW-PLS 4	2	1.4	1.2	1.03

Table 6
Changes of weights in a case study of the CGL fractionator when LW-PLS 4 is applied.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Maximum value	0.32	0.12	0.13	0.09	0.12	0.10	0.09	0.10	0.12
Mean value	0.26	0.09	0.11	0.08	0.09	0.09	0.09	0.08	0.11
Minimum value	0.21	0.07	0.10	0.07	0.08	0.08	0.08	0.08	0.09
Standard deviation	0.03	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.01

References

[1] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, Computers and Chemical Engineering 33 (2009) 795–814.
[2] T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceutics: a tutorial review, International Journal of Pharmaceutics 417 (2011) 280–290.
[3] M. Kano, M. Ogawa, The state of the art in chemical process control in Japan: good practice and questionnaire survey, Journal of Process Control 20 (2010) 969–982.
[4] G. Andersson, P. Kaufmann, L. Renberg, Non-linear modelling with a coupled neural network - PLS regression system, Journal of Chemometrics 10 (1996) 605–614.
[5] G. Baffi, E.B. Martin, A.J. Morris, Non-linear projection to latent structures revisited (the neural network PLS algorithm), Computers and Chemical Engineering 23 (1999) 1293–1307.
[6] S.J. Qin, T.J. McAvovy, Nonlinear PLS modeling using neural networks, Computers and Chemical Engineering 16 (1992) 379–391.
[7] D. Pérez-Marin, A. Garrido-Varo, J.E. Guerrero, J.C. Gutiérrez-Estrada, Use of artifcial neural networks in near-infrared reflectance spectroscopy calibrations for predicting the inclusion percentages of wheat and sunflower meal in compound feedingstuffs, Applied Spectroscopy 60 (2006) 1062–1069.
[8] V.R. Nadadoor, H. Siegler, S.L. Shah, W.C. McCaffrey, A. Ben-Zvi, Online sensor for monitoring a microalgal bioreactor system using support vector regression vector regression, Chemometrics and Intelligent Laboratory Systems 44 (2012) 2101–2105.
[9] I. Barman, C.R. Kong, N.C. Dingari, R.R. Dasari, M.S. Feld, Development of robust calibration models using support vector machines for spectroscopic monitoring of blood glucose, Chemometrics and Intelligent Laboratory Systems 82 (2012) 9719–9726.
[10] D.E. Lee, J.H. Song, S.O. Song, E.S. Yoon, Weighted support vector machine for quality estimation in the polymerization process, Industrial and Engineering Chemistry Research 44 (2005) 2101–2105.
[11] S. Wold, N.K. Wold, B. Skagerberg, Nonlinear PLS modeling, Chemometrics and Intelligent Laboratory Systems 7 (1989) 53–65.
[12] G. Robertsson, Contributions to the problem of approximation of non-linear data with linear PLS in an absorption spectroscopic context, Chemometrics and Intelligent Laboratory Systems 47 (1999) 99–106.
[13] A.I. Abdel-Rahmana, G.J. Lim, A nonlinear partial least squares algorithm using quadratic fuzzy inference system, Journal of Chemometrics 23 (2009) 530–537.
[14] S.Y. Chang, E.H. Baughman, B.C. McIntosh, Implementation of locally weighted regression to maintain calibrations on FT-NIR analyzers for industrial processes, Applied Spectroscopy 55 (2001) 1199–1206.
[15] P. Kadlec, R. Grbić, B. Gabrys, Review of adaptation mechanisms for data-driven soft sensors, Computers and Chemical Engineering 35 (2011) 1–24.
[16] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, Journal of the American Statistical Association 74 (1979) 829–836.
[17] A.T. Walden, P. Prescott, Identification of trends in annual maximum sea levels using robust locally weighted regression, Estuarine, Coastal and Shelf Science 16 (1983) 17–26.
[18] W.S. Cleveland, S.J. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, Journal of the American Statistical Association 83 (1988) 596–610.
[19] T. Naes, T. Isaksson, Locally weighted regression and scatter correction for near-infrared reflectance data, Analytical Chemistry 62 (1990) 664–673.
[20] V. Centner, D.L. Massart, Optimization in locally weighted regression, Analytical Chemistry 70 (1998) 4206–4211.
[21] H.H. Leung, Y.S. Huang, C.X. Cao, Locally weighted regression for desulphurisation intelligent decision system modeling, Simulation Modelling Practice and Theory 12 (2004) 413–423.
[22] Z. Ge, Z. Song, A comparative study of just-in-time-learning based methods for online soft sensor modeling, Chemometrics and Intelligent Laboratory Systems 104 (2010) 306–317.
[23] C. Cheng, M.S. Chiu, A new data-based methodology for nonlinear process modeling, Chemical Engineering Science 59 (2004) 2801–2810.
[24] Z. Ge, Z. Song, Online monitoring of nonlinear multiple mode processes based on adaptive local model approach, Control Engineering Practice 16 (2008) 1427–1437.
[25] Z.Y. Wang, T. Isaksson, B.R. Kowalski, New approach for distance measurement in locally weighted regression, Analytical Chemistry 66 (1994) 249–260.

[26] K. Fujiwara, M. Kano, S. Hasebe, A. Takinami, Soft-sensor development using correlation-based just-in-time modeling, AICHE Journal 55 (2009) 1754–1765.

[27] K. Fujiwara, M. Kano, S. Hasebe, Development of correlation-based clustering method and its application to software sensing, Chemometrics and Intelligent Laboratory Systems 101 (2010) 130–138.

[28] H. Shigemori, M. Kano, S. Hasebe, Optimum quality design system for steel products through locally weighted regression model, Journal of Process Control 21 (2011) 293–301.

[29] S. Kim, M. Kano, H. Nakagawa, S. Hasebe, Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection, International Journal of Pharmaceutics 421 (2011) 269–274.

[30] H. Nakagawa, T. Tajima, M. Kano, S. Kim, S. Hasebe, T. Suzuki, H. Nakagami, Evaluation of infrared-reflection absorption spectroscopy measurement and locally weighted partial least-squares for rapid analysis of residual drug substances in cleaning processes, Analytical Chemistry 84 (2012) 3820–3826.

[31] M. Kano, K. Fujiwara, Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications, Journal of Chemical Engineering Japan (2012), http://dx.doi.org/10.1252/jcej.12we167.

[32] S. Schaal, C.G. Atkeson, S. Vijayakumar, Scalable techniques from nonparametric statistics for real time robot learning, Applied Intelligence 17 (2002) 49–60.