Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-7 (2016) 502–507

Sparse Sample Regression Based Just-In-Time Modeling (SSR-JIT): Beyond Locally Weighted Approach ⋆

Taku Uchimaru * **Manabu Kano** *

* *Department of Systems Science, Kyoto University, Kyoto, Japan (e-mail: manabu@human.sys.i.kyoto-u.ac.jp).*

Abstract: In the present work, a new method for just-in-time (JIT) modeling is proposed. To develop virtual sensors or soft-sensors that can cope with changes in process characteristics as well as nonlinearity, JIT modeling such as locally weighted regression (LWR) and locally weighted partial least squares (LW-PLS) has been investigated and successfully used in various industries. The conventional JIT modeling methods predict output variables by constructing a local model by using past samples located in the neighborhood around the new target sample (query) every time when the output prediction is required; the modeling samples are selected or weighted according to the similarity between the samples and the query. The similarity is usually determined on the basis of the distance from the query. However, the use of distance does not assure the high prediction accuracy. To overcome this limitation of the conventional JIT methods, the proposed method selects past samples that are useful for constructing an accurate local model by using elastic net, which builds a sparse regression model to estimate the query, and uses the derived regression coefficients to evaluate the similarity for conducting LW-PLS. This sparse sample regression based just-in-time modeling (SSR-JIT) has a potential for surpassing the conventional distance-based JIT modeling. In fact, it was demonstrated that SSR-JIT outperformed LW-PLS in the prediction accuracy through two case studies with real industrial data.

Keywords: Virtual sensing, soft-sensor, just-in-time modelling, locally weighted partial least squares regression, regularised regression, elastic net, process analytical technology, near-infrared (NIR) spectroscopy, pharmaceutical processes.

1. INTRODUCTION

Virtual sensing technology, or soft-sensors, has played an important role in various industries to predict and monitor difficult-to-measure product quality or other important variables in real time by using easy-to-measure variables.

Kano and Ogawa (2010) reported a questionnaire survey of process control applications. The questionnaire asked engineers what are problems related with applications of soft-sensors, and the result confirmed that the model maintenance was the most critical issue. In fact, a total of 36% of the engineers pointed out the necessity to cope with changes in process characteristics and operating conditions in order to keep the prediction performance of soft-sensors. It is also very important to lessen the burdens of modeling including data acquisition and preprocessing.

To cope with changes in process characteristics and operating conditions, recursive methods have been widely accepted (Kadlec et al., 2011). The basic idea is updating a statistical model automatically when new data become available. However, the model may adapt excessively and not function in a sufficiently wide range of operating condition when a process is operated within a narrow range for a certain period of time. In addition, the recursive methods cannot cope with abrupt changes in process characteristics, which are common in industry.

To overcome the limitations of recursive methods, just-in-time (JIT) modeling builds a local model from past data around a query, i.e. a new target sample, only when prediction is required. JIT modeling has the following features: 1) when new input and output data become available, they are stored into a database; 2) only when prediction is required, a local model is constructed from samples located in the neighborhood around the query, and output variables are predicted; 3) the constructed local model is discarded after its use for prediction (Kano and Fujiwara, 2013). There are various JIT modeling methods such as local linear regression (LLR) (Gupta et al., 2008), locally weighted regression (LWR) (Cleveland, 1979; Shigemori et al., 2011), and locally weighted partial least squares (LW-PLS) (Kim et al., 2011). Among them, LW-PLS has been successfully applied to various industrial processes in the recent years. In the pharmaceutical industry, for example, LW-PLS has been applied to prediction of active pharmaceutical ingredients (API) content with near infrared (NIR) spectroscopy (Kim et al., 2011), prediction of the amount of residual drug substances in cleaning processes with infrared-reflection absorption spectroscopy (IR-RAS) (Nakagawa et al., 2012a), and NIR-based real-time monitoring of ingredient concentration during blend-

⋆ This study was supported by JSPS KAKENHI 15K06554.

10.1016/j.ifacol.2016.07.392

ing (Nakagawa et al., 2014). Other applications of LW-PLS include inferential control of product quality in the petrochemical industry (Kim et al., 2013a), maize hardness characterization in the food industry (Guelpa et al., 2015), virtual metrology (VM) in the semiconductor industry (Hirai and Kano, 2015), and determination of clinical parameters in human serum samples with Fourier transform infrared (FTIR) spectroscopy (Perez-Guaita et al., 2015).

LWR and LW-PLS construct a local model by prioritizing samples in a database according to the similarity between a query and the samples. To build an accurate local model, the similarity needs to be properly defined. In general, the similarity is defined on the basis of the Euclidean distance or the Mahalanobis distance. However, the prediction performance of distance-based JIT modeling is not always high because samples used for local modeling are selected on the basis of the distance from the query and the correlation among variables is not taken into account. A good model cannot be developed when the correlation among input and output variables is weak even when the distance between samples is small. On the contrary, a very accurate model can be developed when the correlation is strong even if the distance is large. On the basis of this idea, Fujiwara et al. (2009) proposed correlation-based JIT (CoJIT) modeling. CoJIT can achieve the high prediction performance, but the performance strongly depends on the tuning parameters. Thus, the practicability of CoJIT seems limited.

In the present work, a new method, referred to as sparse sample regression based just-in-time modeling (SSR-JIT), is proposed. To overcome the limitation of the distance-based JIT modeling, SSR-JIT selects past samples that are useful for constructing an accurate local model by using elastic net, which builds a sparse regression model to predict a query, and uses the derived regression coefficients as weights of samples for conducting LW-PLS. The performance of SSR-JIT is evaluated through two case studies with real industrial data.

2. CONVENTIONAL METHODS

In this section, PLS, LW-PLS, and elastic net are briefly explained. These methods are used in SSR-JIT.

2.1 Partial Least Squares (PLS)

In general, PLS is preferable to ordinary least squares (OLS) when a linear regression model is built from process data, because PLS can deal with collinearity that prevents from constructing a robust model by using OLS. PLS derives latent variables as linear combinations of input variables and uses them to predict output variables. Each latent variable is determined so that the covariance between the latent variable and the output variable is maximized.

Given data of input variables and an output variable, which are mean-centered and properly scaled, as $\boldsymbol{X} \in \Re^{N\times M}$ and $\boldsymbol{y} \in \Re^{N}$, a PLS model with K latent variables is expressed as follows:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E} \tag{1}$$

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{q} + \boldsymbol{f} \tag{2}$$

where $\boldsymbol{T} \in \Re^{N\times K}$ consisting of latent variables $\boldsymbol{t}_k \in \Re^{N}(k = 1, 2, \ldots, K)$ is a score matrix, $\boldsymbol{P} \in \Re^{M\times K}$ consisting of $\boldsymbol{p}_k \in \Re^{M}$ is a loading matrix of $\boldsymbol{X}$, $\boldsymbol{q} \in \Re^{K}$ is a regression coefficient vector from latent variables to the output variable, and $\boldsymbol{E}$ and $\boldsymbol{f}$ are residuals.

In PLS, the model is constructed in an iterative manner through the NIPALS algorithm (Wold et al., 2001). After $\boldsymbol{X}_1 = \boldsymbol{X}$ and $\boldsymbol{y}_1 = \boldsymbol{y}$ are set, the variable matrices at the kth iteration ($k \geq 2$) are written as

$$\boldsymbol{X}_k = \boldsymbol{X}_{k-1} - \boldsymbol{t}_{k-1}\boldsymbol{p}_{k-1}^{\mathrm{T}} \tag{3}$$

$$\boldsymbol{y}_k = \boldsymbol{y}_{k-1} - \boldsymbol{t}_{k-1}q_{k-1}. \tag{4}$$

The kth latent variable $\boldsymbol{t}_k$ is expressed as

$$\boldsymbol{t}_k = \boldsymbol{X}_k\boldsymbol{w}_k \tag{5}$$

where the kth weighting vector $\boldsymbol{w}_k$, which is the kth column of the weighting matrix $\boldsymbol{W}$, is determined so that the inner product between $\boldsymbol{t}_k$ and $\boldsymbol{y}_k$ is maximized under the constraint $\|\boldsymbol{w}_k\| = 1$. $\boldsymbol{w}_k$, $\boldsymbol{p}_k$, and q_k are derived as follows:

$$\boldsymbol{w}_k = \frac{\boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{y}_k}{\|\boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{y}_k\|} \tag{6}$$

$$\boldsymbol{p}_k = \frac{\boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{t}_k}{\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k} \tag{7}$$

$$q_k = \frac{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{t}_k}{\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k}. \tag{8}$$

This procedure is repeated until k reaches the number of adopted latent variables K. This PLS algorithm is known as PLS1 because the number of output variables is one; PLS2 is available when two or more output variables need to be predicted simultaneously.

2.2 Locally Weighted Partial Least Squares (LW-PLS)

LW-PLS, which is an extension of LWR, uses PLS to construct a local regression model while both LW-PLS and LWR prioritize past samples in a database according to the similarity between a query and the samples (Kim et al., 2011). It has attracted much attention as a tool for virtual sensing since it can cope with not only changes in process characteristics and operating conditions but also collinearity.

Here the algorithm of LW-PLS is explained. x_{nm} and y_{nl} ($n = 1, 2, \ldots, N; m = 1, 2, \ldots, M; l = 1, 2, \ldots, L$) are preprocessed measurements of input and output variables, where M and L are the numbers of input and output variables, respectively. As the preprocess, an adequate scaling is necessary to achieve high prediction performance. The same preprocess should be applied both to samples in the database and to the query. The nth sample is expressed as

$$\boldsymbol{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]^{\mathrm{T}} \tag{9}$$

$$\boldsymbol{y}_n = [y_{n1}, y_{n2}, \ldots, y_{nL}]^{\mathrm{T}}. \tag{10}$$

The input and output variable matrices $\boldsymbol{X} \in \Re^{N\times M}$ and $\boldsymbol{Y} \in \Re^{N}$ consist of these vectors. In LW-PLS, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are stored in a database, and the similarity index ω_n between a query $\boldsymbol{x}_q$ and the nth sample $\boldsymbol{x}_n$ is calculated to construct a local PLS model when output prediction is required for $\boldsymbol{x}_q$. The predicted output $\hat{\boldsymbol{y}}_q$ is derived through the following procedure.

(1) Determine the number of latent variables K and set $k = 1$.
(2) Determine a similarity matrix $\boldsymbol{\Omega}$.

$$\boldsymbol{\Omega} = \mathrm{diag}\{\omega_1, \omega_2, \ldots, \omega_N\} \tag{11}$$

where diag{} denotes a diagonal matrix and ω_n is the similarity index for the nth sample. The details are mentioned below.
(3) Calculate $\boldsymbol{X}_k$, $\boldsymbol{Y}_k$, and $\boldsymbol{x}_{q,k}$.

$$\boldsymbol{X}_k = \boldsymbol{X} - \boldsymbol{1}_N[\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M] \tag{12}$$

$$\boldsymbol{Y}_k = \boldsymbol{Y} - \boldsymbol{1}_N[\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_L] \tag{13}$$

$$\boldsymbol{x}_{q,k} = \boldsymbol{x}_q - [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M]^{\mathrm{T}} \tag{14}$$

$$\bar{x}_m = \frac{\sum_{n=1}^{N} \omega_n x_{nm}}{\sum_{n=1}^{N} \omega_n} \tag{15}$$

$$\bar{y}_l = \frac{\sum_{n=1}^{N} \omega_n y_{nl}}{\sum_{n=1}^{N} \omega_n} \tag{16}$$

where $\boldsymbol{1}_N \in \Re^N$ is a vector of ones. The weighted means $\bar{x}_m$ and $\bar{y}_l$ are derived by using the similarity index ω_n.
(4) Set $\hat{\boldsymbol{y}}_q = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_L]^{\mathrm{T}}$.
(5) Derive the kth latent variable of $\boldsymbol{X}$.

$$\boldsymbol{t}_k = \boldsymbol{X}_k \boldsymbol{w}_k \tag{17}$$

where $\boldsymbol{w}_k$ is the eigenvector of $\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{Y}_k \boldsymbol{Y}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{X}_k$, which corresponds to the maximum eigenvalue, and it is derived from

$$\boldsymbol{w}_k = \frac{\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{Y}_k}{\| \boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{Y}_k \|} \ . \tag{18}$$

(6) Derive the kth loading vector of $\boldsymbol{X}$ and the kth regression coefficient vector.

$$\boldsymbol{p}_k = \frac{\boldsymbol{X}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{t}_k}{\boldsymbol{t}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{t}_k} \tag{19}$$

$$\boldsymbol{q}_k = \frac{\boldsymbol{Y}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{t}_k}{\boldsymbol{t}_k^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{t}_k} \ . \tag{20}$$

(7) Derive the kth latent variable of $\boldsymbol{x}_q$.

$$t_{q,k} = \boldsymbol{x}_{q,k}^{\mathrm{T}} \boldsymbol{w}_k \ . \tag{21}$$

(8) Replace $\hat{\boldsymbol{y}}_q$ with $\hat{\boldsymbol{y}}_q + t_{q,k} \boldsymbol{q}_k$.
(9) If $k = K$, end. Otherwise, set

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k - \boldsymbol{t}_k \boldsymbol{p}_k^{\mathrm{T}} \tag{22}$$

$$\boldsymbol{Y}_{k+1} = \boldsymbol{Y}_k - \boldsymbol{t}_k \boldsymbol{q}_k^{\mathrm{T}} \tag{23}$$

$$\boldsymbol{x}_{q,k+1} = \boldsymbol{x}_{q,k} - t_{q,k} \boldsymbol{p}_k \ . \tag{24}$$

(10) Set $k = k + 1$ and go to step 5.

The definition of similarity plays a crucial role in improving the prediction accuracy of JIT modeling technique including LW-PLS. The similarity index ω_n is defined as

$$\omega_n = \exp\left(-\frac{\varphi d_n}{\sigma_d}\right) \tag{25}$$

$$d_n = \sqrt{(\boldsymbol{x}_n - \boldsymbol{x}_q)^{\mathrm{T}} \boldsymbol{\Theta} (\boldsymbol{x}_n - \boldsymbol{x}_q)} \tag{26}$$

$$\boldsymbol{\Theta} = \mathrm{diag}\{\theta_1, \theta_2, \ldots, \theta_M\} \tag{27}$$

where φ is a localization parameter, σ_d is the standard deviation of d_n, $\boldsymbol{\Theta}$ is a weighting matrix, and θ_m is a weight of the mth input variable. It is clear from Eq. (25) that LW-PLS is equivalent to PLS when $\varphi = 0$ and $\omega_n = 1$. Thus, selecting the localization parameter properly can lead to constructing a LW-PLS model with higher accuracy than or at least the same accuracy as PLS. The localization parameter can be determined through cross-validation. The optimal value of φ is usually found in the range of 0 to 10.

2.3 Elastic Net

Elastic net is a regularized regression method that linearly combines the L1 and L2 penalties, therefore it has the characteristics of both lasso (least absolute shrinkage and selection operator) and ridge regression (Zou and Hastie, 2005). Elastic net can overcome the limitations of lasso, which predicts the regression coefficients $\boldsymbol{\beta}$ as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_1 \|\boldsymbol{\beta}\|_1) \tag{28}$$

$$\|\boldsymbol{\beta}\|_1 = \sum_{m=1}^{M} |\beta_m|. \tag{29}$$

The use of this L1 penalty $\|\boldsymbol{\beta}\|_1$ has several limitations. First, when the number of input variables is larger than that of samples ($M > N$), lasso selects at most N variables. Second, when there is a group of highly correlated variables, lasso tends to select one variable from the group and ignore the other variables. In other words, lasso cannot cope with the collinearity efficiency. To overcome these limitations, elastic net combines lasso with ridge regression, which predicts the regression coefficients as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_2 \|\boldsymbol{\beta}\|^2) \tag{30}$$

$$\|\boldsymbol{\beta}\|^2 = \sum_{m=1}^{M} \beta_m^2. \tag{31}$$

The use of this L2 penalty $\|\boldsymbol{\beta}\|^2$ solves the collinearity problem. The regression coefficients are predicted by elastic net as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_2 \|\boldsymbol{\beta}\|^2 + \lambda_1 \|\boldsymbol{\beta}\|_1). \tag{32}$$

This penalized least squares problem is equivalent to the following optimization problem (Zou and Hastie, 2005).

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \tag{33}$$

$$\text{s.t.} \quad \alpha \|\boldsymbol{\beta}\|_1 + (1 - \alpha) \|\boldsymbol{\beta}\|^2 \leq \gamma \tag{34}$$

$$\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \tag{35}$$

3. SPARSE SAMPLE REGRESSION BASED JUST-IN-TIME MODELING (SSR-JIT)

In this section, the proposed SSR-JIT method is described. SSR-JIT selects past samples that are useful for constructing an accurate local model by using elastic net, which builds a sparse regression model to predict a query, and uses the derived regression coefficients as the similarity index ω_n in Eq. (11) for conducting LW-PLS.

3.1 Sparse Sample Regression (SSR)

In the conventional distance-based JIT methods, the similarity index ω_n is determined on the basis of distances between the query and samples in the database. These methods do not assure high prediction accuracy, because they do not take account of other information such as

correlations among variables. In addition, the conventional methods such as LWR and LW-PLS use all samples in the database to construct a local model; prediction accuracy of the local model will deteriorate if samples that have no relations with the query are used for model construction.

To solve the problem of the distance-based methods, sparse sample regression based just-in-time modeling (SSR-JIT) is proposed in this research. SSR-JIT uses elastic net to select useful samples and determine the similarity index. When a query $\boldsymbol{x}_q \in \Re^M$ is obtained, SSR-JIT solves the following optimization problem to build a sparse regression model to predict the query as a linear combination of past samples in the database.

$$\boldsymbol{\beta}_q = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{x}_q - \boldsymbol{X}^T\boldsymbol{\beta}\|^2 \tag{36}$$

$$\text{s.t.} \quad (1-\alpha)\|\boldsymbol{\beta}\|_1 + \alpha\|\boldsymbol{\beta}\|^2 \leq \gamma \tag{37}$$

where $\boldsymbol{X} \in \Re^{N\times M}$ is the input variable matrix and $\boldsymbol{\beta}_q = [\beta_1, \beta_2, \ldots, \beta_N]^{\mathrm{T}} \in \Re^N$ is the regression coefficient vector to predict the query. A part of $\boldsymbol{\beta}_q$ will be equal to zero because L1 penalty is used in elastic net. Consequently, unnecessary samples can be excluded. Furthermore, β_n represents the importance of the nth sample to express $\boldsymbol{x}_q$. Thus, the absolute value of β_n is useful as the similarity index of the sample in LW-PLS.

3.2 LW-PLS with SSR Coefficients

To build an accurate model with LWR or LW-PLS, the similarity needs to be properly defined. Although the similarity is usually defined on the basis of the Euclidean distance or the Mahalanobis distance (Kim et al., 2011; Leung et al., 2004), the prediction accuracy can be significantly improved by using the similarity index based on the weighted distance. There have been various works that calculate the weighted distance based on the regression coefficients of multiple regression analysis (MRA), PLS, and LW-PLS (Shigemori et al., 2011; Kim et al., 2013b). However, these methods require constructing a regression model in advance to calculate the similarity index, therefore the computational load is heavy. Hazama and Kano (2015) proposed covariance-based LW-PLS (CbLW-PLS) by introducing a new similarity index based on the covariance between input and output variables, and they demonstrated that CbLW-PLS achieved better prediction performance than LW-PLS with other similarity indexes through industrial case studies. In addition, Nakagawa et al. (2012a) proposed to derive weights θ_m from physical properties of target material when estimating its concentration.

In SSR-JIT, the derived sparse sample regression coefficients are used as similarity indexes ω_m to further improve the prediction performance of LW-PLS. Hence, the similarity matrix for the query is determined as follows:

$$\boldsymbol{\Omega} = \mathrm{diag}\{|\beta_1|, |\beta_2|, \ldots, |\beta_N|\} \ . \tag{38}$$

In this way, SSR-JIT selects appropriate samples and determines the similarity index of each sample simultaneously by using elastic net. LW-PLS is used in the following applications although SSR-JIT can be integrated with any JIT modeling method.

4. APPLICATIONS

In this section, the proposed JIT modeling method is validated and compared with PLS and LW-PLS through two case studies: calibration modeling from IR and NIR spectral data.

In the first case study, the concentration of magnesium stearate (MgSt) in the cleaning process was predicted from absorbance of IR spectroscopy. MgSt (Mallinckrodt, USA), which is a standard lubricant in the manufacturing of tablets, was used as an excipient. The solvent used for dissolving MgSt was methanol of HPLC grade (Wako Pure Chemical Industries, Japan). Lactose (DMV, Netherlands), mannitol (Roquette, France), cornstarch (Nisshin Seifun, Japan), and hydroxypropyl cellulose (Nippon Soda, Japan) were blended with MgSt. FT-IR spectrophotometer IR-Prestige 21 (Shimadzu, Japan) was used to measure filmy layer samples on SUS316 test pieces by IR-RAS. The range of the measured wavenumber was 4000-400 cm^{-1}, the resolution was about 4 cm^{-1}, and the scan per measurement was 45 times. VeeMaXII (PIKE Technologies, USA) was used as a measurement unit for IR-RAS. The dataset used in this case study is the same as that investigated by Nakagawa et al. (2012b).

In the second case study, the concentration of API in granules for tableting was predicted from absorbance of NIR spectroscopy. The target drug products consist of six components. Eighteen blending experiments were conducted with different API content using a 3L scale V-blender (Tsutsui Scientific Instruments Co., Ltd.). After each blending experiment, the granules for tableting were taken out and 200 mg of the granules were set in vials, then NIR spectra (2203 points in 800-2500 nm) were measured with Alliance Waters 2690 Separations Module (Waters Corporation). The dataset used in this case study is the same as that investigated by Kim et al. (2011).

In these case studies, the Savitzky-Golay filter (Rinnan et al., 2009) was used for pre-processing. Samples were divided into three sets: samples for model construction (calibration set), samples for parameter tuning (validation set), and samples for evaluating the prediction accuracy of PLS, LW-PLS, and SSR-JIT models (external validation set). Parameters were optimized so that root mean squared error (RMSE) for the validation set was minimized.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \tag{39}$$

where N is the number of samples in the data set, y_n is the nth reference value of the output variable, and $\hat{y}_n$ is the predicted value of y_n. Finally, root mean squared error of prediction (RMSEP) was calculated from the external validation set.

4.1 Estimating MgSt concentration in cleaning process

In the first case study, the concentration of magnesium stearate (MgSt) in the cleaning process was predicted from absorbances at 1868 wavenumbers of IR spectroscopy. 75 samples were used for the calibration, 75 samples were used for the validation, and 75 samples were used for the external validation.

Table 1. Results of estimating MgSt concentration in the cleaning process. PLS, LW-PLS, and SSR-JIT were compared in RMSEP and R. #LV is the number of adopted latent variables, φ is the parameter of LW-PLS, and both γ and α are the parameters of SSR-JIT.

	RMSEP	R	#LV	φ	γ	α
PLS	0.52	0.90	1	—	—	—
LW-PLS	0.40	0.94	1	0.4	—	—
SSR-JIT	0.38	0.95	12	—	0.005	0.00001

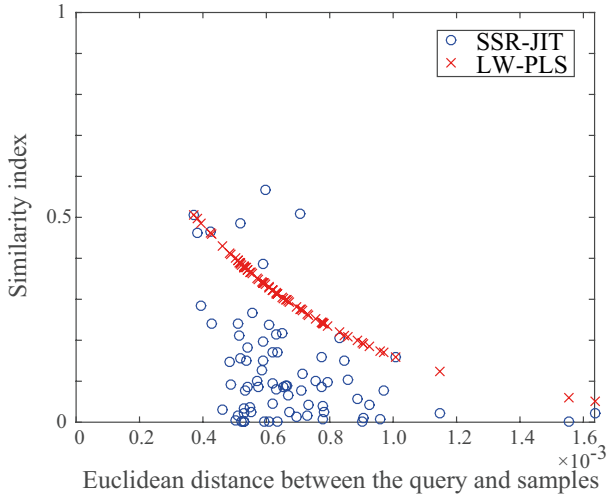

Fig. 1. Comparison of SSR-JIT and LW-PLS in the case study of predicting MgSt concentration: the relationship between the similarity index and the distance.

Table 1 shows the prediction results: RMSEP and the correlation coefficient between the measured values and the predicted values R. LW-PLS achieved the significantly better prediction accuracy than PLS, and SSR-JIT further improved the prediction accuracy.

Figure 1 shows the relationship between the similarity index and the distance when SSR-JIT or LW-PLS was applied to a sample in the external validation set. According to Eq. (25), the similarity index in LW-PLS decreases as the Euclidean distance increases. On the other hand, the similarity index in SSR-JIT does not depend on the Euclidean distance. In this case study, the sparse sample regression coefficients, i.e. the similarity index, of four samples became zero, that is, four samples were excluded by elastic net. As a result, SSR-JIT selected 71 of 75 samples in the calibration set for model construction. The results have clearly shown that the distance is not necessarily an appropriate measure of the importance of samples. The proposed SSR-JIT method can prioritize samples more appropriately than conventional distance-based JIT methods, and it can achieve better prediction performance than LW-PLS.

4.2 Estimating API concentration in granules for tableting

In the second case study, the concentration of API in granules for tableting was predicted from absorbances at 2203 wavelengths of NIR spectroscopy. 300 samples were used for the calibration, 250 samples were used for the validation, and 300 samples were used for the external validation.

Table 2. Results of estimating API concentration in the granules for tableting. PLS, LW-PLS, and SSR-JIT were compared in RMSEP and R. #LV is the number of adopted latent variables, φ is the parameter of LW-PLS, and both γ and α are the parameters of SSR-JIT.

	RMSEP	R	#LV	φ	γ	α
PLS	1.25	0.99	19	—	—	—
LW-PLS	1.22	0.99	16	1.2	—	—
SSR-JIT	1.05	0.99	30	—	0.005	0.005

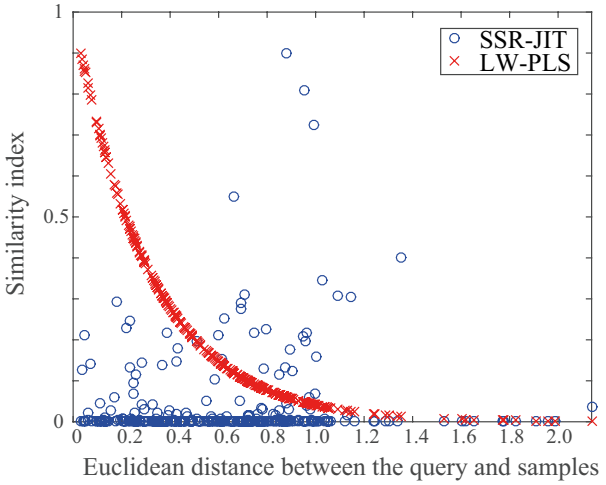

Fig. 2. Comparison of SSR-JIT and LW-PLS in the case study of predicting API concentration: the relationship between the similarity index and the distance.

Table 1 shows the prediction results: RMSEP and the correlation coefficient between the measured values and the predicted values R. SSR-JIT achieved the significantly better prediction accuracy than PLS and LW-PLS.

Figure 2 shows the relationship between the similarity index and the distance when SSR-JIT or LW-PLS was applied to a sample in the external validation set. According to Eq. (25), the similarity index in LW-PLS decreases as the Euclidean distance increases. On the other hand, the similarity index in SSR-JIT does not depend on the Euclidean distance. In particular, the similarity index of several samples became very large when the Euclidean distance was around 1.0. Similar to the previous case study, the results have clearly shown that the distance is not necessarily an appropriate measure of the importance of samples. In this case study, the sparse sample regression coefficients, i.e. the similarity index, of 173 samples became zero, that is, 173 samples were excluded by elastic net. As a result, SSR-JIT selected only 127 of 300 samples in the calibration set for model construction.

The advantage of SSR-JIT over distance-based JIT is that it can achieve high prediction accuracy even when important samples are not close to the query. The sparse sample regression (SSR) is useful to select and prioritize samples appropriately.

5. CONCLUSION

In the present work, sparse sample regression based just-in-time modeling (SSR-JIT) was proposed. The proposed method selects past samples that are useful for constructing an accurate local model by using elastic net and uses the derived sparse sample regression coefficients as the similarity index for conducting LW-PLS. The results of two case studies have demonstrated that SSR-JIT outperforms LW-PLS in the prediction accuracy.

REFERENCES

Cleveland, W.S. (1979). Robust locally and smoothing weighted regression scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836.

Fujiwara, K., Kano, M., Hasebe, S., and Takinami, A. (2009). Soft-sensor development using correlation-based just-in-time modeling. *AIChE Journal*, 55(7), 1754–1765.

Guelpa, A., Bevilacqua, M., Marini, F., O'Kennedy, K., Geladi, P., and Manley, M. (2015). Application of Rapid Visco Analyser (RVA) viscograms and chemometrics for maize hardness characterisation. *Food Chemistry*, 173, 1220–1227.

Gupta, M.R., Garcia, E.K., and Chin, E. (2008). Adaptive local linear regression with application to printer color management. *IEEE Transactions on Image Processing*, 17(6), 936–945.

Hazama, K. and Kano, M. (2015). Covariance-based locally weighted partial least squares for high-performance adaptive modeling. *Chemometrics and Intelligent Laboratory Systems*, 146, 55–62.

Hirai, T. and Kano, M. (2015). Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares. *IEEE Transactions on Semiconductor Manufacturing*, 28(2), 137–144.

Kadlec, P., Grbić, R., and Gabrys, B. (2011). Review of adaptation mechanisms for data-driven soft sensors. *Computers and Chemical Engineering*, 35(1), 1–24.

Kano, M. and Fujiwara, K. (2013). Virtual sensing technology in process industries: Trends and challenges revealed by recent industrial applications. *Journal of Chemical Engineering of Japan*, 46(1), 1–17.

Kano, M. and Ogawa, M. (2010). The state of the art in chemical process control in Japan: Good practice and questionnaire survey. *Journal of Process Control*, 20(9), 969–982.

Kim, S., Kano, M., Hasebe, S., Takinami, A., and Seki, T. (2013a). Long-term industrial applications of inferential control based on just-in-time soft-sensors: Economical impact and challenges. *Industrial and Engineering Chemistry Research*, 52(35), 12346–12356.

Kim, S., Kano, M., Nakagawa, H., and Hasebe, S. (2011). Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. *International Journal of Pharmaceutics*, 421(2), 269–274.

Kim, S., Okajima, R., Kano, M., and Hasebe, S. (2013b). Development of soft-sensor using locally weighted PLS with adaptive similarity measure. *Chemometrics and Intelligent Laboratory Systems*, 124, 43–49.

Leung, H., Huang, Y., and Cao, C. (2004). Locally weighted regression for desulphurisation intelligent decision system modeling. *Simulation Modelling Practice and Theory*, 12(6), 413–423.

Nakagawa, H., Kano, M., Hasebe, S., Miyano, T., Watanabe, T., and Wakiyama, N. (2014). Verification of model development technique for NIR-based real-time monitoring of ingredient concentration during blending. *International Journal of Pharmaceutics*, 471(1-2), 264–275.

Nakagawa, H., Tajima, T., Kano, M., Kim, S., Hasebe, S., Suzuki, T., and Nakagami, H. (2012a). 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(8), 3820–3826.

Nakagawa, H., Tajima, T., Kano, M., Kim, S., Hasebe, S., Suzuki, T., and Nakagami, H. (2012b). 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(8), 3820–3826.

Perez-Guaita, D., Kuligowski, J., Quintás, G., Garrigues, S., and Guardia, M. (2015). Modified locally weighted - Partial least squares regression improving clinical predictions from infrared spectra of human serum samples. *Talanta*, 107, 368–375.

Rinnan, Å., Berg, F., and Engelsen, S. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trac-Trends Anal. Chem.*, 28(10), 1201–1222.

Shigemori, H., Kano, M., and Hasebe, S. (2011). Optimum quality design system for steel products through locally weighted regression model. *Journal of Process Control*, 21(2), 293–301.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2), 301–320.