Chemometrics and Intelligent Laboratory Systems 183 (2018) 90–95

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Adaptive JIT-Lasso modeling for online application of near infrared spectroscopy

Jin Liu, Xiaoli Luan *, Fei Liu

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, Wuxi, 214122, PR China

ARTICLE INFO

Keywords:
Near-infrared spectroscopy
Lasso
Just-in-time learning
Modeling updating

ABSTRACT

Near infrared (NIR) spectroscopy has been widely employed as a non-invasive analytical tool in industry. However, most NIR model are built offline which cannot address changes in process characteristics as well as nonlinearity. To solve this problem, an adaptive JIT-Lasso algorithm was proposed by merging the least absolute shrinkage and the selection operator (Lasso) algorithm into just-in-time (JIT) learning. A time-space similarity measure criterion that combined temporal relevance and spatial relevance was used to further improve the performance of the JIT-Lasso algorithm. This solved both the space nonlinearity and the time-varying issue of the process simultaneously. The proposed model updating approach not only solved the nonlinear and the time-varying issues based on JIT learning framework, but also reduced the computational complexity and improved the model interpretability through Lasso. The effectiveness of the method was demonstrated on a spectroscopic dataset from an industrial petroleum desalination process. Compared with traditional partial least squares, kernel partial least square, locally weighted partial least squares, locally weighted kernel partial least squares,the proposed method achieves better performance.

1. Introduction

Near infrared (NIR) spectroscopy analysis technology is widely applied in the pharmaceutical, petrochemical, agriculture, food, environmental, and polymer fields, on account of its rapidity, non-invasiveness, minimal sample preparation, and environmental friendliness [1–3]. NIR analysis technology records spectra in real time to capture process properties and performs quantitative evaluation of various properties from a single measurement [4–6]. The NIR analysis has great potential for online detection and quick analysis, and its reliable online measurements play a vital role in process automation, monitoring, and optimization [7,8].

The online analysis is affected by variations in the real-life industrial processes, due to factors such as catalyst deactivation, equipment aging, and variations in the measurement conditions [8–11]. This results in performance of the traditional time-invariant models to worsen due to its non-adaptive nature. Model maintenance is indispensable for maintaining a high prediction performance throughout long-term online analysis [12,24,41]. In order to establish a highly-accuracy calibration model with high adaptability and stability for the promotion of online analytical techniques of NIR, some recursive methods have been applied (by prioritizing new available data to update the model) [8,13,14]. When a process is operated within a narrow range for a certain period of region, the model could become over-adapted under a sufficiently wide range of operating conditions [15]. The adapted soft sensor might not work well during a new operating region, until a sufficient period of time has passed, due to the time delay in adapting the recursive method to the new process operating conditions [24]. A more serious drawback of these recursive methods is that they cannot cope with abrupt changes in the process characteristics [16].

The just-in-time (JIT) learning [17] has been used to address the process nonlinearity and the abrupt changes in these process characteristics [18–20]. JIT has also been used during process modeling and monitoring of NIR. LW-PLS was used to estimate the soil properties via NIR spectroscopy [21]. The analysis of sawdust via NIR and locally weighted partial least squares was compared with traditional methods [22]. NIR-based real-time monitoring of ingredient concentration was obtained via LW-PLS [23]. A sparse sample regression based just-in-time modeling (SSR-JIT) was applied to predict the API granules concentration of tableting via NIR spectroscopy [24]. Although PLS has been commonly used to reduce the dimension of wavelength variables in near infrared spectral region (780–2526 nm), the

* Corresponding author.
E-mail address: xiaoli_luan@126.com (X. Luan).

https://doi.org/10.1016/j.chemolab.2018.10.012
Received 9 July 2018; Received in revised form 19 October 2018; Accepted 27 October 2018
Available online 29 October 2018

compressed latent variables still contain uncorrelated information and noise, which can severely corrupt the resulting calibration models and hardly interpretable original features (wavelengths in the case of infrared spectra). Therefore, variable selection during the NIR modeling process is important.

The elimination of the uninformative wavelength variables can result in better prediction performance and less model complexity [25,41]. There are many multivariate methods of variable selection applied in quantitative analysis of NIR such as uninformative variable elimination (UVE) [26], successive projection algorithm (SPA) [27], variable importance in the projection (VIP) [28], simulated annealing (SA) [29], and genetic algorithms (GAs) [30]. The above variable selection methods separate the variable selection as an independent step outside the modeling which results in the increase of the modeling step and the reduction of the convenience of operation. In contrast, Lasso performs variable selections and regression analysis simultaneously, which is usually used in variable selection, modeling, and model updating of the NIR [24,31,38].

The novel adaptive JIT-Lasso algorithm that integrates JIT learning and Lasso algorithm was proposed for the nonlinear process. The JIT-based method could cope with processing the nonlinearity and Lasso selected a few effective variables to reduce the computational complexity and to improve model interpretability. A similarity measurement criterion that solved both the nonlinearity of the space and the time-varying issues of the process simultaneously was applied in the JIT-Lasso method. The proposed adaptive method was updated with effective variables from the relevant samples, which helped to maintain the robustness of the established models. The proposed method is an online analytical technique with high adaptability and stability in the actual industrial process. The usefulness of the proposed method was demonstrated via a petroleum desalination dataset from the NIR spectra.

2. Adaptive JIT-Lasso modeling

The NIR spectrum, measured via a spectrometer, contained a large number of wavelength variables in order to cover the full spectra of high resolution [32]. The full spectral data included the variable information, which also brought collinearity, redundancy, and noise issues that could deteriorate the estimated performance of the calibration model. Manual variable selection required the engineers' personal experience and prior process knowledge. The method was time-consuming as the engineers had to select the input variables, which might not be optimal when a large number of variables were measured and physical and chemical phenomena were not fully understood [33–35]. The appropriate variable selection technique should improve the prediction performance and reduce the complexity of calibration models, due to the removal of the meaningless wavelengths. Many multivariate analysis techniques, including PCA and PLS, were considered methods of "soft" feature selection. The methods use every variable in the data, even those that were noisy and could have an adverse effect on the results. The variables that produced by these methods are difficult to interpret in actual process variables [36].

2.1. Lasso regression algorithm

The least absolute shrinkage and selection operator (Lasso) [37], was an efficient technique for shrinkage and feature selection. It was often considered a "hard" feature selection technique because the variables that were determined to be less informative are eventually omitted from the regression equation [38].

Lasso regression function minimized the residual sum of squares subject to the sum of the absolute value of the coefficients that were less than a constant. Suppose that $X = [x_{1j}, \cdots, x_{ij}]^T \in \Re^{N\times M}$ and $Y = [y_1, \cdots, y_i]^T \in \Re^{N\times 1}$ were the spectra and physical parameter matrices. N was the number of samples and the M denoted the number of wavelength variables. The Lasso was defined as:

$$\beta_{Lasso} = \left[\sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{M}\beta_j x_{ij}\right)^2\right] s.t. \sum_{j=1}^{M}|\beta_j| \le t \tag{1}$$

[1] was equivalent to the following equation

$$\beta_{Lasso} = \left[\sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{M}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{M}|\beta_j|\right] \tag{2}$$

Assuming that $X^TX + \lambda\Omega^-$ was nonsingular, the solution was follows:

$$\widehat{\beta}_{Lasso} = \left(X^TX + \lambda\Omega^-\right)^{-1}X^TY \tag{3}$$

where $\widehat{\beta}_{Lasso}$ was the estimated coefficient in the regression equation and $\Omega = diag(\left|\widehat{\beta}_j\right|)$.

The parameter λ was predetermined and controlled the amount of shrinkage of the regression coefficients. The higher the value of λ, the more elements of the estimated coefficients were set to zero. The smaller λ implied less constraint and more nonzero coefficients. This was appealing from the view of feature selection, as it selected the significant variables and discarded the rest to improve the model interpretability and the precision.

2.2. JIT-Lasso regression algorithm

The Lasso regression is a linear time-invariant model and is not suitable for nonlinear or time-varying processes. The just-in-time (JIT) learning addresses the process nonlinearity and abrupt changes. In the JIT-Lasso learning structure, a local Lasso model was constructed by selecting samples related to the query sample from the historical samples. The local Lasso model was used to obtain the estimated physical parameter value. The local model was discarded right after the prediction was obtained. Therefore, the JIT-based approach can address the process nonlinearity and track the current state of the process well. The process could cope with the abrupt changes, as well as the gradual ones [39]. From a philosophical point of view, the method adopted the "divide and conquer" strategy, where the global model was locally linearized around some samples so that nonlinear process could be modeled [40]. A linear regression Lasso model, based JIT learning could successfully cope with the process nonlinearity.

The proposed JIT-Lasso algorithm updated the model online by implementing the Lasso method in the JIT learning framework. The following section summarized the JIT-Lasso algorithm procedure:

1. When the query sample x_q arrived, the similarity w_i between x_q and the historical sample point x_i were calculated.
2. The similarity matrix W was formed as: $W = diag(w_1, w_2, \cdots, w_N)$.
3. $\tilde{X}$ and $\tilde{Y}$ were calculated as follows:

$$\overline{x} = \sum_{i=1}^{N} w_i x_i \Big/ \sum_{i=1}^{N} w_i \tag{4}$$

$$\overline{y} = \sum_{i=1}^{N} w_i y_i \Big/ \sum_{i=1}^{N} w_i \tag{5}$$

$$\tilde{X} = X - 1_N\overline{x} = \left[\tilde{x}_1^{\mathrm{T}} \quad \tilde{x}_2^{\mathrm{T}} \quad \cdots \quad \tilde{x}_N^{\mathrm{T}}\right]^{\mathrm{T}} \tag{6}$$

$$\tilde{Y} = Y - 1_N\overline{y} = \left[\tilde{y}_1^{\mathrm{T}} \quad \tilde{y}_2^{\mathrm{T}} \quad \cdots \quad \tilde{y}_N^{\mathrm{T}}\right]^{\mathrm{T}} \tag{7}$$

where $1_N \in R^N$ was a vector of ones.

4. The $\beta_{JIT-Lasso}$ was estimated by minimizing the following equation:

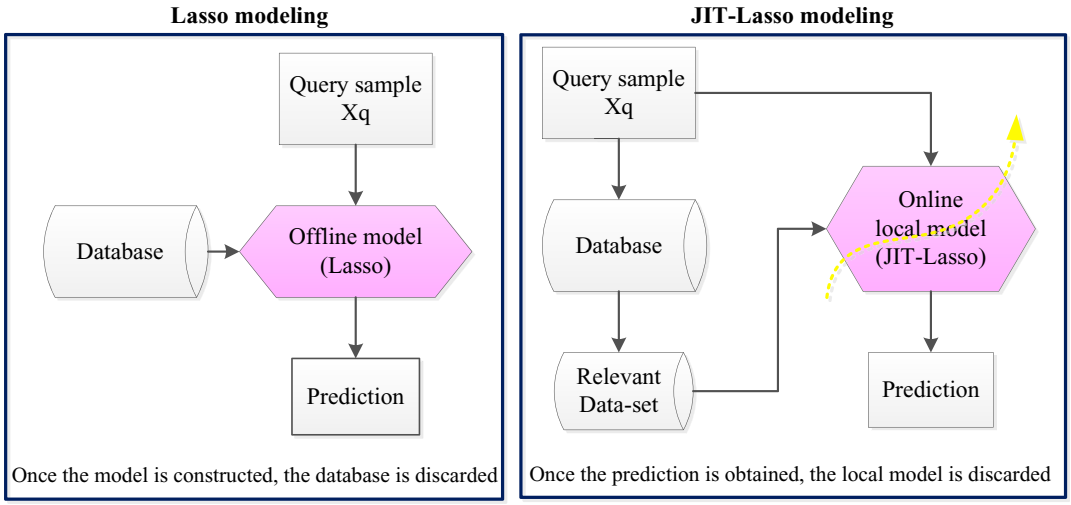

Fig. 1. Comparisons between the Lasso method and the JIT-Lasso method.

$$\sum_{i=1}^{N} W\left(\tilde{y}_i - \sum_{j=1}^{M}\beta_j\tilde{x}_{ij}\right)^2 + \lambda\sum_{j=1}^{M}|\beta_j| \tag{8}$$

Assuming that $X^TWX + \lambda\Omega^-$ was nonsingular, the solution was follows:

$$\widehat{\beta}_{JIT-Lasso} = \left(\tilde{X}^T W\tilde{X} + \lambda\Omega^-\right)^{-1}\tilde{X}^T W\tilde{Y} \tag{9}$$

where $\beta_{JIT-Lasso}$ was the estimated coefficient in the regression equation, and the definition of Ω and λ were the same as the Lasso method.

5. Thus, $\hat{y}_q = x_q\widehat{\beta}_{JIT-Lasso}$ was predicted.

Remark 1. The drawback of the Lasso modeling included that the construction of single global model was offline and rarely produced good approximations under changing conditions. The JIT-Lasso dynamically built models by using relevant data similar to the query sample, which circumvented the nonlinearity of the uncertain factor. Fig. 1 illustrates the difference between the Lasso method and the JIT-Lasso algorithm.

2.3. Similarity measure criterion

To determine the accurate estimation of the JIT-based models, the similarity measurement criterion requires proper definition, according to characteristics of the variables and the processes. The similarity, based on the distance, could be regarded as separating the high relevance samples from the spatial scale. Most industrial processes are continuous and dynamic, where the duration of the process characterizes the various states and phases of the process [41]. We hope the similarity measurement could also be determined from the temporal scale. The distinction between samples are primarily manifested by various distances between the spectra in the NIR field [42]. We proposed a time-space similarity criterion that combined a temporal relevance and a spatial relevance, based on the Euclidean distance. This method had the capability to solve both the nonlinearity of the space and the time-varying issues of the process. The new similarity measure was calculated through [10]-(14):

$$w_i^{space} = \exp\left(-\frac{d_i^{space}}{\sigma_d{}^{space}\psi_{space}}\right) \tag{10}$$

$$w_i^{time} = \exp\left(-\frac{d_i^{time}}{\sigma_d{}^{time}\psi_{time}}\right) \tag{11}$$

Table 1
Statistics of reference values for the training set and the testing set.

Sample sets	Number	Range (mg/L)	Mean (mg/L)	Standard deviations (mg/L)
Training set	262	19.01–49.67	22.50	2.60
Testing set	166	17.67–28.05	22.96	2.27

$$d_i^{space} = \left\|x_q - x_i\right\| \tag{12}$$

$$d_i^{time} = \left\|t_q - t_i\right\| \tag{13}$$

$$w_i = \mu w_i{}^{space} + (1-\mu)w_i{}^{time} \tag{14}$$

where $0 \le \mu \le 1$, and $w_i, w_i{}^{space}$ and $w_i{}^{time}$ $(i = 1, 2, \cdots, N)$ denoted the time-space weight, the space weight and the time weight. σ_d and σ_d were the standard deviations of d_i and d_i. t_q and t_i represented the industrial process duration matched with the query data x_q and historical samples x_i in the database. ψ_{space} and ψ_{time} were two localization parameters.

Remark 2. [14] showed that the proposed time-space similarity measure was equivalent to the single similarity measure when $\mu = 0$ or $\mu = 1$. If the similarity matrix W was an identity matrix, the JIT-Lasso method becomes the equal to the Lasso method.

3. Industry dataset and preprocessing

The NIR spectra used for the case study was obtained from the petroleum desalting process. A total of 428 samples were successively collected from the online operation, with a sampling interval of 20 min. The earliest 262 samples (taken from November 26 to November 29,

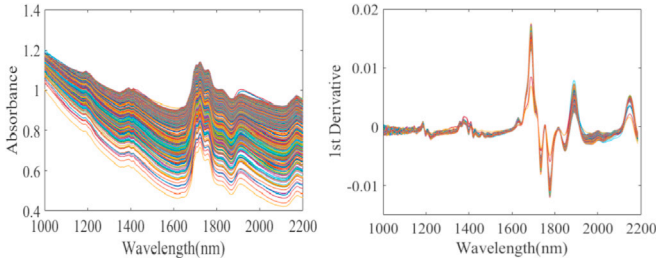

Fig. 2. Original and first order derivative spectra.

Table 2
Quantitative comparison of the prediction results for various soft sensors.

Methods	Type	RMSE	R_p
PLS	Global-model	1.3162	0.3851
KPLS	Global-model	1.2897	0.7208
Lasso	Global-model	1.2932	0.0510
JIT-Lasso A	JIT learning with single similarity	1.2330	0.5464
JIT-Lasso B	JIT learning with single similarity	1.1434	0.7424
JIT-Lasso C	JIT learning with time-space similarity	1.1232	0.7669

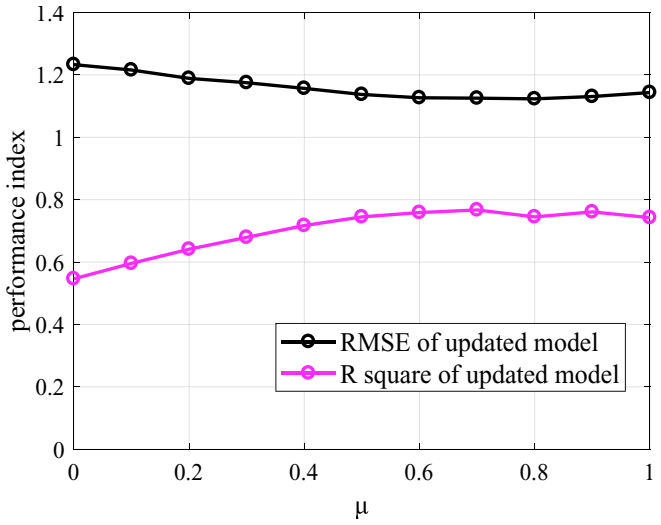

Fig. 3. The prediction performance under different tuning parameter (μ).

2016) were used to form the initial database (training set). The remaining 166 samples (taken from November 30 to December 2, 2016) were used to test the algorithm (testing set). The target property was the chloride content in the oil measured via online analyzers. The sample statistics were summarized in Table 1. The spectra was measured by a Fourier-type NIR spectrometer (Matrix-F, Bruker Optics Inc., Germany) with a wavelength range between 1000 and 2200 nm with a resolution of 2.0 cm^{-1}. The spectra was preprocessed by first order derivative that eliminated the baseline effect and highlighted the characteristics of the raw spectra [43]. The original spectra and the preprocessed spectra are seen in Fig. 2. The data analysis was implemented with Unscrambler 9.6 (CAMO, Oslo, Norway) and Matlab R2017a.

4. Results and discussion

The experiments aimed to verify the necessity of model updating and the effectiveness of the proposed time-space similarity criterion. The μ of [14] had different values from 0 to 1. When $\mu = 0$, the method became a time similarity measurement. When $\mu = 1$, the method became the space similarity measure. Both the time similarity measure and the space similarity measure were conventional single similarity measurement methods. To show the advantage of JIT-Lasso over others, five modeling approaches were compared with it, which are listed as follows:

(1) PLS: A fixed linear model.
(2) KPLS: A fixed linear model.
(3) Lasso: A fixed linear model.
(4) JIT-Lasso A: JIT-Lasso with a time similarity measurement.
(5) JIT-Lasso B: JIT-Lasso with a space similarity measurement.
(6) JIT-Lasso C: JIT-Lasso with a time-space similarity measurement.

The quantitative prediction results of each model are listed in Table 2. The model performance was demonstrated by the root-mean-square error

Table 3
Quantitative comparison of the prediction results for JIT methods.

Methods	Type	RMSE	R_p
LW-PLS	JIT learning	1.3147	0.2762
LW-KPLS	JIT learning	1.2642	0.7071
JIT-Lasso C	JIT learning with time-space similarity	1.1232	0.7669

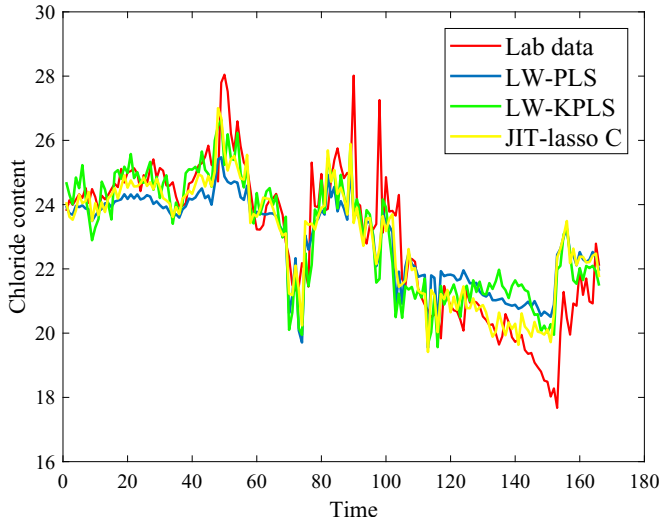

Fig. 4. Prediction results for the LW-PLS, LW-KPLS and the JIT-Lasso C, compared with the reference data (the red line).

(RMSE) of prediction and the correlation coefficient (R_p). For fixed linear models, the PLS and Lasso had poor prediction in terms of the higher error values, because the fixed linear model could not handle the nonlinear and the time-varying characteristics of the process adequately. The fixed nonlinear model KPLS had better performance than fixed linear models because it can explain part of the nonlinearity of process. Specially, JIT-Lasso methods were superior to others obviously, which indicates that JIT learning framework is an effective model updating method for addressing the process nonlinearity. For setting of the parametersμ, the prediction results with different parameter value of μ are displayed in Fig. 3.

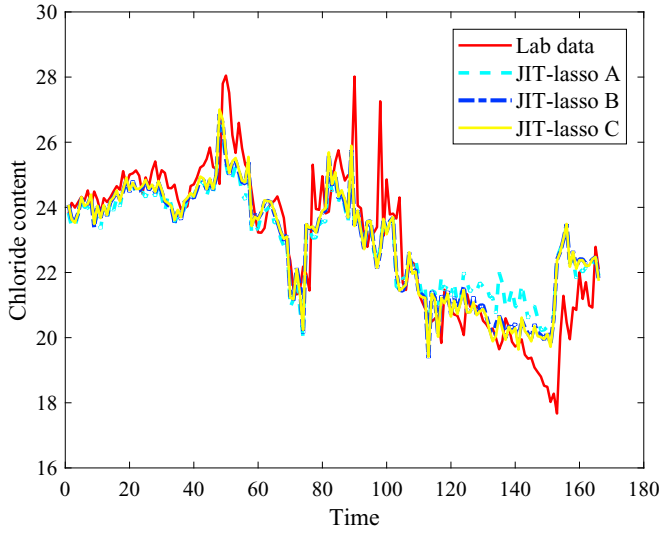

Fig. 5. Prediction results for the JIT-Lasso A-C, compared with the reference data (the red line).

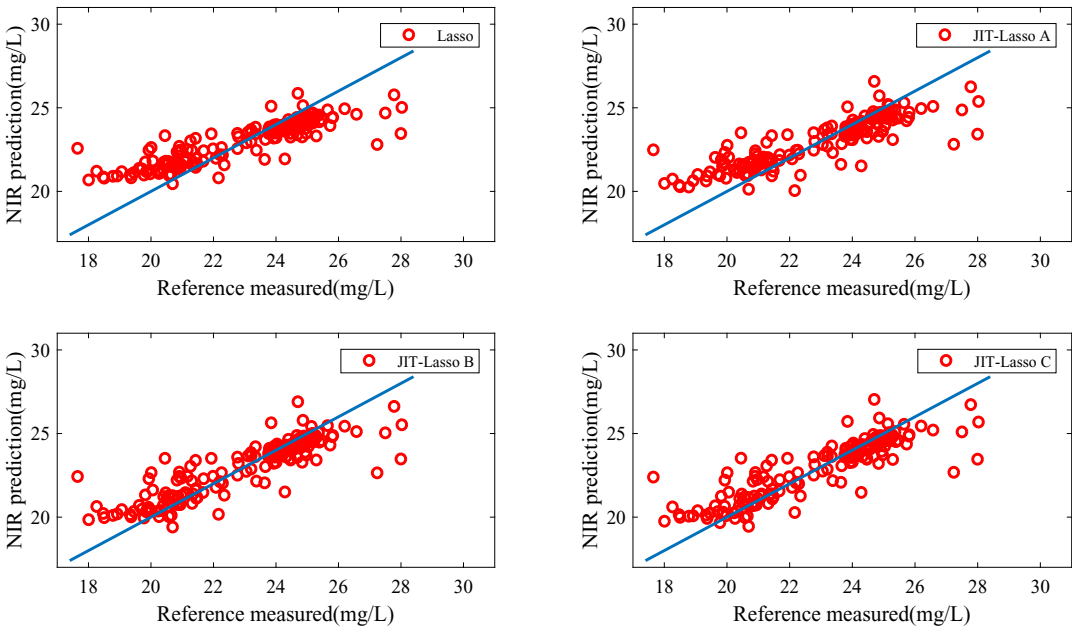

Fig. 6. Scatter plots of the chloride content predictions.

From Fig. 3, when $0 < \mu < 1$, the JIT-lasso methods with the time-space similarity had better performances than the JIT-lasso A and B methods with the single similarity. It indicates that the proposed time-space similarity criterion considering both the different states of the process from the time scale and different distances of samples from the space scale leading to the better performance. The parameter value $\mu = 0.8$was chosen, which had the highest R_p and the lowest RMSE. Although μ was close to 1, it did not mean that the time weight was unnecessary. For processes with serious time varying, a smaller μ would be better. For processes with severe non-linearity and less time varying (same as our case), a larger μ might be more appropriate. The selection of the optimal values of these parameters was dependent on the data set.

To show the advantage of the proposed variable selection method, the following locally weighted PLS and locally weighted KPLS modeling methods were compared.

(1) Locally weighted PLS (LW-PLS) model: PLS applied to JIT learning.
(2) Locally weighted KPLS (LW-KPLS) model: KPLS applied to JIT learning.

The quantitative prediction results of JIT methods are listed in Table 3, which shows that the JIT-Lasso C had better performance than LW-PLS and LW-KPLS. This indicates that the Lasso ("hard" feature selection) had selected more effective variables and removed more unrelated variables than the PLS and KPLS ("soft" feature selection) which cannot eliminate the uninformative wavelength variables of NIR. The detailed prediction results are shown in Figs. 4 and 5.

Fig. 4 shows that the JIT-Lasso C tracked the dynamic process better than LW-PLS and LW-KPLS in the case of large fluctuations in chlorine concentration. Fig. 5 shows that JIT-Lasso B had a better performance than the JIT-Lasso A within the tracking process dynamics, indicating that the process had more spacial non-linearity and less time variation issues. Figs. 4 and 5 show that the JIT-Lasso C obtained a better prediction performance, since the JIT-Lasso C used the time-space similarity measure. There were still some cases where the performance of model was not satisfactory. It should be noted that at around time 110, the separator of the desalination device began to malfunction during the petroleum desalination process. The fault was caused by corrosion of the separator. With the increasing corrosion degree, the failure of the separator eventually led to the sharp increase of water and chloride in desalted crude oil. On the other hand, the fixed database used to update did not contain abnormal operating conditions samples in our experiment. Therefore, at around time 150 the prediction of the method proposed looks different from lab data. However, at around time 170, the proposed method tracked the dynamic process again. This indicates that the proposed method is difficult to track the dynamic process at the preliminary stage of the sudden changing of operating condition, such as the fault. Scatter plots of the four modeling methods (Lasso, JIT-Lasso A, JIT-Lasso B, and JIT-Lasso C) are shown in Fig. 6, demonstrating that the JIT-Lasso C exceeded all other algorithms.

5. Conclusions

The JIT-Lasso algorithm provided a novel adaptive modeling framework for nonlinear processes. The algorithm addressed the nonlinear problems based on JIT learning, and simultaneously reduced the computation and improved model interpretability by Lasso. In order to further improve the prediction performance of JIT-Lasso, the new similarity measure criterion that accounted for the space nonlinearity and the time-varying issues of the process was applied. The feasibility and the effectiveness of the proposed algorithm was illustrated by an industrial NIR case study of the petroleum desalting process. Other regularization methods that merged into the JIT learning should obtain a good performance. The results indicated that the near-infrared spectroscopy is a valuable tool for the on-line process modeling. The method was suitable and had potential application value in industrial processes.

Conflicts of interest

There is no conflict of interest.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61722306, 61833007), and the national first-class discipline program of Light Industry Technology and Engineering (LITE2018-25).

References

[1] R.M. Balabin, S.V. Smirnov, Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data, Anal. Chim. Acta 692 (1–2) (2011) 63–72.

[2] 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, Anal. Chem. 84 (8) (2012) 3820–3826.

[3] Y. Roggo, P. Chalus, L. Maurer, C. Lemamartinez, A. Edmond, N. Jent, Evaluation of infrared-reflection absorption spectroscopy measurement and locally weighted partial least-squares for rapid analysis of residual drug substances in cleaning processes, Anal. Chem. 84 (8) (2012) 3820–3826.

[4] E.A. Lidia, R. Charles, J. Hurburgh, A Tutorial on near infrared spectroscopy and its calibration, Crit. Rev. Anal. Chem. 40 (4) (2010) 246–260.

[5] X.B. Zou, J.W. Zhao, M.J. Povey, M. Holmes, H.P. Mao, Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta 667 (1–2) (2010) 14–32.

[6] R.A. Rossel, S.R. Cattle, A. Ortega, Y. Fouad, In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy, Geoderma 150 (3–4) (2009) 253–266.

[7] J.S. Alford, Bioprocess control: advances and challenges, Comput. Chem. Eng. 30 (10) (2006) 1464–1475.

[8] Q.P. Mei, Y.K. Tang, T.F. Li, L.Z. Yao, Q. Yang, H.J. Zhang, X.H. Liu, Characterization of a ssss adaptive calibration model using near-infrared spectroscopy and partial least squares with a Kalman filter, Anal. Lett. 51 (8) (2018) 1176–1193.

[9] Z.Q. Ge, Z.H. Song, A comparative study of just-in-time-learning based methods for online soft sensor modeling, Chemometr. Intell. Lab. Syst. 104 (2) (2010) 306–317.

[10] M. Chen, S. Khare, B. Huang, A unified recursive just-in-time approach with industrial near infrared spectroscopy application, Chemometr. Intell. Lab. Syst. 135 (14) (2014) 133–140.

[11] S. Kim, M. Kano, H. Nakagawa, S. Hasebe, Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection, Int. J. Pharm. 421 (2) (2011) 269–274.

[12] H. Min, X.L. Luo, Enhanced just-in-time soft sensor calibration method using data density estimation, Chemometr. Intell. Lab. Syst. 161 (2017) 79–87.

[13] S. Jia, H. Li, Y. Wang, R.Y. Tong, Q. Li, Recursive variable selection to update near-infrared spectroscopy model for the determination of soil nitrogen and organic carbon, Geoderma 268 (2016) 92–99.

[14] J.F. Wang, Recursive PLS modeling of aromatics and olefins in Gasoline, Adv. Mater. Res. 740 (2013) 267–272.

[15] K. Hazama, M. Kano, Covariance-based locally weighted partial least squares for high-performance adaptive modeling, Chemometr. Intell. Lab. Syst. 146 (2015) 55–62.

[16] K. He, H. Cheng, W.L. Du, F. Qian, Online updating of NIR model and its industrial application via adaptive wavelength selection and local regression strategy, Chemometr. Intell. Lab. Syst. 134 (8) (2014) 79–88.

[17] C. Cheng, M.S. Chiu, A new data-based methodology for nonlinear process modeling, Chem. Eng. Sci. 59 (13) (2004) 2801–2810.

[18] K. Fujiwara, M. Kano S, Soft-sensor development using correlation-based just-in-time modeling, AIChE 55 (7) (2009) 1754–1765.

[19] S. Kim, M. Kano, S. Hasebe, A. Takinami, T. Seki, Long-term industrial applications of inferential control based on just-in-time soft-sensors: economical impact and challenges, Ind. Eng. Chem. Res. 52 (35) (2013) 12346–12356.

[20] Y. Liu, Z. Gao, P. Li, H.Q. Wang, Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes, Ind. Eng. Chem. Res. 51 (11) (2012) 4313–4327.

[21] C.D. Christy, S.A. Dyer, Estimation of soil properties using a combination of spectral and scalar sensor data, in: 2006 IEEE Instrumentation and Measurement Technology Conference Proceedings, 2007, pp. 729–734.

[22] A.S. Rathore, K.V. Gernaey, C.R.C. Calado, Y. Seongkyu, Process analytical technologies in biopharmaceutical process development, Chemical Technology and Biotechnology 90 (2) (2015) 213–214.

[23] H. Nakagawa, M. Kano, S. Hasebe, T. Miyano, Verification of model development technique for NIR-based real-time monitoring of ingredient concentration during blending, Int. J. Pharm. 471 (2) (2014) 264–275.

[24] T. Uchimaru, M. Kano, Sparse sample regression based just-in-time modeling (SSR-JIT): beyond locally weighted approach, IFAC Papersonline 49 (7) (2016) 502–507.

[25] X.B. Zou, J.W. Zhao, M.J. Povey, M. Holmes, H.P. Mao, Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta 667 (1–2) (2010) 14–32.

[26] V. Centner, D. Massart, O. deNoord, S. deJong, B. Vandeginste, C. Sterna, Elimination of uninformative variables for multivariate calibration, Anal. Chem. 68 (21) (1996) 3851–3858.

[27] M. Araujo, T. Saldanha, R. Galvao, T. Yoneyama, H. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, Chemometr. Intell. Lab. Syst. 57 (2) (2001) 65–73.

[28] I. Chong, C. Jun, Performance of some variable selection methods when multicollinearity is present, Chemometr. Intell. Lab. Syst. 78 (1) (2005) 103–112.

[29] H. Swierenga, P.D. Groot, A.D. Weijer, M. Derksen, L. Buydens, Improvement of PLS model transferability by robust wavelength selection, Chemometr. Intell. Lab. Syst. 41 (2) (1998) 237–248.

[30] R. Leardi, A. Gonzalez, Genetic algorithms applied to feature selection in PLS regression: how and when to use them, Chemometr. Intell. Lab. Syst. 41 (2) (1998) 195–207.

[31] K. Fujiwara, M. Kano, Efficient input variable selection for soft-senor design based on nearest correlation spectral clustering and group Lasso, ISA (Instrum. Soc. Am.) Trans. 58 (2015) 367–379.

[32] M. Chen, S. Khare, B. Huang, H.T. Zhang, E. Lau, E.B. Feng, Recursive wavelength-selection strategy to update near-infrared spectroscopy model with an industrial application, Industrial & Engineering Chemistry Research 52 (23) (2013) 7886–7895.

[33] L. Cui, P. Xie, J.W. Sun, T. Yu, J.Q. Yuan, Data-driven prediction of the product formation in industrial 2-keto-l-gulonic acid fermentation, Computers & Chemical Engineering 36 (1) (2012) 386–391.

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

[35] A.K. Pani, V.K. Vadlamudi, H.K. Mohanta, Development and comparison of neural network based soft sensors for online estimation of cement clinker quality, ISA Transactions 52 (1) (2013) 19–29.

[36] M.R. Delgado, E.Y. Nagai, A neuro-coevolutionary genetic fuzzy system to design soft sensors, Soft Computing 13 (5) (2009) 481–495.

[37] R.J. Tibshirani, Regression shrinkage and selection via the LASSO, Royal Statistical Society 58 (1) (1996) 267–288.

[38] M.D. Dyar, M.L. Carmosino, E.A. Breves, M.V. Ozanne, S.M. Clegg, R.C. Wiensb, Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples, Spectrochimica Acta Part B Atomic Spectroscopy 70 (1) (2012) 51–67.

[39] K. Chen, J. Ji, H.Q. Wang, Y. Liu, Z.H. Song, Adaptive local kernel-based learning for soft sensor modeling of nonlinear processes, Chemical Engineering Research & Design 89 (10) (2011) 2117–2124.

[40] W.M. Shao, X.M. Tian, P. Wang, Supervised local and non-local structure preserving projections with application to just-in-time learning for adaptive soft sensor, Chinese Journal of Chemical Engineering 23 (12) (2015) 1925–1934.

[41] H.P. Jin, X.G. Chen, J.W. Yang, L. Wu, Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes, Computers & Chemical Engineering 71 (2014) 77–93.

[42] R.F. Shan, Y. Zhao, M.L. Fan, X.W. Liu, W.S. Cai, X.G. Shao, Multilevel analysis of temperature dependent near-infrared spectra, Talanta 131 (2015) 170–174.

[43] A. Rinnan, F.V.D. Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, Trends in Analytical Chemistry 28 (10) (2009) 1201–1222.