CHEMICAL ENGINEERING TRANSACTIONS

VOL. 70, 2018

A publication of AIDIC

The Italian Association of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Timothy G. Walmsley, Petar S. Varbanov, Rongxin Su, Jiří J. Klemeš

ISBN 978-88-95608-67-9; **ISSN** 2283-9216

DOI: 10.3303/CET1870240

Soft Sensor Modelling based on Just-in-Time Learning and Bagging-PLS for Fermentation Processes

Congli Mei[a,*], Yuhang Ding[b], Xu Chen[b], Yao Chen[b], Jiangpin Cai[a]

[a]College of Electrical Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, P.R.China
[b]Department of Automation, Jiangsu University, Zhenjiang 212013, P.R.China
clmei@126.com

For process modelling of complex and dynamic industrial processes, it has been proven that just-in-time learning (JITL) method has some advantages. The key to success of JITL modelling is how to select appropriate relevant samples for modelling. However, the size of most relevant samples selected by using traditional similarity indexes is difficult to be determined and still needs further research. To handle the problem, a novel relevant sample selection method is proposed in this work by constructing a similarity index based on Gaussian kernel function. In the method, confidence value is used to replace traditional error limit as a threshold value for relevant sample selection. Meanwhile, the bagging strategy is used to resample samples from selected samples to avoid setting the size of selected samples. In this work, the common modelling method on industrial processes, partial least square (PLS), is used to build regression models. To verify the proposed method, a mathematical model of a nonlinear system and the benchmark of a Penicillin fermentation process simulation were both studied. Results show that compared to traditional JITL methods, the proposed method has significant advantages in terms of the robustness and prediction precision of the model.

1. Introduction

Due to the advancement of intelligent instruments, massive process data is collected and stored every day. These data have been used to build data-driven soft sensors (Souza et al., 2016). The most common data-driven modelling techniques for developing soft sensors are multivariate statistical techniques such as multivariate linear regression principal component regression and partial least squares (PLS). Nevertheless, these linear methods cannot function well while being applied to highly nonlinear processes like batch processes. Thus, many efforts have been attempted to nonlinear approaches (Souza et al., 2016), such as artificial neural networks (ANN), kernel partial least squares (KPLS), support vector regression (SVR), neuro-fuzzy system and Gaussian process regression (GPR) (Mei et al., 2017).
However, there exists some issues remained to be solved in developing soft sensors for batch processes. One particular drawback of many traditional soft sensors is their non-adaptive nature (Kadlec et al., 2011). Traditionally, the structure and parameters of the predictive models are fixed. It means that those models do not change while being applied to real-life processes. However, many process factors such as the environment, process recipes and the state of equipment often change. Moreover, there is still a lack of effective maintenance methods of soft sensors (Abonyi et al., 2015). In addition, the traditional model training methods become ineffective while facing mass process data.
Recently, JITL has attracted increasing attention in soft sensor development for its potential in coping with these problems (Chen et al., 2017). By applying JITL, a local model is constructed from the samples similar to the query data. Thus, on the one hand, JITL based soft sensors can cope with abrupt changes as well as gradual ones (Mei et al., 2017). On the other hand, it can deal with nonlinearity since it builds a local model repeatedly. Compared to the traditional modelling methods by using global modelling, JITL based method exhibits a local structure. In the method, a local model is built from the historical data set selected by a similarity measure to the query data when the estimation is required.

Please cite this article as: Mei C., Ding Y., Chen X., Chen Y., Cai J., 2018, Soft sensor modelling based on just-in-time learning and bagging-pls for fermentation processes , Chemical Engineering Transactions, 70, 1435-1440 DOI:10.3303/CET1870240

In traditional JITL-based soft sensor, distance-based similarity index and its variants are often used for relevant sample selection. However, there are still several theoretical and practical issues that have yet to be overcome. For instance, the sample scale of relevant samples is difficult to be determined. Although the significant efforts of many researcher have been done in the past in the development and improvement of JIT model, the challenge remains unsolved.
To handle the problem, a novel JITL-based soft sensor based on bagging technology is proposed. Different from traditional JITL-based soft sensors, Gaussian function is used as a similarity index. Confidence values of similarity is used to avoid determining sample scale. To enhance the robustness of the soft sensor model, the bagging strategy is used in modelling. In this work, the common modelling tool, PLS (Wang et al., 2015), used to evaluate the proposed modelling strategy.

2. Proposed soft sensor

2.1. Gaussian function-based similarity criterion

The key step of JITL modelling is the selection of relevant samples (Saptoro, 2014). The selected samples from the database is neighbouring data around the query data. The neighbourhood is defined as any data having similarity with the query point. To evaluate the similarity, distance-based measure, especially Euclidian distance is commonly used due to its simplicity.
In recent years, there are several variants of distance-based similarity factor: Euclidean distance based, weighted Euclidian distance based and Mahalanobis distance based. JITL modelling is carried out by the relevant samples according to the similarity factor and specified distance measure.
However, the parameter, relevant sample scale, of traditional distance-based measures is difficult to be determined because of the complexity of process data. In this study, Gaussian function is used as the similarity factor which can be formulated as follows:

$$K\left(x_q, x_i\right) = \exp\left(-\frac{\left\|x_q - x_i\right\|^2}{2\sigma^2}\right) \qquad (1)$$

For Gaussian function-based similarity, traditional sample scale setting by using distance-based measures is avoided and confidence value of similarity is introduced to the selection of relevant samples. Usually, the confidence value is set to 0.9. It means x_i is selected as a relevant sample if $K\left(x_q, x_i\right)$ is greater than 0.9. In (1), σ in (1) can be set by statistical computation of x_i or according to the precision of instruments.

2.2. Bagging method

Bagging (Chen and Ren, 2009), short for “Bootstrap AGGregatING”, is a method of obtaining more robust and accurate models using bootstrap re-samples of training data. The procedure for bagging consists of two stages. First bootstrap samples are obtained from the original data to form a set of training sets, from which multiple models are developed. Then these models are combined in some way to make predictions. Bagging can be applied to both regression and classification models, whilst the focus is on regression in this study. It was shown that bagging is especially suitable for “unstable” models, i.e. the models that are sensitive (in terms of model parameters and predictive performance) to small changes in training data. In practice, bagging has been applied to a large number of model structures, including regression trees, regression with variable selection and neural networks.
In JIT-based soft sensor modelling, the basic idea of using bagging technology is straightforward. Instead of making predictions from a single regression model that is fit to the most relevant data of samples, a number of models are developed to characterize the same relationship between input and output variables of samples. Each model is developed from a bootstrap re-sampled set of relevant data. Then the predictions from the multiple models are combined to improve model accuracy and robustness. What’s more, in the modelling the parameters, error limit and relevant sample scale, are not needed to be predetermined.

2.3. PLS model

PLS (Höskuldsson, 1988) is a well-known multivariate statistical technique for modelling the relationship between *p* process variables, $\mathbf{x} \in \Re^{n\times p}$, and l predicted variable, $\mathbf{y} \in \Re^{n\times l}$, with *n* samples, as shown as follows,

$$X = TP^T + E \qquad (2)$$

$$Y = TQ^T + F \qquad (3)$$

$$T = XW(P^TW)^{-1} \tag{4}$$

Where $T \in \Re^{n\times A}$, $P \in \Re^{A\times p}$, and $E \in \Re^{n\times p}$ are the score, loading and residual matrices of $\mathbf{x}$; $Q \in \Re^{A\times l}$ and $F \in \Re^{n\times l}$ are the loading and residual matrices of $\mathbf{y}$; $W \in \Re^{A\times p}$ is the weight matrix; and A is the number of principal components retained. PLS is not only to establish the maximum variance of the second variables, $\mathbf{x}$, but also to maximize the variability of the primary variables, $\mathbf{y}$, explained by the correlation by $\mathbf{x}$ and $\mathbf{y}$. When the original variables are highly correlated, redundant, noisy and of high dimensionality, the orthogonal scores can be obtained through decompositions of $\mathbf{x}$, and $\mathbf{y}$, which would contain sufficient information on $\mathbf{x}$ and predictive information. In other words, it removes the correlation, noise, etc, between the original variables by projection and reduction. PLS models are more stable than the models built upon the original variables, since the regression is done on the scores instead of original variables.

2.4. Procedures of the proposed sensor modeling

As above mentioned, it is difficult to determine error limit and sample scale in traditional distance-based measures. Therefore, a more general similarity measure based on Gaussian function is proposed in this work. In selecting relevant samples to the query data, confidence value is used to indicate the degree of similarity. Detail modelling procedures of the proposed soft sensor are depicted as follows:

Step 1, Collect and standardize training data and test data.

Step 2, Search relevant samples to construct a similar set *Z* in the database based on Gaussian function-based similarity criterion. Set value of the Gaussian function-based similarity criterion, which is within 0 and 1. Usually it is larger than 0.9.

Step 3, Sample *N* data points from *Z* as data set Z_1. Repeat the procedure for *K* times and obtain *K* resample data set: Z_1, …, Z_K.

Step 4, Build K separate models on these resample data sets.

Step 5, Give predictions by combing the *K* models.

3. Case study

3.1. A nonlinear system model

In this section, a mathematical model of a nonlinear system (Zeng et al., 2012) is used to evaluate the proposed soft sensor. The details of the nonlinear system are given as follows.

$$y(t) = \frac{x_1(t)}{1+0.5\sin(x_2(t))} + \frac{x_2(t)}{1+0.5\sin(x_1(t))} + e(t) \tag{5}$$

$$x_1(t+1) = \left(\frac{x_1(t)}{1+x_1^2(t)} + 1\right) + \sin(x_2(t)) \tag{6}$$

$$x_2(t+1) = x_2(t)\cos(x_2(t)) + \exp\left(-\frac{x_1^2(t)+x_2^2(t)}{8}\right) \times x_1(t) + \frac{u^3(t)}{1+u^2(t)+0.5\cos(x_1(t))+x_2(t)} \tag{7}$$

$$u(t) = a\sin(0.5\pi t) + \sin(0.08\pi t) \tag{8}$$

Where, $x_1(t)$ and $x_2(t)$ are state variables, $y(t)$ is output variable and $u(t)$ is input variable. In the simulation, 3 operating working modes are assumed when *a* = 1,2 and 3. Under each operating mode, 400 data points are collected. Among them, 350 data points are selected as training set and 50 data points are selected as test set. The threshold value of similarity is set to 0.9.

Figure 1 gives the results of different soft sensors under 3 working modes. It shows that the JIT-bagging-based PLS soft sensors track the real curves more closely under all conditions. It means that the performance of JIT-based PLS soft sensor can be improved by using the bagging method. This is attributed to good robustness of bagging-based modelling strategy.

For quantitative comparisons, root mean square error (RMSE) is usually used as an index. Table 1 gives quantitative comparisons of different models. From the table, it can be observed that the JIT-bagging based PLS soft sensors outperform the JIT based soft sensors with different values of *a*.

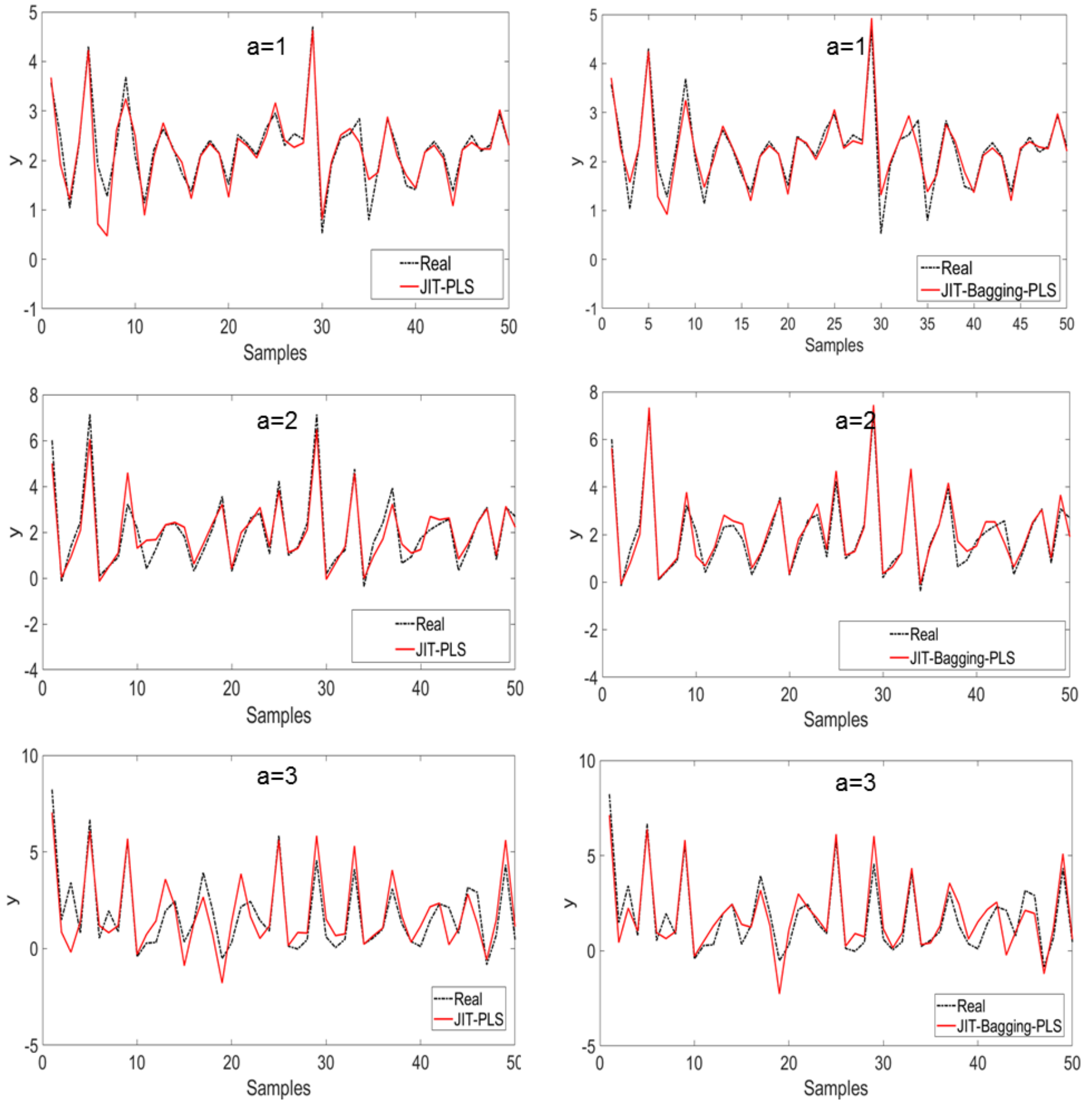

Figure 1: Comparisons of JIT and JIT-bagging based PLS soft sensors in a nonlinear system

Table 1: Comparisons of RMSE values of JIT and JIT-Bagging-based PLS soft sensors in a nonlinear system

Methods	a=1	a=2	a=3
JIT-PLS	0.2993	0.4924	1.0060
JIT-bagging-PLS	0.2554	0.3935	0.7785

3.2 Penicillin fermentation process

This case is based on the simulation software of Penicillin fermentation process (Birol et al., 2002) to produce data, and to carry out the research work of soft sensor modelling. The simulator contains a fermenter where the biological reaction takes place. The detailed description of process variables is given in (Birol et al., 2002). For different demands, the simulator provides several settings including the controller, process duration, sampling rates, etc. In this study, we set the sampling intervals is 1 hour and the fermentation period is set to 400 hours. Then, the Pensim software randomly run 10 batches under normal conditions. 7 batch data are used as training data sets, and the remaining 3 batch data are used as test data sets. There are totally 16 measure and variables in the simulation plant. The penicillin concentration is difficult to measure online and chosen as output variable. 11 variables among them are highly related to it are selected as input variables.

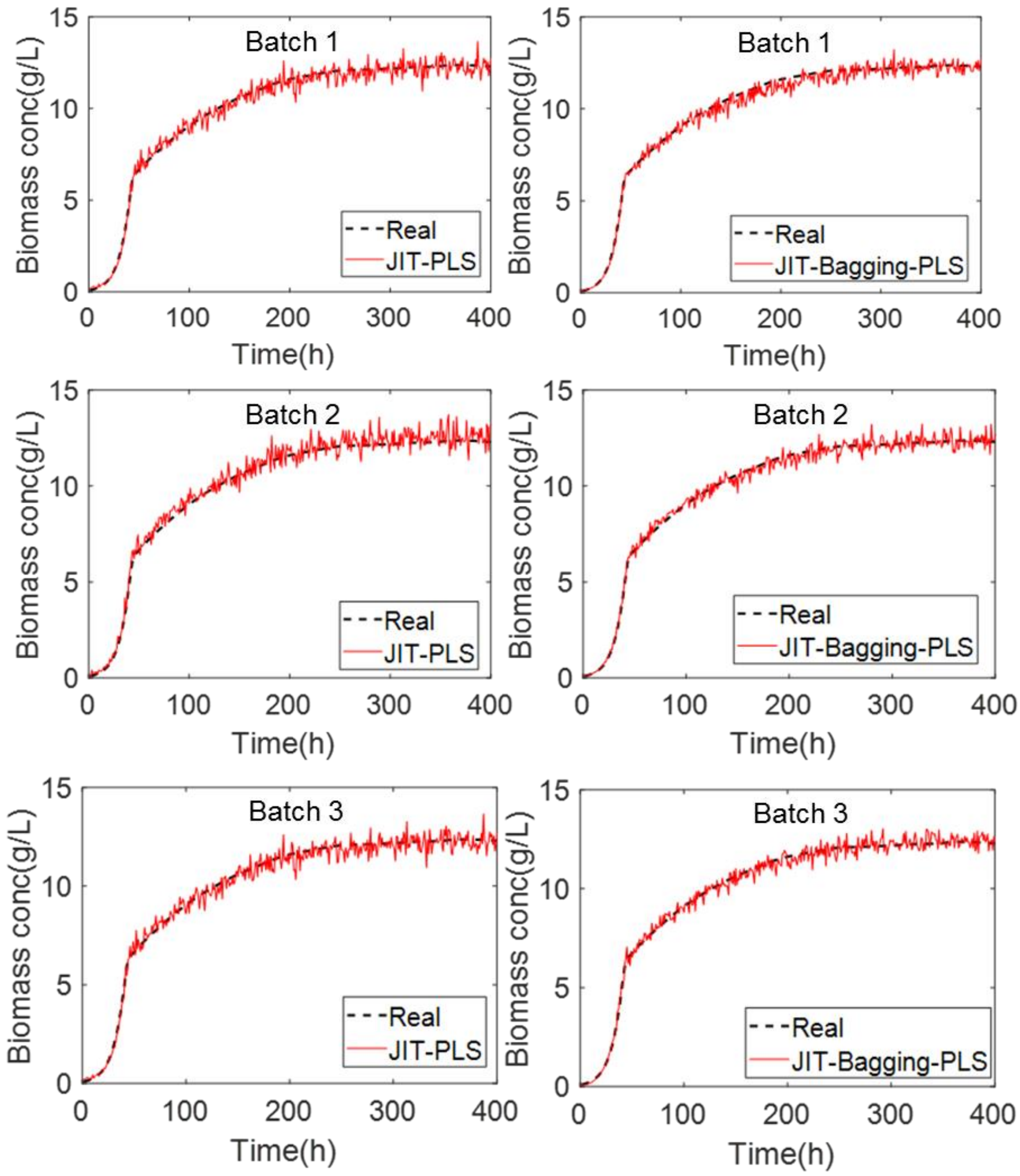

Figure 2: Comparisons of JIT and JIT-bagging based PLS soft sensors in the Penicillin fermentation process

Figure 2 gives predictions of JIT and JIT-bagging based PLS soft sensors. From Figure 2, it can be observed that the predicted curves of JIT-bagging based PLS soft sensors track the real curves more closely.

For quantitative comparisons, Table 2 gives RMSE values of different methods. It can be observed that JIT-bagging based soft sensors have smaller RSME values than those of JIT-PLS soft sensors. It can be concluded that it is an effective way to enhance the performance of JIT-based soft sensors by using the bagging method.

Table 2: Comparisons of RMSE values of JIT and JIT-Bagging-based PLS soft sensors

Methods	Batch 1	Batch 2	Batch 3
JIT-PLS	0.3687	0.4278	0.3645
JIT-bagging-PLS	0.3246	0.2794	0.2921

4 Conclusions

JITL has ability of coping with those processes with strong nonlinearity. In JITL modelling, the design of similarity measures is crucial. However, for traditional distance-based similarity measures, it is difficult to determine the number of selected relevant samples. To handle the problem, a new Gaussian function-based similarity measure is used in this work. For the novel similarity measure, the bagging method is used to build submodels for avoiding set threshold values of traditional distance-based similarity measure and good robustness. To evaluate the proposed soft sensor, a complex nonlinear mathematical model and an industrial Penicillin fermentation process were used. Results show that the proposed JIT-bagging-based soft sensor performs better than traditional JIT-based soft sensor.

Acknowledgement

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (Grant No. 61573322), the Natural Science Foundation of Jiangsu Province of China (grant no. BK 20160540) and the Natural Science Research Foundation of Higher Education of Jiangsu Province (17KJB510008).

References

Abonyi, J., Farsang, B., Kulcsar, T., 2015, Data-driven development and maintenance of soft-sensors. In IEEE International Symposium on Applied Machine Intelligence and Informatics, 239-244.

Birol, G., Ündey, C., Çinar, A., 2002, A modular simulation package for fed-batch fermentation: penicillin production. Computers & Chemical Engineering, 26, 1553-1565.

Chen, T., Ren, J., 2009, Bagging for Gaussian process regression. Neurocomputing, 72, 1605-1610.

Chen, X., Chen, X., She, J., Wu, M., 2017, A hybrid just-in-time soft sensor for carbon efficiency of iron ore sintering process based on feature extraction of cross-sectional frames at discharge end. Journal of Process Control, 54, 14-24.

Höskuldsson, A., 1988, PLS regression methods, Journal of Chemometrics, 2, 211-228.

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

Mei, C., Chen, Y., Jiang, H., Ding, Y., Chen, X., Liu, G., 2017, Just-in-time modeling with a combination of input and output similarity criterions for soft senor modeling in fermentation processes, Chemical Engineering Transaction, 61, 1045-1050.

Mei, C., Chen, Y., Zhang, H., Chen, X., Liu, G., 2017, Development of a multi-model strategy based soft sensor using gaussian process regression and principal component analysis in fermentation processes, Chemical Engineering Transaction, 61, 385-390.

Saptoro, A., 2014, State of the art in the development of adaptive soft sensors based on just-in-time Models, Procedia Chemistry, 9, 226-234.

Souza, F. A. A., Rui, A., Mendes, J., 2016, Review of soft sensor methods for regression applications, Chemometrics & Intelligent Laboratory Systems, 152, 69-79.

Wang, Z. X., He, Q. P., Wang, J., 2015, Comparison of variable selection methods for PLS-based soft sensor modeling. Journal of Process Control, 26, 56-72.

Zeng, J., Wang, J., Guo, J. Y., 2012, Local multi-model method based on similarity of vector, Application Research of Computers, 29, 1631-1634.