An Improved JITL Method for Soft Sensing of Multimodal Industrial Processes for Search Efficiency

Yushan Chen*, Yuan Lin, Tianyu Zheng
Hangzhou Dianzi University, Hangzhou, Zhejiang, China 310018

*Corresponding author: yushanchen666@hdu.edu.cn

Abstract. In the industrial process, due to product change, working condition switch, or controller adjustment, process data often presents multimodal characteristics. Data-driven approaches are often based on single-modal assumptions, which may fail to describe process characteristics. The traditional just-in-time learning (JITL) method can continuously update the model to describe the multimodal data, but it takes much time and cannot meet the real-time requirements. In this paper, an improved JITL method is proposed to find similar samples quickly. The new samples are divided into the main category first, and then find the similar samples to improve the search efficiency. The effectiveness of the method is proved by a case of an industrial soft sensor case combined with partial least squares (PLS). Compared with the basic JITL, the root mean square error (RMSE) of the proposed method is reduced by 0.09, and the running speed is increased by 8.8 times.

Keywords: Soft sensor, just in time learning, multimode, partial least squares, data-driven.

1. Introduction
In the actual industrial process, pursuing product quality improvement is a long-term and industrially valuable task. However, many key process variables are challenging to obtain due to the cost of the device or the complexity of the environment. With artificial intelligence and data storage technology development, the soft sensor has attracted more and more attention. Data-driven soft sensor methods have many attractive properties: (1) They offer a low-cost alternative to expensive hardware sensors; (2) They allow real-time estimation of data, overcoming the time delays introduced by slow hardware sensors, thus improving the performance of the control algorithms; (3) They play an indispensable role in quality control.

In the final decades, data-based soft sensor modeling methods have already been researched intensively, such as support vector machines (SVM) [1], artificial neural networks (ANN) [2], partial least squares (PLS) [3]. Defined as a convex quadratic optimization problem, SVM benefits from low computational costs and accessible optimization options. Nevertheless, when large-scale samples are input, the construction of the model is challenging to implement. ANN builds models by establishing relationships between data and adjusting various network parameters. It has been demonstrated that it is particularly appropriate for creating delicate sensors and has been broadly utilized. However, ANN still suffers from an uncontrolled convergence speed and local optima.
PLS is a multivariate statistical data analysis method. Since the proposed PLS has been recognized, and it became a popular method in the soft sensor. The reason for the popularity resides in the fact that the method can manage massive dimensional co-linear data and the fact that the resulting model considers the covariance between multiple measurement variables X and one test variable y. Besides, PLS has high efficiency and better interpretation, which is more suitable for industrial occasions with high requirements on real-time performance.

However, due to equipment wear, process load, or process adjustment, industrial processes often show multimode characteristics when switching between different modes, whose characteristics are significantly changed along with the operating conditions. Traditional data-driven models are based on the assumption that data follows a specific distribution, which cannot be used to handle multimodal problems.

In recent years, the multimodal problem has been widely concerned. However, this method requires much prior knowledge, which is not satisfied in most scenarios. More and more attention has been paid to adaptive methods in recent years, and the just-in-time-learning (JITL) is a typical adaptive method. JITL method has been lately applied in-process monitoring and soft sensor development. In JITL modeling, a local model is built from past data around a query point only when an estimated value is requested. Unlike conventional approaches, the JITL-based approach exhibits a local model structure that can be considered global models. When the global model is built offline, the local model is built online based on JITL. Another advantage of JITL is that it can easily handle nonlinear without much prior knowledge. However, in online applications, JITL will spend plenty of time searching for the nearest samples, which reduces the efficiency of the algorithm.

In this paper, a multimode soft sensor method based on just-in-time learning and PLS is proposed. The remainder of this article is structured as follows. The second section briefly introduces the K-means clustering algorithm and PLS. The methods mentioned in this article are introduced in Section 3. An industrial example verifies the effectiveness of the algorithm in the next section. Finally, some conclusions were drawn.

2. Related works

2.1. Just-in-time learning

Just-in-time learning (JITL) is a typical adaptive method, which has been proposed to solve multimodal problems. The critical point of JITL is to develop a local model for every query sample. This approach is based on the idea that similar inputs produce similar output. Similar samples are selected as training data for the local model. Since JITL can build new models for every new sample, the multimodal problems can be handled here. Generally speaking, in a JITL scheme to build an online model, there are three main steps.

Step 1: When a query sample comes, relevant samples that are located in a neighbor region around the query sample are selected as training samples for local modeling.

Step 2: A local model is built and trained based on these relevant samples.

Step 3: Obtain the output based on the new model, then discard the local model and build a new local model in the same way when a new query sample arrives.

On the other hand, JITL can be regarded as building a local linearization model to deal with nonlinear problems. Besides, since the JITL requires frequent model updates, there is a high requirement for the computational complexity of the model. In general, the linear models are widely applied for local modeling, which ensures the efficiency of modeling, and the fragile nonlinear characteristics in the data will not have a significant impact on the linear model.

2.2. PLS

When there are multiple correlations between independent variables, or when the sample has fewer dimensions than the variable, the traditional regression analysis method is no longer applicable. However, these problems can be solved by partial least squares (PLS). The purpose is to forecast or
evaluate many dependent variables from a set of predictors or independent variables. This forecast is accomplished by extracting from the indicators a set of orthogonal variables called latent factors, which have the most excellent prescient power. The mathematical representation of the PLS model is expressed as:

\[ X = TP^T + E \]  
\[ Y = UQ^T + F \]

Where \( X \) is a \( n \times m \) matrix of predictors and \( Y \) is a \( n \times p \) matrix of responses. \( T \) and \( U \) are two \( n \times l \) matrices that are projections of \( X \) (scores, components, or the factor matrix) and projections of \( Y \) (scores) \( P \), \( Q \) are \( m \times l \), and \( p \times l \) orthogonal loading matrices. Matrices \( E \) and \( F \) are the error terms, which are assumed to be an independent and identical normal distribution. Decompositions of \( X \) and \( Y \) are made to maximize the covariance of \( T \) and \( U \).

3. method

3.1. Motivation

The traditional methods are based on the assumption that the distribution of data is stationary and thus, we can generalize from the sample to the whole process. Industrial processes show typical multimode characteristics due to the change of process characteristics, such as catalyst deactivation, equipment aging, and change of raw materials, etc. Therefore, the traditional methods are not sufficient for these industrial processes. The first step of multimodal methods is mode recognition, which determines the accuracy of the proposed model. The wrong judgment may lead to significant errors. Compared with multimodal methods, adaptive methods are more suitable for dealing with this kind of problem since they can automatically adjust and update the model according to new samples. JITL is a typical adaptive method that can track the current running state by constructing a local model online. The prediction performance of JITL is mainly dependent upon the samples that are selected for local modeling. When the amount of data is large, it will take much time to find similar data, which may not meet the real-time requirement for online applications. Therefore, it is necessary to improve the search efficiency and establish the model quickly to ensure the speed of the adaptive method.

3.2. An improved JITL soft sensor method

Generally speaking, the search efficiency of the traditional JITL is relatively low. This section divides the original historical data into different clusters by clustering algorithm and then marks the new samples as corresponding clusters. It is done for the following two reasons. We can judge the cluster to which the new sample belongs according to the similarity after clustering and then search similar samples in this cluster. It helps to improve the efficiency of the search. Secondly, new samples can be quickly divided into corresponding clusters to supplement historical data, and then the soft sensor based on the PLS model can be established.

The key of this method is determining the number of clusters and how to deal with the influence of the addition of new samples on the original data distribution. Here the Silhouette Coefficient is adopted to evaluate the clustering effect and data analysis [4]. The specific steps are summarized as follows.

Step 1: Clustering samples. The dataset is where \( N \) is the number of samples and \( M \) is the data dimension. It is divided into \( C \) clusters by clustering algorithm by K-means. The divided data is denoted as

\[ X = \{X_1 = \{x_{11}^1,x_{12}^1,...,x_{1n}^1\}, X_2 = \{x_{21}^2,x_{22}^2,...,x_{2n}^2\},...,X_c = \{x_{c1}^c,x_{c2}^c,...,x_{cn}^c\}\} \]
Where $X_i$ represents the $i$th cluster and $x_i^n$ represents the $n$th sample of the $i$th cluster. The clustering effect is evaluated by the Silhouette Coefficient, which is expressed as follows:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \text{ for } a_i > b_i$$

(3)

where $S_i$ represents silhouette coefficient, $a_i$ is the average distance from the $i$th object to all other objects in a cluster, and $b_i$ is the average distance from the $i$th object to any cluster not containing the object.

**Step 2:** Searching for similar samples. Suppose $x_{\text{new}}$ as the new sample, determine which cluster it belongs to, and then search $K$ similar samples in this cluster. The similarity measure for sample $x_i$ and sample $x_j$ is as follows

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik}^k - x_{jk}^k)^2}$$

(4)

**Step 3:** Building a local model. Suppose $X_{\text{local}}$ and $Y_{\text{local}}$ are the data selected in Step 2. By maximizing the variance of predictors and response variables. The first latent variable can be obtained:

$$\max_{||c||=1,||w||=1} \text{cov}(X_{\text{local}} w, Y_{\text{local}} c)$$

Where the loading weights $c_i$, $w_i$ are the first left and right singular vectors of $X_{\text{local}}^T Y_{\text{local}}$. Furthermore, the first latent variable is defined as: $t_i = X_{\text{local}} w_i$. The $X_{\text{local}}$ and $Y_{\text{local}}$ are deflated by using the least square estimation. The residual errors can be computed as follows:

$$X_{\text{local}} = t_i a_i + X_{\text{local}}^l$$

(5)

$$Y_{\text{local}} = t_i b_i + Y_{\text{local}}^l$$

(6)

The next components $t_2, \ldots, t_h$ can be similarly computed as formula (1) through iterating the residual errors $X_{\text{local}}^l$ and $Y_{\text{local}}^l$. This deflation pattern can ensure the loading weights $W_{p \times H}$ and components $T_{n \times H}$ are orthogonal. Hence, the estimation of $\beta$ within the PLS calculation can be expressed:

$$\hat{\beta} = (W^T X_{\text{local}}^T W)^{-1} W^T X_{\text{local}}^T Y_{\text{local}}$$

(7)

**Step 4:** Updating clustering centers. In the first step, the new sample has been labeled and placed in the corresponding cluster. Next, recalculate the Silhouette Coefficient. If the new sample label is significantly different from the original clustering effect, it means that the addition of the new sample has an impact on the original data distribution, which will cause the Silhouette Coefficient to decrease. When the Silhouette Coefficient is less than a certain threshold, it means that the original clustering effect cannot sufficiently represent the existing data distribution, and we need to cluster the filled data
again. Perform K-means clustering again, using the current cluster center as the initialization center. Finally, return to Step 2. The flowchart of the proposed method is shown in Fig. 1.

Fig. 1 The flowchart of the proposed method

4. Study case
The data in this article comes from a three-phase flow facility at Cranfield University. Furthermore, this equipment aims to study the impact of multiphase flow supply on small industrial equipment. The process mixes oil, water, and gas and separates them through many devices and operations. When the reaction is over, they are processed by the equipment and returned to the tank. In this article, the sampling interval for all data is 1 second. The pressure of the three-phase separator is always 0.1 MPa. The variables used in this case are listed in Table 1. Here we predict the pressure in the top of the riser (Variable 3) through other variables.

| Variable nr | Location                                    | Measured magnitude                     | Unit      |
|-------------|---------------------------------------------|----------------------------------------|-----------|
| 1           | PT312                                       | Air delivery pressure                  | MPa       |
| 2           | PT401                                       | Pressure in the bottom of the riser    | MPa       |
| 3           | PT408                                       | Pressure in top of the riser           | MPa       |
| 4           | PT403                                       | Pressure in top separator              | MPa       |
| 5           | PT501                                       | Pressure in 3 phase separators         | MPa       |
| 6           | PT408                                       | Diff. pressure (PT401-PT408)           | MPa       |
| 7           | PT403                                       | Differential pressure over VC404       | MPa       |
| 8           | FT305                                       | Flow rate input air                    | Sm$^{3}$/s|
| 9           | FT104                                       | Flow rate input water                  | kg/s      |
| 10          | FT407                                       | Flow rate top riser                    | kg/s      |
| 11          | LI405                                       | Level top separator                    | m         |
| 12          | FT406                                       | Flow rate top separator output         | kg/s      |
| 13          | FT407                                       | Density top riser                      | kg/m$^{3}$|
| 14          | FT406                                       | Density top separator output           | kg/m$^{3}$|
| 15          | FT104                                       | Density water input                    | kg/m$^{3}$|
| 16          | FT407                                       | Temperature top riser                  | °C        |
| 17          | FT406                                       | Temperature top separator output       | °C        |
| 18          | FT104                                       | Temperature water input                | °C        |
| 19          | LI504                                       | Level gas-liquid 3 phase separator     | %         |
| 20          | VC501                                       | Position of valve VC501                | %         |
| 21          | VC302                                       | Position of valve VC302                | %         |
| 22          | VC101                                       | Position of valve VC101                | %         |
| 23          | PO1                                         | Water pump current                     | A         |

By contrast, we use the primary JITL method and the global model based on PLS to prove the superiority and effectiveness of the proposed method. Firstly, 70% of the collected data were selected as the training set and the remaining 30% as the verification set. The global model is built with PLS,
and its effect is shown in Fig. 2. As can be seen from Fig. 2, for multimodal industrial processes, the global model cannot describe all the situations well, so the error is large. When the sample size is relatively small, the global model is not enough to describe these details, leading to the failure of prediction. By looking for N samples that are most similar to the new sample, JITL establishes a local model to accurately depict the model to which the current sample belongs and achieves a more accurate effect, which is shown in Fig. 3. Here we denote the N equals 100. The disadvantage of the basis JITL method is that it takes a long time to find similar samples, which is challenging to meet the real-time requirements of online applications. Utilizing clustering, first quickly find the new sample belongs to the primary category, and then find the most similar sample through JITL, which significantly reduces the search time of similar samples. The results of the proposed method are shown in Fig. 4. The accuracy and time expenditure by the three methods are listed in Table 2.

| Method               | root-mean-square error | time expenditure |
|----------------------|------------------------|------------------|
| PLS                  | 0.765                  | 0.039s           |
| JITL                 | 0.138                  | 1101.893s        |
| The proposed method  | 0.048                  | 125.297s         |
5. Conclusion
In this paper, a multimode soft sensor method based on just-in-time learning and PLS is proposed. Firstly, samples are divided into different clusters by clustering to overcome the inefficient sample search problem of JITL. When a new sample enters, first roughly judge which cluster it belongs to through cluster analysis results, then search the nearest neighbor samples in the corresponding cluster, which improves the algorithm's efficiency. It is worth noting that the division of clusters here is only used as a reference for JITL to find neighbor samples and has little effect on the judgment of the final modal. Afterward, PLS is applied to model a local soft sensor model with the nearest samples. Based on the obtained partial model, the new sample is analyzed to obtain the value of the variable to be measured. An industrial case verifies the effectiveness of the proposed method. Benefit from the above steps, the proposed method reduces the root mean square error (RMSE) of basic PLS and basic JITL by 0.717 and 0.09. The running speed of the proposed method is 8.8 times faster than the basic JITL. The disadvantage of this method is that the selection of the number of clusters and the parameters of iteration conditions still depends on experience adjustment, which can be improved by studying the adaptive way to determine the parameters.

References
[1] Jianlin WANG, et al."Online Estimation of Biomass in Fermentation Process Using Support Vector Machine 1 1 Supported by the National Natural Science Foundation of China (No.20476007)." Chinese Journal of Chemical Engineering 14.3(2006).
[2] Pisa, I.; Santin, I.; Vicario, J.L.; Morell, A.; Vilanova, R. ANN-Based Soft Sensor to Predict Effluent Violations in Wastewater Treatment Plants. Sensors 2019, 19, 1280.
[3] Ren, M.; Song, Y.; Chu, W. An Improved Locally Weighted PLS Based on Particle Swarm Optimization for Industrial Soft Sensor Modeling. Sensors 2019, 19, 4099.
[4] P. Anitha, Malini M. Patil. RFM model for customer purchase behavior using K-Means algorithm [J]. Journal of King Saud University - Computer and Information Sciences, 2019.