IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 7, 2020, accepted February 11, 2020, date of publication February 18, 2020, date of current version February 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974783

Soft Sensor Modeling for Unobserved Multimode Nonlinear Processes Based on Modified Kernel Partial Least Squares With Latent Factor Clustering

XIAOGANG DENG, (Member, IEEE), YONGXUAN CHEN, PING WANG, AND YUPING CAO
College of Control Science and Engineering, China University of Petroleum, Qingdao 266580, China

Corresponding author: Xiaogang Deng (dengxiaogang@upc.edu.cn)

This work was supported in part by the Research Fund for the Creative Research Team of Young Scholars at Universities in Shandong Province of China under Grant 2019KJN019, in part by the National Natural Science Foundation of China under Grant 61403418, and in part by the Key Programme of Research and Development of Shandong Province of China under Grant 2018GGX101025.

ABSTRACT To cope with the soft sensor modeling of unobserved multimode nonlinear processes, this paper proposes a modified kernel partial least squares (KPLS) by integrating latent factor clustering (LFC), called LFC-KPLS. In the proposed method, the process data are first divided into several batches orderly, and then projected onto the latent space by using the nonlinear functional expansion technology. In the latent space, partial least squares method is applied to compute the regression coefficients between the input variables and output variable of each batch. These regression coefficients, called the latent factors, can describe the functional relationships in the unobserved multimode data. Therefore, the latent factors are used for mode clustering so that the process data with similar functional relations can be clustered in one mode together. For each mode, the nonlinear soft sensor is established based on KPLS. To assign the mode of the online query sample, a mode identification strategy based on Bayesian inference is designed for the soft sensor online prediction. Finally, two cases studies are adopted to validate the proposed method.

INDEX TERMS Soft sensor, nonlinearity, unobserved multimode, kernel partial least squares, latent factor clustering.

I. INTRODUCTION

Realtime monitoring and control of quality variables play a vital role in the complicated industrial processes [1], [2]. However, some important quality variables, such as the freezing point of diesel oil in the refinery fractionators, and the concentration of reactant production in the chemical reactors, are often difficult to measure directly through the hardware sensors. Even if the online quality analyzers are installed in some units, they have the shortcomings of high price and frequent maintenance. In most cases, these quality variables are obtained only by the offline laboratory analyses. It has the disadvantage of a significant time interval (often from 4 to 12 hours) so that the realtime control on the quality variable is not practicable. Therefore, the soft sensor technique, which builds a virtual software sensor by mining the mathematical relationship between the easy-to-measure process variables and the difficult-to-measure quality variables, has been extensively implemented in the industrial plants. The present soft sensor modeling methods can be divided into two categories: model-based and data-based. The former builds the soft sensor based on the accurate physical and chemical mechanisms, which are often difficult to obtain in many complicated processes. The latter performs the data mining in the historical running data without the use of accurate mechanical models, which is more popular in recent years because of the available abundant process data [3].

Some typical data-driven soft sensor modeling methods include partial least squares (PLS), Gaussian mixture regression (GMR) and extreme machine learning (ELM) [4]–[7], etc. Among these methods, PLS has gained great attention in the soft sensor field because of its effectiveness. Sharmin *et al.* [8], Zheng and Funatsu [9], Zheng and Song [10] discussed the successful applications of PLS in

The associate editor coordinating the review of this manuscript and approving it for publication was Ming Luo.

different industrial units. However, the basic PLS models are intrinsically linear while many real industrial processes are with strong nonlinearity. To deal with the nonlinear soft sensor modeling issue, a number of nonlinear extensions of PLS have been proposed. In the early studies, quadratic PLS [11] and neural network PLS [12] were developed, which utilize the quadratic polynomial and neural networks to model the inner nonlinear relation, respectively. Later, Bang *et al.* [13] applied the fuzzy inference system to assist the nonlinear PLS modeling. Considering that Gaussian process regression (GPR) has the powerful nonlinear fitting ability, Liu *et al.* [14] designed a GPR-PLS method and tested it on a real wasterwater treatment process (WWTP).

In recent years, kernel PLS (KPLS) has been developed as one effective nonlinear PLS method [15]. Different to other nonlinear PLS versions, KPLS avoids the explicit nonlinear optimization via the kernel trick. Because of the simpleness and effectiveness, KPLS has attracted enough attention in the nonlinear soft sensor field. Zhang *et al.* [16] applied KPLS to an industrial oil refinery fatory and demonstrated its performance advantage over the linear PLS. To deal with the batch process soft sensor modeling, Wang *et al.* [17] developed a new multiway KPLS method, which uses the feature vector selection to reduce the number of kernel vectors for low computation loads. The KPLS modeling is based on the sufficient training data. However, in some new industrial process, the training data is often very limited. To cope with this problem, Chu *et al.* [18] combined transfer learning idea with KPLS modeling for an improved joint-Y KPLS (JYKPLS) method, which transfers the rich information from similar old processes to the new process model. To deal with the collinear characteristic and enhance the model prediction performance, Tang *et al.* [19] built a selective ensemble KPLS (SENKPLS) method, where one double-layer genetic algorithm is employed to optimize the parameters of sub-models.

Apart from the process nonlinearity, the multimode operation is another common situation in industrial processes. Due to the market demand changes, the process disturbances, and the changeover of catalyst, etc., the process operation modes are often changing. In this case, the single global nonlinear soft sensor may not provide the best predictions. Therefore, how to design the multimode nonlinear soft sensor model is a valuable problem deserving deep discussions. The researchers have developed many solutions for this problem, which can be divided into two categories. One category develops the local models by appling the just-in-time learning (JITL) strategy, while another category builds mutiple models by the divide-and-rule (DAR) strategy. The JITL method is also called lazy learning method because it only collects the historical data as the training dataset and does not need the offline model training. When an online query sample is available, JITL constructs a local model by searching the most relevant samples in the offline training dataset. For dealing with the soft sensor modeling of multiphase batch process, Jin *et al.* [20] developed a JITL KPLS method, where a hybrid similarity including the sample similarity and phase similarity is used to select the relevant training samples and then the local KPLS soft sensor is built for each query sample. To consider both the modeling accuracy and the efficiency, Chen *et al.* [21] proposed a JITL method with selective updating based on approximated linearity dependence (ALD) and applied it to the soft sensor of roller kiln temperature. The DAR method firstly identifies the process modes by applying the data clustering technologies, and then builds multiple local soft sensors corresponding to the different clusters. At the online prediction procedure, the new sample is assigned to one certain mode based on some similarity index such as the distance similarity. The commonly used data clustering methods include the K-means method, and the fuzzy C means (FCM) method. Zhao *et al.* [22] proposed an improved K-means based ensemble KPLS method. Yuan *et al.* [23] utilized FCM to obtain different local clusters and built the locally weighted PCR model for the query sample. Gholami *et al.* [24] presented a soft sensor by combining the FCM clustering with the support vector regression. Wang *et al.* [25] designed a nonlinear multimode process soft sensor, which applies a self organizing framework to build the multimode KPLS and applies the conditional probability density analysis to identify the sample mode. To overview the present multimode soft sensor methods, both the JITL and the DAR strategies can handle the soft sensor modeling for many complicate processes including nonlinear and/or multimode processes effectively. However, JITL involves a larger computation loads because the local model is online built for each query. Therefore, this paper focuses on the DAR based soft sensor modeling method.

Although the present DAR based soft sensor methods have achieved the significant success in the nonlinear multimode processes. However, there are still some challenging problems worthy of extensive study. One important problem is the soft sensor modeling for the unobserved multimode nonlinear processes. Almost all the past works focus on the observed multimode processes, where there is often an underlying assumption that the different operating modes can be distinguished by investigating the magnitudes of the measured variables, that is the input variables of the soft sensor. However, the unobserved multimode process, firstly discussed by Liu [26], is a different kind of multimode process where the operating mode switching can not be directly measured. For example, in the refinery units, when the crude oil types or properties change, the process variables are kept at the similar operation points, but the product quality variables appears with multiple modes. Another example is about the reactor. In some chemical reactors, the catalyst activation energy degrades as time goes, which also brings the unobserved multimode data. In these cases, different process modes come with the similar measured variables, but the inner mechanism between predictors and quality-related variables has changed. For the unobserved multimode proceess, it is difficult to perform the mode division by the distance similarity based clustering method.

According to the above discussions, we propose a new soft sensor for unobserved multimode nonlinear process based on a modified KPLS by integrating latent factor clustering, called LFC-KPLS. The contribution of the proposed method is three fold. First, a soft sensor modeling framework is designed for the unobserved multimode nonlinear processes. To our best knowledge, we are the first to discuss the soft sensor modeling method of the unobserved multimode nonlinear processes. Second, a latent factor clustering is designed based on the functional extension technique. Different to the traditional clustering methods, LFC clusters the multimode data by measuring the similarity of nonlinear data relationship, but not the similarity of sample distance. Third, a mode identification method for the online query sample is proposed by applying the Bayesian inference to compute the posterior probability.

The remainder of this paper is organized as follows. Section II overviews the preliminaries including the KPLS and the FCM. Then the proposed methodology is introduced in the Section III. Section IV gives two case studies of one numerical system and the simulated continuous stirred tank reactor. The last section offers some conclusions.

II. PRELIMINARIES

A. KERNEL PARTIAL LEAST SQUARES

Kernel partial least squares (KPLS) combines the kernel technique with PLS for a nonlinear regression model [15]. For the given input matrix $\boldsymbol{X} \in \boldsymbol{R}^{n\times m}$ and the output vector $\boldsymbol{y} \in \boldsymbol{R}^{n}$ with n samples, KPLS first projects the nonlinear original input data $\boldsymbol{X}$ into the linear latent space $\psi(\boldsymbol{X})$ and then performs the linear PLS modeling between $\psi(\boldsymbol{X})$ and $\boldsymbol{y}$, which brings a PLS regression model as

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e} = \psi(\boldsymbol{X})\boldsymbol{b} + \boldsymbol{e} \tag{1}$$

where $\psi(.)$ is the assumed nonlinear transformation, $\boldsymbol{b}$ is the regression coefficient, $\hat{\boldsymbol{y}} = \psi(\boldsymbol{X})\boldsymbol{b}$ is the output prediction value, while $\boldsymbol{e}$ is the prediction error vector.

As the nonlinear mapping function $\psi(.)$ is usually unknown and can not be explicitly expressed, Eq. (1) can not be directly used for the output prediction. To deal with this problem, we expand the regression coefficient vector by the input data matrix as

$$\boldsymbol{b} = \psi(\boldsymbol{X})^T\boldsymbol{\beta} \tag{2}$$

Combining the Eqs. (1) and (2) leads to a nonlinear PLS model based on the kernel matrix, which is given as

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{\beta} + \boldsymbol{e} \tag{3}$$

where $\boldsymbol{K} = \psi(\boldsymbol{X})\psi(\boldsymbol{X})^T$ is the kernel matrix with its (i, j)-th element k_{ij} defined by

$$k_{ij} = \psi(\boldsymbol{x}_i)^T\psi(\boldsymbol{x}_j) = ker(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{4}$$

where $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ represent the i-th and j-th vector in the matrix $\boldsymbol{X}$, respectively, and $ker(\cdot,\cdot)$ denotes kernel function computation. The commonly used kernel function is the Gaussian kernel function, expressed by [17]

$$ker(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}) \tag{5}$$

where σ is the kernel width parameter.

Algorithm 1 The Solution Procedure of KPLS Model

```
1: Given the input matrix $\boldsymbol{X}$, the output vector $\boldsymbol{y}$, and the retained kernel score vector number $L$.
2: Randomly initialize $\boldsymbol{u}$ (usually, $\boldsymbol{u}$ can be set to the output variable $\boldsymbol{y}$).
3: Compute the input score vector $\boldsymbol{t} = \boldsymbol{K}\boldsymbol{u}$ and normalize it by $\boldsymbol{t}/||\boldsymbol{t}||$.
4: Obtain the weight coefficient $c = \boldsymbol{y}^T\boldsymbol{t}$.
5: Calculate the output score vector $\boldsymbol{u} = \boldsymbol{y}c$.
6: Repeat the steps 3 to 6 until convergence.
7: Deflate the kernel matrix and the output vector as $\boldsymbol{K} = (\boldsymbol{I} - \boldsymbol{t}\boldsymbol{t}^T)\boldsymbol{K}(\boldsymbol{I} - \boldsymbol{t}\boldsymbol{t}^T), \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{t}\boldsymbol{t}^T\boldsymbol{y}$.
8: Go back to step 3 until all the $L$ score vectors $\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_L]$ and $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_L]$ are found.
```

The solution of KPLS can be done by the classic NIPALS algorithm [15], [16], listed in Algorithm 1. Based on the KPLS algorithm, the regression coefficient vector $\boldsymbol{\beta}$ can be established by

$$\boldsymbol{\beta} = \boldsymbol{U}(\boldsymbol{T}^T\boldsymbol{K}\boldsymbol{U})^T\boldsymbol{T}^T\boldsymbol{y} \tag{6}$$

where $\boldsymbol{T}$ is the input score matrix and $\boldsymbol{U}$ is the output score matrix.

For the test input vector $\boldsymbol{x}_t$, its corresponding output prediction is given by

$$\hat{y}_t = \boldsymbol{k}_t\boldsymbol{\beta} \tag{7}$$

where $\boldsymbol{k}_t = (\psi(\boldsymbol{X})\psi(\boldsymbol{x}_t))^T$ is the kernel vector corresponding to the test vector.

B. FUZZY C-MEANS CLUSTERING

Fuzzy C-means (FCM) is a well-known data clustering method and has been widely used for unsupervised data pattern recognition [23], [24]. It groups all the training data into C clusters with varying membership degrees. FCM is be viewed as the improvement of the traditional K-means clustering. Different to the K-means method where each data point only belongs to one cluster, FCM assigns each datapoint to all clusters with different membership degrees. It has been demonstrated that FCM outperforms the basic K-means method in many cases.

Given the sample set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i \in \boldsymbol{R}^m$ is one sample, FCM is to find the cluster centroid $\boldsymbol{o}_1, \boldsymbol{o}_2, \cdots, \boldsymbol{o}_C$ based on the following optimization objective

$$\min J = \min \sum_{i=1}^{n}\sum_{j=1}^{C}(\mu_ij)^r||\boldsymbol{x}_i - \boldsymbol{o}_j||^2 \tag{8}$$

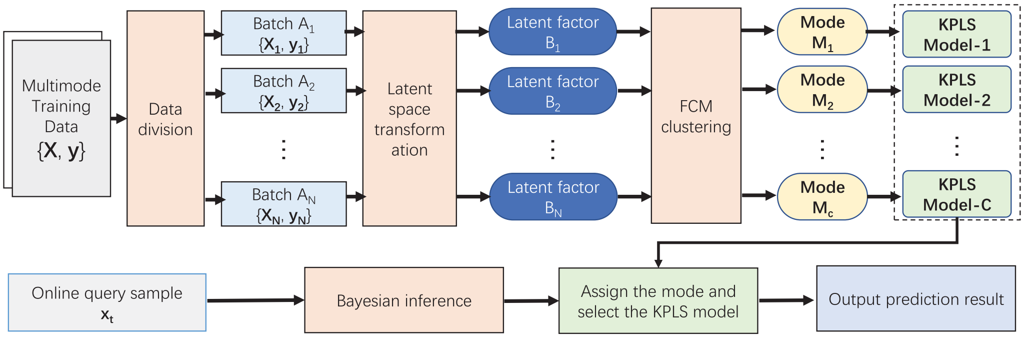


FIGURE 1. The schematic of the proposed method.

where μ_{ij} is the membership degree of $\boldsymbol{x_i}$ belonging to the cluster $\boldsymbol{o}_j$, and r is the fuzziness exponent usually set to be one real number greater than 1.

To solve this optimization function, an iteration procedure is applied, which is described as follows.

Algorithm 2 The Solution Procedure of FCM

```
1: Randomly initialize the cluster centers $\{\boldsymbol{o}_1, \boldsymbol{o}_2, \cdots, \boldsymbol{o}_C\}$.
2: Compute the membership matrix $\boldsymbol{U} = u_{ij}, 1 \leq i \leq n, 1 \leq j \leq C$ by the following expression
       $u_{ij} = \frac{1}{\sum_{k=1}^{C} (\frac{||\boldsymbol{x}_i - \boldsymbol{o}_j||}{||\boldsymbol{x}_i - \boldsymbol{o}_k||})^{2/(m-1)}}$    (9)
3: Update the cluster centers by the equation
       $\boldsymbol{o}_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} \boldsymbol{x}_i}{\sum_{i=1}^{n} u_{ij}^{m}}$    (10)
4: Repeat the steps 2 and 3 until convergence.
```

In this algorithm, the cluster number C is an important parameter which is needed to be pre-specified. In the cases with enough prior knowledge, it can be determined by experience. Without available knowledge, it can be set based on the data-driven methodology [27].

III. THE PROPOSED LFC-KPLS METHOD

As mentioned in the introduction section, the unobserved multimode processes data are distance indivisible. To develop a soft sensor method for the unobserved multimode process, three problems are involved. (1) How do we design a soft sensor modeling framework? (2) For the distance-indivisible data, how do we develop a cluster algorithm to recognize the different modes? (3) For the online query sample, how do we identify its mode? Aiming at these questions, we are to propose one latent factor clustering based KPLS (LFC-KPLS) method for unobserved multimode process soft sensing. Next, the details of the modeling framework, training data clustering and online query sample mode identification are introduced.

A. THE SOFT SENSOR MODELING FRAMEWORK

The whole schematic of the proposed LFC-KPLS method is displayed in the Fig. 1. During the offline modeling stage, the multimode KPLS model is built by the following steps. Firstly, the training data are divided into serval batches along the time orderly. Then, for each batch, the data are projected into the latent space by the nonlinear function expression, and the latent factors are computed to indicate the relationship between inputs and output. Thirdly, FCM is applied to the latent factors to obtain different modes, and for each mode, a KPLS model is developed as the local soft sensor submodel. In the online application stage, one online query sample is collected and its mode is identified by the Bayesian inference technology. Based on the identified mode, the corresponding KPLS model is chosen to generate the output prediction.

B. LATENT FACTOR CLUSTERING

For the unobserved multimode process, the process data from the different modes have different input-output data relationships, but may be very close in terms of the input sample distance. Therefore, the traditional FCM algorithm, depending on the distance similarity in Eq. (8), can not distinguish the data modes correctly, and it is necessary to develop a new data clustering method. The new clustering method should measure the data similarity based on the input-output data relationship, which means the regression coefficient $\boldsymbol{b}$ in Eq. (1). However, an assumed nonlinear mapping function $\psi(.)$ is applied so that it is difficult to measure the similarity of the $\boldsymbol{b}$ directly. Thus, the key point focuses on the handling of nonlinear function $\psi(.)$.

To deal with the above problem, this section proposes a new data clustering method called latent factor clustering(LFC). LFC first projects the original data onto a latent space by some explicit expanded nonlinear functions, which are used to substitute the implicit nonlinear function $\psi(.)$. Then LFC computes the input-output relationship factor in the latent space, called latent factor. Based on the latent factors, the FCM is applied to cluster the different data modes. The details are clarified as follows.

For one input variable x, a simple nonlinear functional expression is given as

$$G(x) = [g_1(x)\ g_2(x)\ \cdots] \tag{11}$$

where $g_i(x)$ represents some kind of nonlinear transformation. In this paper, we adopt six nonlinear functions including $g_1(x) = x$, $g_2(x) = \frac{1}{1+e^{-x}}$, $g_3(x) = sin(x)$, $g_4(x) = sin(\pi x)$, $g_5(x) = cos(x)$, $g_6(x) = cos(\pi x)$ [28].

For the training dataset $\boldsymbol{X} \in \boldsymbol{R}^{n\times m}$ with multimode property, it is divided into serval batches $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N$ with the same size by applying moving window technology. Each batch is denoted as $\boldsymbol{X}_i \in \boldsymbol{R}^{w\times m}$, where w is the length of data window and meets $n = Nw$. $\boldsymbol{X}_i$ can be expressed by

$$\boldsymbol{X}_i = \begin{bmatrix} x_{i,11} & x_{i,12} & \cdots & x_{i,1m} \\ x_{i,21} & x_{i,22} & \cdots & x_{i,2m} \\ \ddots & \ddots & \vdots & \ddots \\ x_{i,w1} & x_{i,w2} & \cdots & x_{i,wm} \end{bmatrix} \tag{12}$$

where $x_{i,jk}$ represents the (j, k)-th sample in the data window $\boldsymbol{X}_i$.

Before applying the nonlinear latent space transformation, the input and output vectors should be normalized for the same magnitude range by the following way.

$$\tilde{x}_{i,jk} = \frac{x_{i,jk} - x_{i,jmin}}{x_{i,jmax} - x_{i,jmin}} \tag{13}$$

$$\tilde{y}_i = \frac{y_i - y_{i,min}}{y_{i,max} - y_{i,min}} \tag{14}$$

where $x_{i,jmin}$, $x_{i,jmax}$ is the minimum and maximal values of the j-th column input variable, respectively, and $y_{i,min}$, $y_{i,max}$ is the minimum and maximal values of the output variable, respectively.

Then the corresponding latent space description is given by

$$G(\boldsymbol{X}_i) = \begin{bmatrix} G(\tilde{x}_{i,11}) & G(\tilde{x}_{i,12}) & \cdots & G(\tilde{x}_{i,1m}) \\ G(\tilde{x}_{i,21}) & G(\tilde{x}_{i,22}) & \cdots & G(\tilde{x}_{i,2m}) \\ \ddots & \ddots & \vdots & \ddots \\ G(\tilde{x}_{i,w1}) & G(\tilde{x}_{i,w2}) & \cdots & G(\tilde{x}_{i,wm}) \end{bmatrix} \tag{15}$$

The output vector corresponding to the matrix $\boldsymbol{X}_i$ is denoted as $\tilde{\boldsymbol{y}}_i$, which is the linear expression of the input matrix $G(\boldsymbol{X}_i)$ depicted by

$$\tilde{\boldsymbol{y}}_i = G(\boldsymbol{X}_i)\boldsymbol{b}_i + \boldsymbol{e}_i \tag{16}$$

To solve the above problem by the basic PLS algorithm will lead to the latent factor $\boldsymbol{b}_i$. Similar operations on all the data batches bring a series of latent factors $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N$. We further apply the FCM on these latent factors and the data clusters are obtained.

To sum up, the novelty of LFC lies in two aspects. (1) LFC clusters the data based on the input-output relationship factor, but not the original sample distance. (2) LFC provides a practicable nonlinear transformation by applying the explicit expanded nonlinear functions, which may not approximate the kernel function perfectly, but at least provides a viable solution to deal with the unknown $\psi(.)$.

C. ONLINE MODE IDENTIFICATION BASED ON BAYESIAN INFERENCE

For the multimode soft sensor, one important question is to identify which mode the new query sample belongs to. In the distance clustering based multimode soft sensor, the assignment of new sample is determined by the spatial similarity. Usually, two ways are used. One way depends on the distance between the new sample and the cluster centers, which assigns the new sample to the cluster with the minimum distance. The other approach applies the K nearest neighbor method, which recognizes the cluster according to the K nearest samples. However, for unobserved multimode processes, both mode identification methods lose their feasibility.

To handle the above problem, this section proposes an online mode identification strategy based on Bayesian inference. Assuming that the clustering on the training dataset brings C modes $\{M_1, M_2, \cdots, M_C\}$, the occurrence probability of the mode M_j regarding the query sample $\boldsymbol{x}_i$ is obtained by

$$p(M_j|\boldsymbol{x}_i) = \frac{p(\boldsymbol{x}_i|M_j)p(M_j)}{p(\boldsymbol{x}_i)} \tag{17}$$

where the $p(\boldsymbol{x}_i)$ is the occurrence probability of $\boldsymbol{x}_i$, which can be computed by

$$p(\boldsymbol{x}_i) = \sum_{j=1}^{C} p(\boldsymbol{x}_i|M_j)p(M_j) \tag{18}$$

where $p(M_j)$ is the prior probability of the mode M_j, while $p(\boldsymbol{x}_i|M_j)$ is the conditional probability of the sample $\boldsymbol{x}_i$ under the mode M_j. The prior probability $p(M_j)$ can be estimated by the training data or decided by the expert experience. The conditional probability $p(\boldsymbol{x}_i|M_j)$ is designed as:

$$p(\boldsymbol{x}_i|M_j) = exp(-(e_i^{(j)})^2/2) \tag{19}$$

where $e_i^{(j)}$ represents the estimation error of the j-th soft sensor model on the sample $\boldsymbol{x}_i$. Theoretically, we compute this error based on the the sample $\boldsymbol{x}_i$'s estimated output $f_j(\boldsymbol{x}_i)$ and the real output y_i. However, in real applications, the real output y_i is unknown at the i-th sample instant. Therefore, we apply the model estimation results at the $(i-1)$-th time instant to substitute the above expression, which results in

$$e_i^{(j)} = y_{i-1} - f_j(\boldsymbol{x}_{i-1}), \tag{20}$$

In fact, this applies an underlying assumption that the continuous two samples belong to the same mode. Considering the real industries often run under the same mode for a long period, this assumption is practicable.

Finally, the mode $M(\boldsymbol{x}_i)$ of the query sample $\boldsymbol{x}_i$ can be determined as the mode with the maximum posterior probability $p(M_j|\boldsymbol{x}_i)$, that means

$$M(\boldsymbol{x}_i) = arg \max_{M_j} p(M_j|\boldsymbol{x}_i) \tag{21}$$

D. SOFT SENSING PROCEDURE BASED ON LFC-KPLS

The proposed LFC-KPLS soft sensing procedure for unobserved multimode processes involves two stages: offline modeling and online application. During the offline modeling stage, the LFC-KPLS model is developed based on the training data, while at the online application stage, the query sample is collected and its corresponding output is given based on the developed soft sensor model. The details are listed as follows.

Offline Modeling Stage:

- Gather the training dataset $\boldsymbol{X}$, and standardize it with their mean and variance.
- Perform the latent factor clustering on the standardized training data to divide them into the C data modes $\{\boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^C\}$.
- For each mode $\boldsymbol{X}^i (1 \leq i \leq C)$, the local KPLS model is developed.

Online Application Stage:

- Collect the query sample $\boldsymbol{x}_i$ at the i-th sample instant, and standardize it with the mean and variance of the training data.
- Identify the mode $M(\boldsymbol{x}_i)$ of $\boldsymbol{x}_i$ using the Bayesian inference technology.
- Project the query sample onto the corresponding KPLS model and obtain the soft sensor output prediction.

IV. CASE STUDY

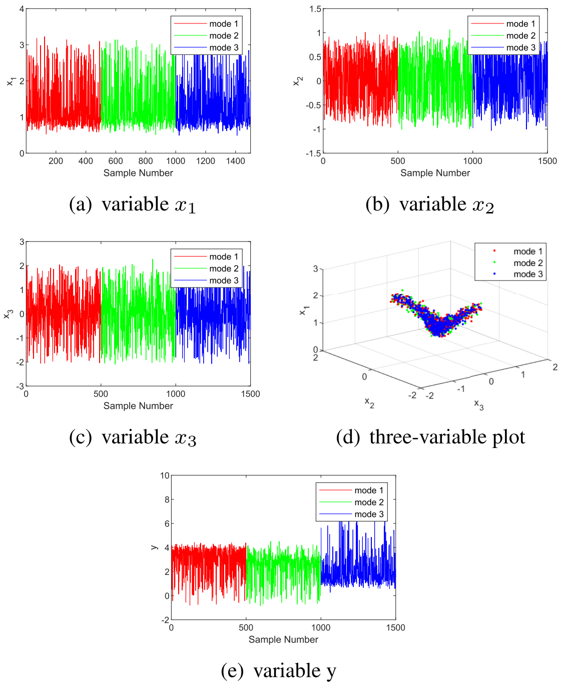
This section applies two case studies to validate the proposed multimode soft sensor method. One is the numerical example, while another is about the continuous stirred tank reactor (CSTR) system. The prediction performance of the proposed method is evaluated by the index of the root means squared error (RMSE). The better algorithm should be with the smaller RMSE.

A. A NUMERICAL SYSTEM

To test the proposed method, a numerical system is designed as follows [16]. Three nonlinearly-related input variables are expressed by

$$\begin{cases} x_1 = t^2 - t + 1 + e_1, \\ x_2 = sin(t) + e_2, \\ x_3 = t^3 + e_3. \end{cases} \tag{22}$$

where t is the random source variable with the uniform distribution in the range of [-1,1], $e_i (1 \leq i \leq 3)$ is the Gaussian noise with zero mean and the variance of 0.01. Based on the input variables, the output variables under three modes are computed as



(a) variable x_1 (b) variable x_2

(c) variable x_3 (d) three-variable plot

(e) variable y

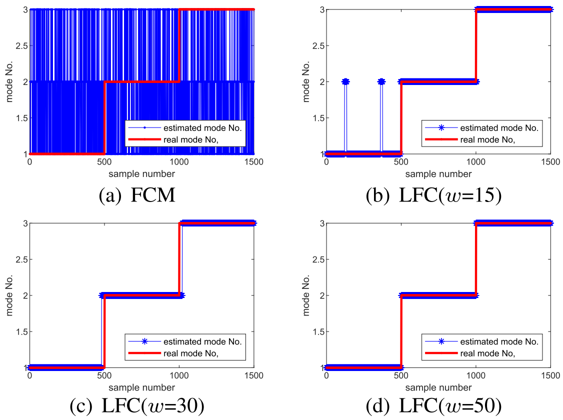
FIGURE 2. The plot of input-output variable.

$$y = \begin{cases} x_1 + x_1x_2 + 3cos(x_3) + e_4, & mode\ 1 \\ sin(x_1) + 2cos(x_2) + x_3 + e_4, & mode\ 2 \\ x_1^2 + sin(x_1x_2) + x_3 + e_4. & mode\ 3 \end{cases} \tag{23}$$

where e_4 is the output noise with the same characteristic to the input noise.

For each mode, 500 samples are simulated as the training dataset, while the other 300 samples are generated to constitute the testing dataset. The input variables of the training data are plotted in Fig. 2a-c, where the first 500 samples (No.1-500) belong to mode 1, the middle 500 samples (No.501-1000) are from mode 2, while the last samples (No. 1001-1500) belong to mode 3. It is seen that the input variables of all modes follow the similar distribution. A three-dimensional plot of the input variables is plotted in the Fig. 2d, which indicates clearly that the input variables are distance indivisible. Furthermore, the output variable of the training data is given in the Fig. 2e. We see that there is no obvious distinction in view of the output variable. By analyzing the characteristics of the input and output variables, the numerical system is a typical unobserved multimode system. For this kind of system, it is a challenging problem to build the corresponding soft sensor.

To deal with the soft sensor modeling of unobserved multimode system, this paper proposes the improved KPLS method by incorporating LFC. We first validate the effectiveness of LFC. The training data are used to test whether the proposed method can identify three different modes correctly. For a comparison, the basic FCM clustering is also



(a) FCM (b) LFC(w=15)

(c) LFC(w=30) (d) LFC(w=50)

FIGURE 3. Mode identification results.

TABLE 1. Mode identification rates (%) by FCM and LFC methods.

$Method$	Mode identification rate
FCM	32.68%
LFC(w=15)	96.00%
LFC(w=30)	97.33%
LFC(w=50)	100.0%

used for mode identification. The mode identification results are plotted in Fig. 3 and summarized in Table 1. Fig. 3a lists the mode identification results of FCM, where many samples are misclassified with a low mode identification rate of 32.68%. When the LFC is applied, the training samples are divided into several batches. If the batch size is set to 15, 30, 50, respectively, the mode identification results are given in Figs. 3b to 3d, correspondingly. It is observed that if the batch size is set as 15, some samples from mode 1 is wrongly recognized as the mode 2. When the batch size w is chosen as 30, most of the samples are correctly identified besides some samples in the mode switch procedure. In this case, the mode identification rate is 97.33%. For a large batch size $w = 50$, all the samples are correctly identified with 100% mode identification rate. No matter what value is used, the LFC outperforms the basic FCM method. In the practice, the determination of the batch size is based on the user experience.

Next we analyze the prediction performance of the proposed method. For the method comparison, the basic KPLS method and the FCM-KPLS method are also applied to build the soft sensors. For all the used methods, the kernel width parameter σ and the kernel score vector number L are optimized by the intelligent difference evolutionary (DE) algorithm. The prediction charts of three methods are shown in Figs. 4, 5 and 6. TABLE 2 quantitatively compares the RMSE values of different soft sensors. By Fig. 4, the basic KPLS method can not predict the change of the output effectively, which has a large RMSE of 1.2038. When FCM-KPLS is used, as it is not able to distinguish the different modes, its prediction performance is also unsatisfactory. The RMSE of FCM-KPLS is even increased to 1.2088. That shows unreasonable mode partition can worsen the soft sensor performance. With the proposed method, LFC can recognize the multiple modes correctly and the soft sensor can provide the remarkable performance improvement compared to the basic KPLS and FCM-KPLS method. The RMSE is reduced to 0.3173. To sum up, by projecting data into nonlinear latent space, the latent factor clustering based KPLS method can solve the unobserved multimode soft sensor modeling issue effectively.

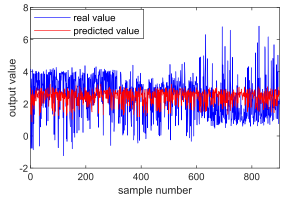


FIGURE 4. The numerical system prediction results based on KPLS.

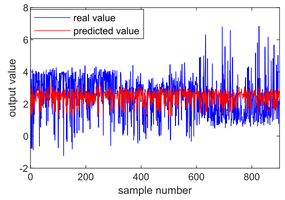


FIGURE 5. The numerical system prediction results based on KFCM-KPLS.

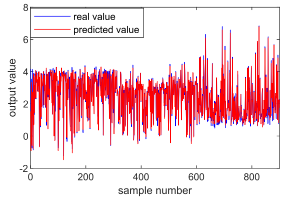


FIGURE 6. The numerical system prediction results based on LFC-KPLS.

TABLE 2. Model prediction performance by KPLS, FCM-KPLS and the proposed method for the numerical system.

Method	Prediction RMSE
KPLS	1.2038
FCM-KPLS	1.2088
The proposed method	0.3173

B. THE CONTINUOUS STIRRED TANK REACTOR SYSTEM

The continuous stirred tank reactor system (CSTR) [29] is a well-known industrial process and its diagram is illustrated in Fig. 7. It has the characteristics of nonlinearity and multimode because of the complex chemical reaction mechanism and the process condition change. In CSTR system, the reactant A is transformed into the product B through a irreversible chemical reaction. The concentration of reactant A in the output is one key quality variable, which is chosen as the output y of the soft sensor. Eight auxiliary variables x_1 to x_8 are

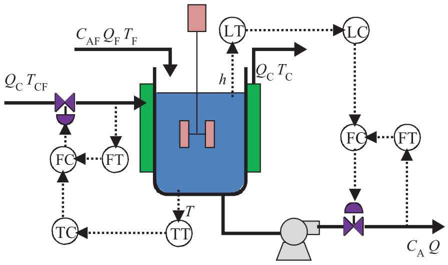


FIGURE 7. The CSTR system diagram.

TABLE 3. The variable list of the CSTR system.

Variable	Description
x_1	The reactor feed flow rate Q_f
x_2	The reactor feed temperature T_f
x_3	The reactor level h
x_4	The reactor temperature T
x_5	The reactor outlet flow rate Q
x_6	The coolant flow rate Q_c
x_7	The coolant feed temperature T_{cf}
x_8	The coolant outlet temperature T_c
y	The concentration of reactant A in the reactor outlet

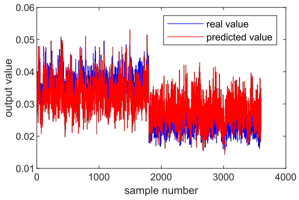


FIGURE 8. The CSTR system prediction results based on KPLS.

selected as the inputs of the soft sensor and their descriptions are demonstrated in table 3.

Mechanical simulation is carried out to generate the system data, which involves two different operation modes. A total of 7200 samples are collected from the process simulator and one half is used as the training dataset while the other half is applied as the testing dataset. The training data set is firstly processed by the nonlinear latent space clustering to identify the different process modes. We set the batch size as 72 samples (2% of the whole training data). Therefore, 50 batches are obtained in the latent space. Three methods of KPLS, FCM-KPLS and LFC-KPLS are applied to the CSTR system soft sensor modeling. Then the testing data including 3600 samples are projected on these three models for performance comparison. Figs. 8 to 10 show the prediction results of three different soft sensors and table 4 quantitatively compares the prediction RMSE. By the Fig. 8, it is observed that the prediction output of KPLS has a clear bias with the real output. The corresponding RMSE is 0.5926×10^{-2}. When FCM-KPLS is applies, it can not identify the correct data mode and therefore has a close performance with the basic KPLS method. As LFC can recognize the data modes effectively, the proposed LFC-KPLS method reduces the RMSE to 0.2407×10^{-2}. The Fig. 10 demonstrates that the output predictions are very close to the real values. Generally, the applications on the CSTR system show that the proposed method can build the more precise soft sensor in the case of unobserved multimode data.

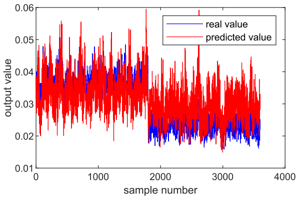


FIGURE 9. The CSTR system prediction results based on KFCM-KPLS.

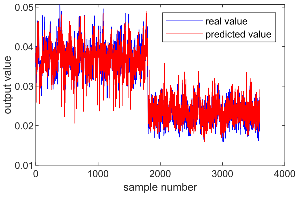


FIGURE 10. The CSTR system prediction results based on LFC-KPLS.

TABLE 4. Model prediction performance by KPLS, FCM-KPLS and the proposed method for the CSTR system.

Method	Prediction RMSE (10^{-2})
KPLS	0.5926
FCM-KPLS	0.5988
The proposed method	0.2407

V. CONCLUSION

In this paper, a novel soft sensor modeling method called LFC-KPLS is developed for the unobserved multimode nonlinear processes. The proposed method designs a modeling framework for the unobserved multimode framework, which can be generalized to many other similar data-driven soft sensor modeling method. Besides the design of the modeling framework, the other two important aspects in the proposed method are the offline mode clustering for the unobserved multimode data, and the online mode identification for the query sample. Two case studies, including one numerical system and the continuous stirred tank reactor system (CSTR), are applied to test the proposed method. The application results demonstrate that the LFC can identify the data modes more effectively than the basic FCM method, and the proposed soft sensor has a higher prediction precision compared to the traditional KPLS and FCM-KPLS methods. However, some limitations of the proposed method should be also noted. As the moving window technique is applied to divide the data batches, this method is based on the underlying assumption that the process mode could last a period of time and does not change suddenly, and the offline historical data are enough plentiful for model training. In the case of limited training data, some new methods should be investigated in the future work.

REFERENCES

[1] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.

[2] L. Yao and Z. Ge, "Big data quality prediction in the process industry: A distributed parallel modeling framework," *J. Process Control*, vol. 68, pp. 1–13, Aug. 2018.

[3] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[4] J. Zheng, Z. Song, and Z. Ge, "Probabilistic learning of partial least squares regression model: Theory and industrial applications," *Chemometric Intell. Lab. Syst.*, vol. 158, pp. 80–90, Nov. 2016.

[5] Z. Ge, "Mixture Bayesian regularization of PCR model and soft sensing application," *IEEE Trans. Ind. Electron.*, vol. 62, no. 7, pp. 4336–4343, Jul. 2015.

[6] W. Shao, Z. Ge, and Z. Song, "Soft-sensor development for processes with multiple operating modes based on semisupervised Gaussian mixture regression," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 5, pp. 2169–2181, Sep. 2019.

[7] L. Yao and Z. Ge, "Distributed parallel deep learning of hierarchical extreme learning machine for multimode quality prediction with big process data," *Eng. Appl. Artif. Intell.*, vol. 81, pp. 450–465, May 2019.

[8] R. Sharmin, U. Sundararaj, S. Shah, L. Vande Griend, and Y.-J. Sun, "Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant," *Chem. Eng. Sci.*, vol. 61, no. 19, pp. 6372–6384, Oct. 2006.

[9] K. Zheng and K. Funatsu, "Partial constrained least squares (PCLS) and application in soft sensor," *Chemometric Intell. Lab. Syst.*, vol. 177, pp. 64–73, Jun. 2018.

[10] J. Zheng and Z. Song, "Semisupervised learning for probabilistic partial least squares regression model and soft sensor application," *J. Process Control*, vol. 64, pp. 123–131, Apr. 2018.

[11] S. Wold, N. Kettaneh-Wold, and B. Skagerberg, "Nonlinear PLS modeling," *Chemometric Intell. Lab. Syst.*, vol. 7, nos. 1–2, pp. 53–65, Dec. 1989.

[12] S. J. Qin and T. J. McAvoy, "Nonlinear PLS modeling using neural networks," *Comput. Chem. Eng.*, vol. 16, no. 4, pp. 379–391, Apr. 1992.

[13] Y. H. Bang, C. K. Yoo, and I.-B. Lee, "Nonlinear PLS modeling with fuzzy inference system," *Chemometric Intell. Lab. Syst.*, vol. 64, no. 2, pp. 137–155, Nov. 2002.

[14] H. Liu, C. Yang, B. Carlsson, S. J. Qin, and C. Yoo, "Dynamic nonlinear partial least squares modeling using Gaussian process regression," *Ind. Eng. Chem. Res.*, vol. 58, no. 36, pp. 16676–16686, Aug. 2019.

[15] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel Hilbert space," *J. Mach. Learn. Res.*, vol. 2, pp. 97–123, Mar. 2001.

[16] X. Zhang, W. Yan, and H. Shao, "Nonlinear multivariate quality estimation and prediction based on kernel partial least squares," *Ind. Eng. Chem. Res.*, vol. 47, no. 4, pp. 1120–1131, Feb. 2008.

[17] X. Wang, P. Wang, X. Gao, and Y. Qi, "On-line quality prediction of batch processes using a new kernel multiway partial least squares method," *Chemometric Intell. Lab. Syst.*, vol. 158, pp. 138–145, Nov. 2016.

[18] F. Chu, X. Cheng, R. Jia, F. Wang, and M. Lei, "Final quality prediction method for new batch processes based on improved JYKPLS process transfer model," *Chemometric Intell. Lab. Syst.*, vol. 183, pp. 1–10, Dec. 2018.

[19] J. Tang, J. Zhang, Z. Wu, Z. Liu, T. Chai, and W. Yu, "Modeling collinear data using double-layer GA-based selective ensemble kernel partial least squares algorithm," *Neurocomputing*, vol. 219, pp. 248–262, Jan. 2017.

[20] H. Jin, X. Chen, J. Yang, and L. Wu, "Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes," *Comput. Chem. Eng.*, vol. 71, pp. 77–93, Dec. 2014.

[21] N. Chen, J. Dai, X. Yuan, W. Gui, W. Ren, and H. N. Koivo, "Temperature prediction model for roller kiln by ALD-based double locally weighted kernel principal component regression," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 2001–2010, Aug. 2018.

[22] M. Zhao, S. Ma, and J. Ren, "An improved ensemble adaptive kernel PLS soft sensor model and its application," in *Proc. 37th Chin. Control Conf. (CCC)*, Wuhan, China, Jul. 2018, pp. 8098–8103.

[23] X. Yuan, J. Zhou, Y. Wang, and C. Yang, "Fuzzy C-means cluster based on local weighted principal component regression for soft sensor of an industrial hydrocracking process," in *Proc. 12th Asian Control Conf. (ASCC)*, Kitakyushu, Japan, Jun. 2019, pp. 242–247.

[24] A. Gholami, M. Shahbazian, and G. Safian, "Soft sensor development for distillation columns using fuzzy C-means and the recursive finite newton algorithm with support vector regression (RFN-SVR)," *Ind. Eng. Chem. Res.*, vol. 54, no. 48, pp. 12031–12039, Nov. 2015.

[25] L. Wang, J. Zeng, X. Liang, Y. He, S. Luo, and J. Cai, "Soft sensing of a nonlinear multimode process using a self organizing model and conditional probability density analysis," *Ind. Eng. Chem. Res.*, vol. 58, no. 31, pp. 14267–14274, Jul. 2019.

[26] J. Liu, "Developing a soft sensor with online variable reselection for unobserved multi-mode operations," *J. Process Control*, vol. 42, pp. 90–103, Jun. 2016.

[27] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets Syst.*, vol. 158, no. 19, pp. 2095–2117, Oct. 2007.

[28] B. Zhu, Z.-S. Chen, Y.-L. He, and L.-A. Yu, "A novel nonlinear functional expansion based PLS (FEPLS) and its soft sensor application," *Chemometric Intell. Lab. Syst.*, vol. 161, pp. 108–117, Feb. 2017.

[29] X. Deng and X. Tian, "Sparse kernel locality preserving projection and its application in nonlinear process fault detection," *Chin. J. Chem. Eng.*, vol. 21, no. 2, pp. 163–170, Feb. 2013.

XIAOGANG DENG (Member, IEEE) received the B.Eng. and Ph.D. degrees from the China University of Petroleum, China, in 2002 and 2008, respectively. He is currently an Associate Professor with the College of Control Science and Engineering, China University of Petroleum. From October 2015 to October 2016, he was a Visiting Scholar with the Department of Electronics and Computer Sciences, University of Southampton, Southampton, U.K. His research interests include statistical process monitoring, industrial process modeling and simulation, and data-driven fault diagnosis.

YONGXUAN CHEN received the bachelor's degree in engineering from the China University of Petroleum, in 2017, where he is currently pursuing the master's degree with the College of Control Science and Engineering. His research interests include machine learning theory and industrial process quality modeling.

PING WANG received the B.Eng. degree in engineering and the Ph.D. degree in control theory and control engineering from the China University of Petroleum, in 2005 and 2012, respectively. He is currently a Lecturer with the College of Control Science and Engineering, China University of Petroleum. His research interests include complex industrial process modeling, and advanced control and optimization.

YUPING CAO received the B.Eng. degree in engineering and the Ph.D. degree in control theory and control engineering from the China University of Petroleum, in 2004 and 2010, respectively. She is currently a Lecturer with the College of Control Science and Engineering, China University of Petroleum. Her research interests include process monitoring, fault diagnosis, and fault prediction.