DTU Library

A Semi-supervised Soft-sensor of Just-in-time Learning With Structure Entropy Clustering and Applications for Industrial Processes Monitoring

Li, Dong; Liu, Yiqi; Huang, Daoping; Xu, Chong

Published in:
IEEE Transactions on Artificial Intelligence

Link to article, DOI:
10.1109/TAI.2022.3217028

Publication date:
2023

Document Version
Peer reviewed version

A Semi-supervised Soft-sensor of Just-in-time Learning with Structure Entropy Clustering and Applications for Industrial Processes Monitoring

Dong Li, Yiqi Liu, *Scenior Member, IEEE*, Daoping Huang and Chong Xu

***Abstract* — To monitor industrial processes properly, soft-sensors are widely used to predict significant but difficult-to-measure quality variables. However, the prediction performances of traditional data-driven soft-sensors are usually unacceptable once suffering from high-nonlinear, high-dimension and imblance data issues. Therefore, a semi-supervised soft-sensor, which is learned by a just-in-time method with structure entropy clustering (SS-JITL-SEC), is proposed aiming to improve prediction performance with a simpler way. Inspired by a divide and conquer strategy, a novel SEC method is proposed to achieve several clusters and then to translate the highly complex and nonlinear modeling problems into simple and linear ones. Moreover, the training dataset is extended through a mixed semi-supervised (SS) labeling approach. Finally, dissimilarity-based just-in-time learning (JITL) works together with the resulting clustering sub-datasets to formulate a local adaptive prediction model. Two datasets from different types of wastewater treatment plants are used to verify the effectiveness of the proposed soft-sensor. The results show that the SS-JITL-SEC soft-sensor can achieve better prediction performance than other standard counterparts, and even for effective process monitoring with the resulted residuals.**

***Impact Statement* — Proper processes monitoring of difficult-to-measure quality-related variables is imperative for safe and stable operation of industrial processes, particularly under the case of suffering from significantly dynamic, highly dimensional behaviors during supervised learning. Data-driven soft-sensors together with adaptive learning and semi-supervised learning are currently the alternatives to achieve this goal. The novelty of present work is to propose a just-in-time learning for semi-supervised soft-sensor together with a structure entropy clustering algorithm. Inspired by a divide and conquer strategy, a complex model learning problem can be simplified. Then the resulted soft-sensor can be used for online monitoring of difficult-to-measure variables. The proposed case studies demonstrate that this soft-sensor is able to overcome the limitations of standard modeling problem for the complex processes using insufficient samples. With the proposed soft-sensors, we believe that the proposed methodology could be applied for a wider range of fields, such as chemical plants, wind power plants and so on.**

***Index Terms*—Clustering, Semi-supervised learning, Adaptive algorithms, Industrial processes monitoring**

I. Introduction

WITH the development of industrial processes, proper monitoring of significant but difficult-to-measure quality variables has recently received more and more attention [1-4]. Online measurement equipment usually fails in industrial processes due to harsh working conditions and high costs of maintenance, even though they can achieve acceptable results [5-7]. Therefore, soft-sensors are studied and act as an alternative to back up a hardware sensor or to address the industrial process monitoring problems by justifying the residuals between the predictive and real values. Though refining the hidden information from easy-to-measure variables and analyzing mechanism knowledge of industrial processes, various types of soft-sensors can be constructed and applied [8-10].

In general, soft-sensors are categorized as mechanism modeling and data-driven modeling [11]. Mechanism modeling

This work was supported by the Horizon 2020 Framework Programme-Marie Skłodowska-Curie Individual Fellowships (891627), also partially supported by the National Natural Science Foundation of China (61873096, 62073145), the Basic and Applied Basic Research Foundation of Guangdong Province (2020A1515011057, 2021B1515420003), the Guangdong International Scientific Cooperation Research Foundation (2020A0505100024, 2021A0505060001).. (Corresponding author: Yiqi Liu, e-mail: aulyq@scut.edu.cn.)

D. Li is with the Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: lddscut@163.com).

Y. Liu is with the Key Laboratory of Autonomous Systems and Networked Control, Ministry of Education, School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China; He is also with the Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 227, 2800 Kgs. Lyngby, Denmark (e-mail: aulyq@scut.edu.cn).

D. Huang is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: audhuang@scut.edu.cn).

C. Xu is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China and School of Physics& Electronic Information, Gannan Normal University, Ganzhou 341000, China (e-mail: jxgz2009xc@126.com).

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2022.3217028

aims to establish prediction models by analyzing mechanism knowledge and translating them into mathematical relationships [12-14]. However, due to the complexity of industrial processes as well as lack of mechanism knowledge, mechanism modeling is usually difficult to be widely used, particularly, when unknown or unexpected working conditions happen. Fortunately, data-driven modeling provides a convenient and efficient alternative to predict the difficult-to-measure quality variables. Data-driven modeling describes the inherent relationships between input and output variables without the need to understand all the mechanism details [2, 15-18]. In recent years, data-driven modeling methods, such as partial least squares (PLS) [19], multiple linear regression (MLR) [20] and artificial neural networks (ANNs) [21], have been successfully used for industrial processes monitoring. However, data-driven modeling of more complex industrial processes implies more requirements of training samples and training time. To deal with these problems, Liu et al. proposed a bagging method to boost the utilization efficiency of training samples [22]. But more replicated samples do not mean more useful information. Li et al. proposed a co-training approach to enrich the training samples by labeling the unlabeled data as labeled data [23]. However, the computational intensity using the cross-validation process is unacceptable. Mohamed et al. combined mechanistic modeling with data-driven modeling to build a hybrid prediction model [24]. Despite the hybrid prediction model reducing dependence on training samples, the application scope of mechanistic modeling is limited and hard to generalize.

Clustering methods provide an alternative to reduce the difficulty of modeling and decrease the time-consuming [25]. Inspired by the divide and conquer strategy, clustering methods herein are able to divide samples into several sub-datasets by analyzing the information interaction [26]. By doing so, each sub-dataset involves the most relevant information between them, making complex and nonlinear modeling problems translate into simple and linear problems. Thus, the prediction model can be built more effectively and achieve better prediction performance with less training data. Oyelade et al. proposed a k-means clustering algorithm for monitoring students' academic performance [27]. Depending on the similarity measurement with Euclidean distance, clustering sub-datasets with distinct non-lapping boundaries were produced and then worked together with the simple regression models to achieve effective prediction. But some initial parameters of the k-means clustering algorithm must be given properly in advance, such as the clustering centers and the number of centers. Also, the computational costs of continually iterative training processes are often unmanageable. Lu et al. proposed a Gaussian mixture model (GMM) clustering method to extract sub-patterns of heating load [28]. The motivation of GMM clustering is to make full use of the joint probability distribution within samples to decouple the intra-relationship of sub-patterns. However, the method usually has terrible clustering results for samples with high-dimension and uneven spatial distribution. Therefore, designing a reliable and efficient clustering method is an imperative step to achieve samples clustering with high-dimension and uneven spatial distribution.

Even though clustering methods can simplify the modeling problems for industrial process monitoring, reduce the difficulty of modeling and decrease the time-consuming, the highest priority is still to collect the appropriate training samples. Due to the lack of measurement devices and the high cost of collection, only a limited amount of labeled data, which contain both input and output variables, are available. On the contrary, a large amount of unlabeled data, which only contain input variables, are taken as useless for granted. Therefore, semi-supervised (SS) learning algorithms are necessarily studied to increase the utilization of unlabeled data, further addressing the problem of imbalance between labeled data and unlabeled data [29-31]. The main purpose of SS learning is to enrich the initial training samples by labeling the unlabeled data as new labeled data, then to increase training data size. Sun et al. proposed a self-training strategy for the classification of samples [32]. In this paper, the self-training strategy was used to select and label the unlabeled samples optimized in the outer loop. However, the mutual information between labeled and unlabeled samples is usually ignored, which makes the error be accumulated in prediction models and has a negative influence on further prediction performance. Li et al. proposed a co-training approach to reduce the accumulation of errors [23]. The main principle is to divide the training samples into two independent parts and to estimate the output variables of unlabeled samples, respectively, then the appropriate unlabeled samples can be selected by cross-validation between two parts. Unfortunately, the cross-validation process is time-consuming. Tao et al. proposed a scalable multi-view semi-supervised algorithm for classification problems [33]. The algorithm improves computational efficiency by directly constructing a global regression model for each sample, and can solve large-scale multi-view semi-supervised classification problems. However, the algorithm is susceptible to outliers and limited to classifications, rather than regressions.

In addition, traditional prediction models are usually trained offline but used online, making model prediction performance deteriorate. Therefore, it becomes more necessary than ever to devote to equipping a soft-sensor with adaptive ability. One of the commonly used methods is resorting to the moving window (MW) technique. Zhou et al. used the MW technique to develop a model to predict fuel cell degradation [34]. The purpose of the MW technique is to iteratively update the model structure and parameters by changing the informative regions online. But it is susceptible to abnormal samples in the training process. Wu et al. proposed a time difference (TD) approach to ensure online prediction for variables in wastewater treatments [35]. The approach can adapt to gradual drifts of both secondary variables and corresponding targets simultaneously. However, this method is also easily affected by abnormal samples and only suitable to deal with the prediction problem of continuous data. Therefore, Yuan et al. proposed a multi-similarity measurement-driven ensemble just-in-time learning (E-JITL) for adaptive soft-sensors [36]. In E-JITL, different distance measurement strategies, such as Euclidean and Mahalanobis distances, are adopted for samples selection and then local prediction models can be established. However, each distance measurement strategy often focuses on a certain aspect of samples information and has its own limitations. Thus, traditional distance measurement strategies may be insufficient or even fail to capture the similarity between samples.

To solve the aforementioned problems, this paper proposed a semi-supervised soft-sensor of just-in-time learning with structure entropy clustering (SS-JITL-SEC). The main contributions can be reflected in the following aspects:

1) A dissimilarity measurement strategy is defined for clustering and modeling. It is able to calculate distance values by analyzing the dissimilarity between two sets of samples, which cannot only effectively alleviate the negative influence of outliers, but also deal with high-nonlinear and high-dimension data issues in the process industries.

2) Depending on the proposed dissimilarity measurement strategy, a novel structure entropy clustering (SEC) method is derived for dynamic data feature extraction. Compared to the existing clustering methods, initial parameters of the SEC method do not need to be given in advance, therefore avoiding the heavy workforce of prior knowledge setting as well as the continuous and costly iterative learning processes.

3) Since a large amount of unlabeled data are generally discarded wastefully, a mixed SS labeling method is proposed to label the unlabeled data to extend the original training dataset. The method considers the specific distribution scenarios of each unlabeled data, and then estimates the output variable of each unlabeled data with proper approaches.

4) To prevent deterioration of prediction performance because of the significant variations within test data, the dissimilarity measurement strategy is assimilated into JITL and works with the resulted clustering sub-dataset to formulate a local adaptive prediction model, thereby ensuring the high prediction quality sequentially.

The organization of the remainder of this paper is as follows: In Section II, the basic concepts of structure entropy and JITL are briefly introduced. Section III presents the proposed dissimilarity measurement strategy, SEC method as well as the framework of the SS-JITL-SEC soft-sensor in detail. Section IV demonstrates the effectiveness of the proposed soft-sensor through a real full-scale wastewater plant and an oxidation ditch (OD) wastewater plant. Section V discusses the advantages and disadvantages of the SS-JITL-SEC soft-sensor. Finally, Section VI provides the conclusion and future works.

II. Basis Overview

A. Structure Entropy

Entropy is usually used to describe the chaos of atom distribution in physics. A smaller entropy value indicates a more well-organized system. Similarly, in information theory, entropy describes the uncertainty degree for an information source in the signal. Therefore, entropy is a measurement of chaos in various systems [37]. When the system has n different states, the probability of every state is defined as P_i ($i = 1,2,\dots,n$). The entropy of the system is:

$$E = -\sum_{i=1}^{n} P_i ln P_i \tag{1}$$

where P_i must satisfy $0 \le P_i \le 1$ and $\sum_{i=1}^{n} P_i = 1$.

To assimilate entropy into clustering for samples analysis, the distribution of samples can be regarded as similar to the atom distribution. Thus, structure entropy can be used to describe the internal information of samples [38]. The structure entropy of the system is defined:

$$E_i = \sum_{j\in X}^{i\neq j}(S_{ij} log_2 S_{ij} + (1 - S_{ij}) log_2(1 - S_{ij})) \tag{2}$$

where X is a sample set, $i = 1,2,\dots,n$, and S_{ij} is the similarity between x_i and x_j.

$$S_{ij} = e^{-\alpha D_{ij}} \tag{3}$$

where D_{ij} is the Euclidean distance between x_i and x_j. α is the curvature of the exponential function:

$$\alpha = -ln 0.5/\bar{D} \tag{4}$$

where $\bar{D}$ is the average Euclidean distance among all the samples. It is obvious that $0 \le S_{ij} \le 1$, when $S_{ij} = 0.5$, $E_i = 1$ is the maximum. When $S_{ij} = 0$ or 1, $E_i = 0$ is the minimum.

According to the above definition and analysis, if the samples are either farther or closer to each other, their structure entropy values should be smaller. In other words, a smaller structure entropy value means that the sample may be more likely located in the distribution center. Outliers will also have small structure entropy values, but they are not the distribution centers. Therefore, discriminating and removing outliers is a critical step for subsequent clustering and modeling.

B. Just-in-time Learning (JITL)

JITL is an effective tool to prevent the deterioration of prediction performance [39]. According to the similarity between samples, the most similar training samples are selected and a local adaptive model is constructed with these training samples to achieve accurate prediction and reliability of soft-sensors [40]. The major framework of JITL is shown as follows:

Firstly, x_t represents the test data, d_{lim} is a similarity threshold value. According to Euclidean distance values between x_t and original training samples, the local training samples space Ω_k can be determined:

$$\Omega_k = \{x_i, d(x_t, x_i) \le d_{lim}, i = 1,2,\dots,k \tag{5}$$

where k represents the size of the sample space Ω_k. $d(\cdot,\cdot)$ is the distance measurement between two data. In general,

$$d(x_t, x_i) = \|x_t - x_i\|_2 \tag{6}$$

where $\|\cdot,\cdot\|_2$ represents the 2-norms. Then, the local function of x_t is described as follows:

$$f(x_t, \theta) = \theta_0 + \theta_1(x_t - x_i) + \cdots + \theta_l(x_t - x_i)^l = \theta G(x_t - x_i) \tag{7}$$

where $\theta = (\theta_0, \theta_1, \dots, \theta_l)$, $G(x_t - x_i) = (1, (x_t - x_i), \dots, (x_t - x_i)^l)'$. The process of optimization is shown as follows:

$$\theta = argmin \sum_{i=1}^{k}\{y_i - \sum_{j=0}^{l}\theta_j(x_t - x_i)^j\} f_i \tag{8}$$

where y_i is the output of x_i. f_i is the function between $d(x_t, x_i)$ and d_{lim}:

$$f_i = K(\frac{d(x_t, x_i)}{d_{lim}}) \tag{9}$$

where f_i describes the contribution of x_i for the local model, it is usually determined by the kernel function K. Therefore, the prediction values $\hat{y}_t = x_t'\theta$.

In this paper, we use the new dissimilarity measurement strategy to replace the traditional 2-norms distance

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2022.3217028

measurement, making the JITL improve and named dissimilarity-based JITL.

III. A Semi-Supervised Soft-Sensor of Just-in-Time Learning with Structure Entropy Clustering (SS-JITL-SEC)

To enhance the prediction performance of soft-sensors in industrial processes monitoring, this paper proposes an SS-JITL-SEC soft-sensor. Depending on the dissimilarity measurement strategy, an SEC method is firstly derived for dynamic data feature extraction. Then, a mixed SS labeling method is proposed to enrich the original training dataset by labeling unlabeled data as new labeled data. Finally, the dissimilarity measurement strategy is assimilated into JITL and works with the resulted clustering sub-dataset to build a local adaptive prediction model, ensuring high prediction performance with the change of test data.

A. Dissimilarity Measurement Strategy

Dissimilarity measurement strategy is a new distance measurement strategy, which calculates the distance by analyzing the dissimilarity between two high-dimension samples. Given samples $x_i \in X, x_i = (x_{i1}, x_{i2}, \dots, x_{il}), 0 < i \leq n$, l presents the dimension of the samples, and n is the number of samples. The dissimilarity measurement is defined as follows:

$$d(x_i, x_j) = (l - s(x_i, x_j))/l \tag{10}$$

where d is the distance value between two samples, and s is the total similarities between x_i and x_j:

$$s(x_i, x_j) = \sum_{p=1}^{l} \theta(x_{ip}, x_{jp}) \tag{11}$$

where θ is the similarity between x_{ip} and x_{jp}:

$$\theta(x_{ip}, x_{jp}) = \begin{cases} 1, |x_{ip} - x_{jp}| \leq r \\ 0, |x_{ip} - x_{jp}| > r \end{cases} \tag{12}$$
$$i \neq j, 1 \leq i, j \leq n, 1 \leq p \leq l$$

where r is a similarity threshold value, x_{ip} and x_{jp} are the p-th column of x_i and x_j.

According to the properties of distance measurement, it is easy to prove that the proposed dissimilarity measurement strategy can satisfy non-negativity and symmetry, but it cannot follow the triangle inequality. Therefore, the proposed dissimilarity measurement strategy is only defined as a generalized distance measurement strategy. Fortunately, it satisfies the properties of a linear space, such as the calculation law of addition and multiplication, so it can be used to define a linear system space for subsequent clustering and modeling.

Compared with the traditional Euclidean distance measurement, because $\theta(x_{ip}, x_{jp})$ has the same weight influence on the final distance value of (x_i, x_j), the negative influence of outliers is reduced. Thus, it is a more rational distance measurement strategy and the reliability of the distance value is enhanced.

B. Structure Entropy Clustering (SEC)

In this paper, the dissimilarity measurement strategy replaces the traditional Euclidean distance measurement as a basis for clustering. For high-dimension samples, the calculation formula of structure entropy is still Eq. (2). Depending on the calculated structure entropy value of each sample, we can achieve dynamic samples clustering. Thus, the clustering method is named structure entropy clustering (SEC). The detailed steps are as follows:

The first step is to find and determine clustering centers. Initial clustering centers of traditional clustering methods are given in advance and depend on the given prior knowledge. And then the clustering centers are adjusted during the iterative training process, which is often unmanageable and time-consuming. However, the SEC method acquires the structure entropy value of each sample by one-circle calculation. According to the properties of structure entropy, if the structure entropy value approaches 0, the sample may be a clustering center. Therefore, it is cheap and efficient to determine the clustering centers of the SEC method and avoid the repeated iteration training process.

The samples with the minimum structure entropy values are selected as the clustering centers. Then, we need to determine

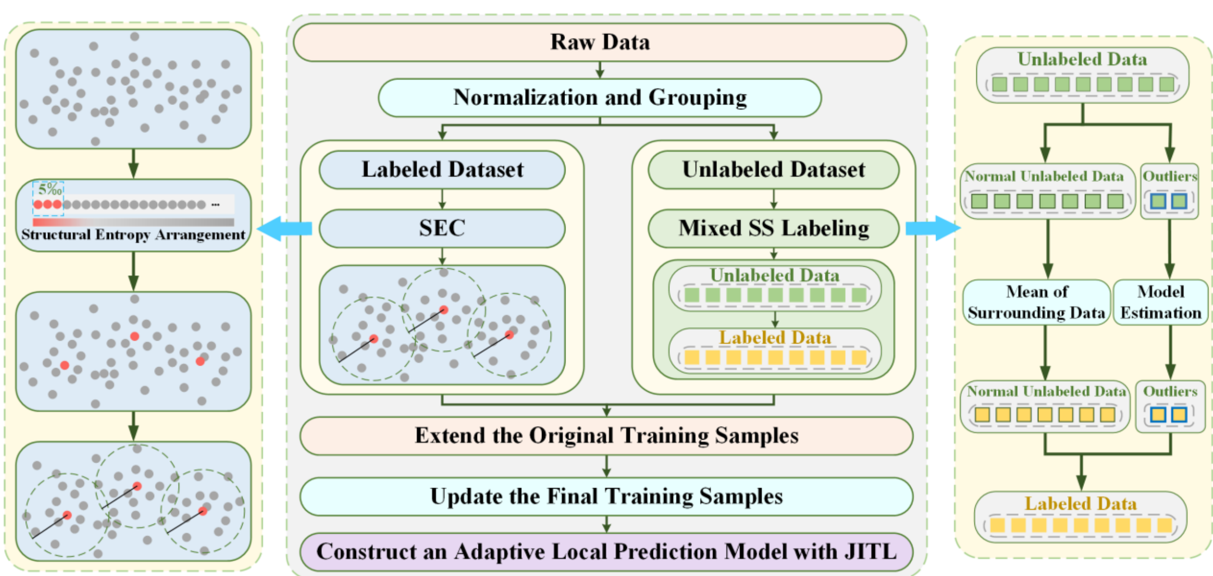

Fig. 1. The framework of SS-JITL-SEC soft-sensors

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2022.3217028

the number of clustering centers. Each sample x should satisfy the Gaussian distribution assumption $f(x)$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{13}$$

where σ represents the standard deviation, μ is the mean. Depending on the Gaussian distribution assumption, we determine the number of clustering centers: 5‰ of the number of total samples.

After the clustering centers and their number are determined, we calculate the dissimilarity measurement distance between clustering centers and other samples. Then, these samples are grouped depending on the dissimilarity distance values. If the dissimilarity distance value is less than a control threshold value Q, the sample will be grouped into the corresponding clustering sub-dataset. Otherwise, the sample will be removed. Where Q represents a specified control threshold value. In fact, the smaller control threshold value Q is, the more samples are in each clustering sub-dataset, but the worse clustering results may be. Therefore, a proper control threshold value Q is the primary factor influencing the clustering results.

In addition, the structure entropy values of outliers also approach 0. But they are not suitable to serve as clustering centers, even they will cause failed clustering results. Therefore, we define a threshold value inf of the size of clustering sub-dataset to eliminate the negative influence from outliers. If the size of resulted clustering sub-dataset is less than inf, the clustering center will be removed, and then we will re-determine new clustering centers.

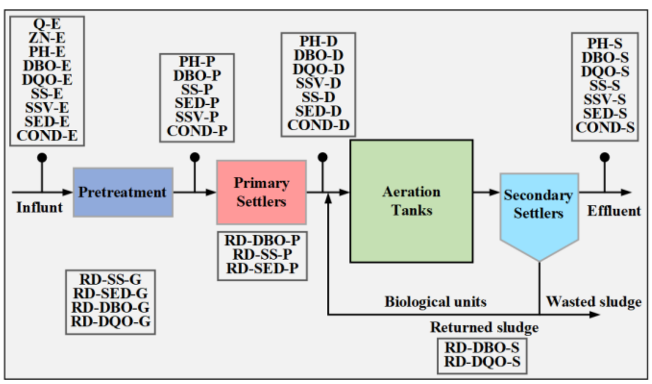

Fig .2. The schematic of the real full-scale wastewater plant

C. A Mixed Semi-Supervised (SS) Labeling Method

In general, a large amount of unlabeled data are discarded wastefully. An alternative, semi-supervised (SS) labeling method, deals with the problem by labeling the unlabeled data as labeled data to extend original training samples. However, traditional SS labeling methods usually ignore the negative influence of the surrounding labeled data on unlabeled data even easily fall into errors due to the outliers. Therefore, a mixed SS labeling method is proposed to address these issues. The detailed process is as follows:

Firstly, the dissimilarity distance values between unlabeled data and clustering centers are calculated through Eq. (10) - Eq. (12). Depending on the dissimilarity distance values, unlabeled data with the minimum dissimilarity distance value have the most similarity to the clustering sub-dataset. Then, we calculate the dissimilarity distance values between the unlabeled data and labeled data from this clustering sub-dataset. If the dissimilarity distance value is less than w, the labeled data will be selected to estimate the output value of the unlabeled data. Where w represents a specified control threshold value. Finally, the mean of these selected labeled data is taken as the estimated output value of the unlabeled data.

$$\hat{y}_i = \frac{1}{p}\sum_1^p y_j, d(x_i, x_j) \le w \tag{14}$$

where $\hat{y}_i$ is the estimated output value of x_i, (x_j, y_j) is the labeled data in the clustering sub-dataset and p is the number of the selected labeled data.

However, when the unlabeled data are outliers, due to the lack of surrounding labeled data, the output value can be only estimated by the training model. Thus, the negative influence of outliers is reduced even is eliminated.

The mixed SS labeling method correctly labels all the unlabeled data. Also, since the clustering sub-datasets are updated, the mixed SS labeling method is used online. Therefore, the proposed SS labeling method can make full and effective use of the unlabeled data to extend the original training samples.

TABLE I
RMSE AND R OF THE PREDICTION RESULTS IN THE FIRST CASE STUDY

Models	RMSE	R	Models	RMSE	R
k-means PLS	10.81	88.36%	**k-means MLR**	9.32	89.96%
SS PLS	7.68	86.78%	**SS MLR**	7.62	86.91%
Offline PLS	3.95	92.90%	**Offline MLR**	3.91	93.00%
SS-JITL-SEC PLS	3.63	93.56%	**SS-JITL-SEC MLR**	**3.56**	**93.67%**

D. JITL-SS-SEC Soft-Sensors

In this paper, the framework of the proposed soft-sensor is illustrated in Fig. 1. To understand the proposed soft-sensor more clearly, the detailed process of SS-JITL-SEC soft-sensors is as follows:

1) The first step is data pre-processing. The raw data are normalized to [0,1]. The normalized data will be beneficial to subsequent clustering and modeling. Then, they are divided into a labeled dataset L and an unlabeled dataset U depending on whether the output variable is contained.

2) The second step is dynamic samples clustering. The labeled dataset L is clustered by using the SEC method. According to the structure entropy value of each sample, clustering centers $l_1, l_2, \ldots l_m$ is determined firstly, where m is the number of clustering centers, and $m = 0.005 \times num(L)$, num represents the number of all the samples. Then, depending on the dissimilarity distance values between the clustering centers and other labeled data as well as newly defined a control threshold value Q, we can select the training samples with high similarity to derive the clustering sub-datasets $L_1, L_2, \ldots, L_m$.The first step is data pre-processing. The raw data are normalized to [0,1]. The normalized data will be beneficial to subsequent clustering and modeling. Then, they are divided into a labeled dataset L and an unlabeled dataset U depending on whether the output variable is contained.

3) The third step is extending the original training samples. We can use the mixed SS labeling method to label the unlabeled dataset U. The dissimilarity distance values between unlabeled data $x_u \in U$ and clustering centers are calculated through the proposed dissimilarity measurement strategy, and then the minimum dissimilarity distance value is found. According to the property of dissimilarity measurement value, x_u has the most similarity to these labeled data from the corresponding clustering sub-dataset. Thus, the output $\hat{y}_u$ of x_u can be estimated through the mean of the selected surrounding labeled data with the most similarity. However, if x_u is an outlier, surrounding labeled data are insufficient, so the output $\hat{y}_u$ can be estimated through training models. Finally, we labeled all the unlabeled data and used them to extend the original training samples.

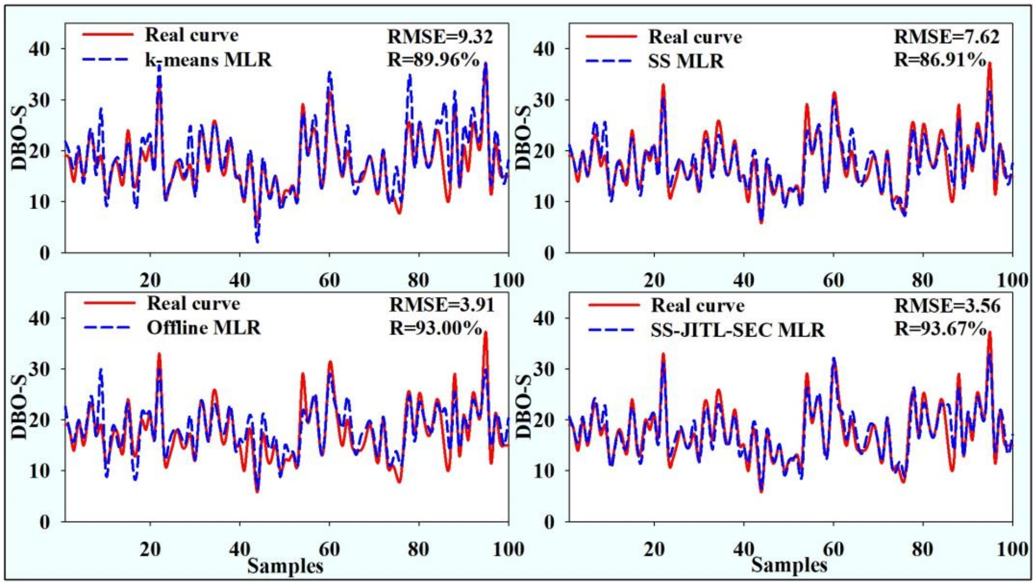

Fig. 3. The real and predictive curves in the first case study

4) The last step is the building of the prediction model with the final training samples. To prevent deterioration of prediction performance with the variations within test data, a dissimilarity-based JITL is used to update the prediction model. Firstly, the dissimilarity distance values between test data x_t and clustering centers $l_1, l_2, \ldots l_m$ are calculated. Then, we select the resulted clustering sub-dataset corresponding to the minimum dissimilarity distance value as the final training samples. Finally, a local adaptive prediction model is constructed with the training samples online. Because the training samples have the most similarity to the test data x_t, the reliability of the model is improved.

In the SS-JITL-SEC soft-sensor, the SEC method is more efficient than traditional clustering methods and avoids continually iterative training processes. At the same time, the mixed SS labeling method can correctly label all the unlabeled data, which are both normal data and outliers, more complete use of unlabeled data. Therefore, the SS-JITL-SEC soft-sensor cannot only translate the complex and nonlinear modeling problems into simple and linear problems but also enriches the information of training data used for modeling. Finally, the dissimilarity-based JITL constructs a local adaptive prediction model with the variations of test data, ensuring the high prediction performance of soft-sensor.

IV. Case Studies

The SS-JITL-SEC soft-sensor can be verified in this section by two case studies: a real full-scale wastewater plant and an oxidation ditch (OD) wastewater plant with field data. In two case studies, partial least squares (PLS) and multiple linear regression (MLR) work together with SS-JITL-SEC to construct the final prediction models. Both PLS and MLR belong to one of the most commonly useful linear regression algorithms, and they are simple, convenient as well as are appropriate for most of linear regression problems. However, they have high requirements for the process variables. For example, the output variables should have a strong linear relationship with the input variables. Thus, this would seriously limit their range of applicability.

To demonstrate the superiorities of the SS-JITL-SEC soft-sensor, the k-means clustering, standard SS labeling method and offline counterparts replace the SEC method, the mixed SS labeling and the dissimilarity-based JITL algorithm as comparisons, respectively, and other sections maintain unity.

The root means square error (RMSE) and correlation coefficient (R) are used to present the prediction performance. RMSE is the square root of the average of the squared errors between the real and predictive values and is used to describe the overall prediction error. RMSE and R are defined as follows:

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad i = 1,2, \ldots n \tag{15}$$

$$R(Y, \hat{Y}) = \frac{cov\ (Y, \hat{Y})}{\sqrt{var\ [Y]\ var\ [\hat{Y}]}} \tag{16}$$

where $Y = (y_1, y_2, \ldots, y_n)$ is the real value, $\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$ is the predictive value, n is the number of samples; cov is the covariance of Y and $\hat{Y}$; var is the variance.

To verify the result of industrial process monitoring, we make a Shewhart control chart to describe the performance of

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2022.3217028

monitoring [41]. These soft-sensors are constructed with

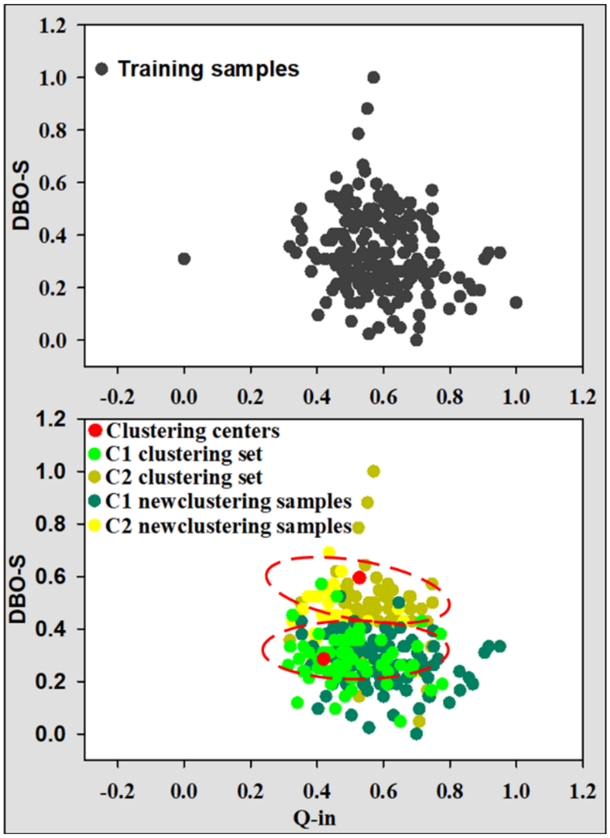

Fig. 4. The result of clustering and extending in the first case study

normal training data, so the residuals between real values and predictive values should comply with Gaussian distribution $N(\mu, \sigma^2)$ [42]. In this paper, absolute values of the residuals are taken as the final values. Then, according to different monitoring requirements, the upper limitations of residual, σ, 2σ and 3σ are determined. By tracking and comparing residual curves, achieve industrial process monitoring.

A. A Real Full-scale Wastewater Plant

1) Background: In this case study, the data were collected from a real full-scale activated sludge-based wastewater plant. Organic matter and nutrients were removed by the wastewater plant and corresponding technological process is shown in Fig. 2. Due to the complex biochemical process, the data exhibit highly dynamical behaviors. Sequentially, these dynamic data add more complexity to the process monitoring. In the collected dataset, there are 400 samples and the sample rate is 1 day. Each set of data consists of 38 process variables. The detailed description of the variables can be found in [43]. The first 200 sets of data are labeled data for modeling, and the last 100 sets of data are used as test data. For the remaining 100 sets of data, they are used as unlabeled data.

2) Prediction Results and Analysis: The prediction results concerning RMSE and R are tabulated in Table I, and the main purpose is to compare the prediction performances of different soft-sensors. Firstly, by comparing the SS-JITL-SEC with the k-means clustering algorithm, both SS-JITL-SEC PLS and SS-JITL-SEC MLR can achieve better prediction results. The main reason is that training data have the best similarity in the clustering sub-dataset of the SEC method, which is beneficial to simplify the modeling problem, and linear models can approach the dataset locally. Then, the standard SS method is used as a comparison baseline to demonstrate the advantage of mixed SS labeling method. Obviously, the prediction results of SS-JITL-SEC soft-sensors are better than using the standard SS method, and the RMSE decreased 52.73% and 53.28%, respectively. This is because the mixed SS labeling method labels all the unlabeled data as new labeled data and considers the specific distribution scenarios of each unlabeled data, and then estimates the output variable of each unlabeled data with proper approaches, leading to more new and reasonable labeled data available for rebuilding the prediction model. In addition, because the SS-JITL-SEC soft-sensor can be updated online, the prediction results are certainly better than the offline soft-sensors. Finally, we notice that the SS-JITL-SEC MLR achieves the best prediction results in terms of RMSE and R, 3.564 and 93.67%. This proves the MLR is more suitable for linear problems than the PLS in this case study.

To describe prediction performances more intuitively, the real and predictive curves of soft-sensors using MLR are presented in Fig. 3. As we can see, the predictive curve of SS-JITL-SEC MLR fits the real curve with the best performance, especially in the peaks and valleys. This mainly lies on the fact that the SEC method can select the most similar data including the peaks and valleys, which is beneficial for the prediction of

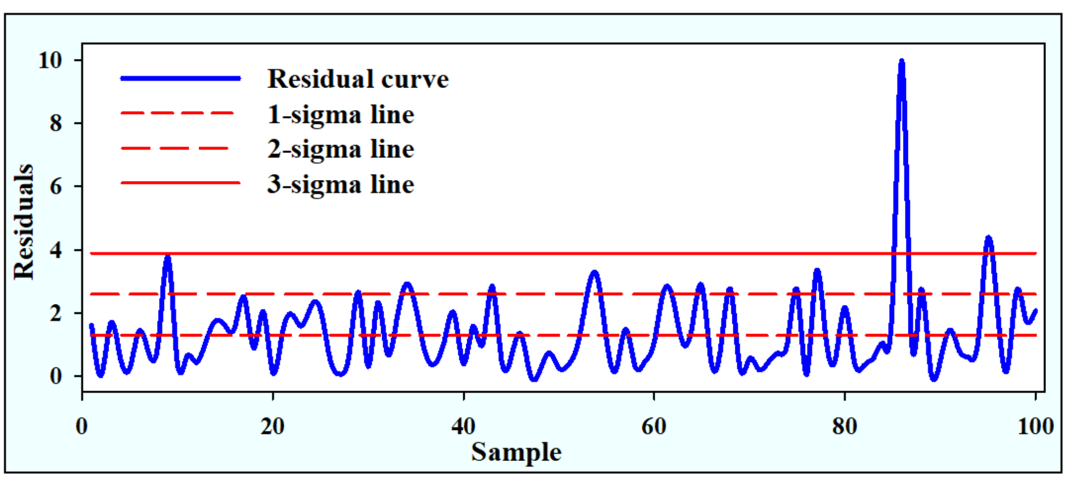

Fig. 5. The Shewhart control chart in the first case study

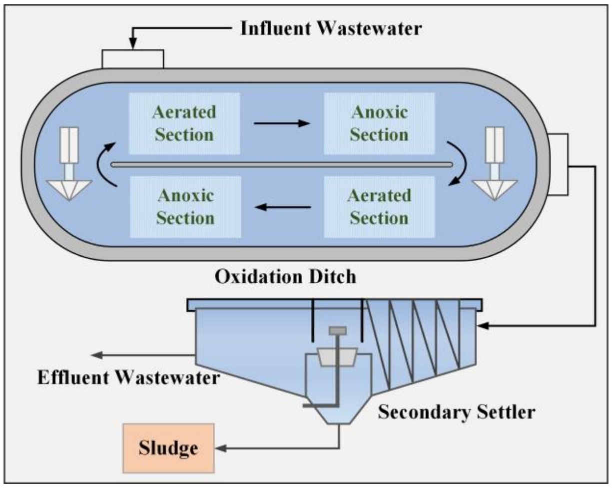

Fig .6. Schematic of the OD wastewater plant

peaks and valleys. Also, the mixed SS labeling method also enriches the samples information of the peaks and valleys. Finally, the dissimilarity-based JITL continuously updates the prediction model, enabling the soft-sensors to cope with abrupt noises and outliers, such as the peaks and valleys.

Fig. 4 profiles the change of training dataset after using the SS-JITL-SEC. In Fig. 4, the original training dataset is divided by using the SEC method into two clustering sub-datasets, C1 and C2. As we can see, the data from the same sub-dataset have the most similarity, which will be beneficial to reducing the difficulty of modeling. Then new labeled data are produced by the mixed SS labeling method and added into the original labeled dataset to extend C1 and C2. Thus, the information of training data is enriched implies the prediction performance of model can be improved. Finally, depending on the characteristics of test data, we use the dissimilarity-based JITL to select the proper clustering sub-dataset and construct a local adaptive prediction model, which would avoid degradation of prediction performance.

3) Process Monitoring: The Shewhart control chart is shown in Fig. 5 for industrial process monitoring. In Fig. 5, the blue line represents the residuals between the real and the predictive values, the three red lines are the upper control limits of residual to recognize the deviations. Depending on different attentions for process monitoring, σ, 2σ and 3σ represent the percentage of qualified data, 68.27%, 95.45% and 99.73%, respectively. As shown in Fig. 5, if samples come across the upper control limits, we must diagnose and deal with these problems in time, reduce failure loss. In this case study, the output variable, DBO represents the concentration of biological oxygen demand. The more DBO value is, the worse the quality of effluent is. Therefore, predicting and monitoring DBO values as well as the residuals between the predictive and real values is significant for achieving reasonable process monitoring.

TABLE II
RMSE AND R OF THE PREDICTION RESULTS IN THE SECOND CASE STUDY

Models	RMSE	R	Models	RMSE	R
k-means PLS	0.0086	97.73%	**k-means MLR**	0.0090	97.55%
SS PLS	0.0087	97.83%	**SS MLR**	0.0098	97.67%
Offline PLS	0.0090	97.81%	**Offline MLR**	0.0102	97.64%
SS-JITL-SEC PLS	0.0064	98.25%	**SS-JITL-SEC MLR**	**0.0067**	**98.14%**

B. An Oxidation Ditch (OD) Wastewater Plant

1) Background: The second case is an OD wastewater plant. OD process is a modified activated sludge biological treatment process. Fig. 6 shows a schematic of the reactor for the plant. It consists of an OD and a secondary sedimentation tank. The aerated section and anoxic section are alternate in the OD. In this case study, the effluent biological oxygen demand (BOD) is the output variable and the other 13 easy-to-measure process variables are the input variables. More details can refer to Table S I in the Supporting Information. In addition, 400 sets of data are collected. The first 200 samples in the data set and another 100 samples are labeled and unlabeled datasets, respectively. The remaining 100 samples is used as a test dataset.

2) Prediction Results and Analysis: The prediction results are shown in Table II in terms of RMSE and R. To prove the advantages of the SS-JITL-SEC soft-sensor in this case, the

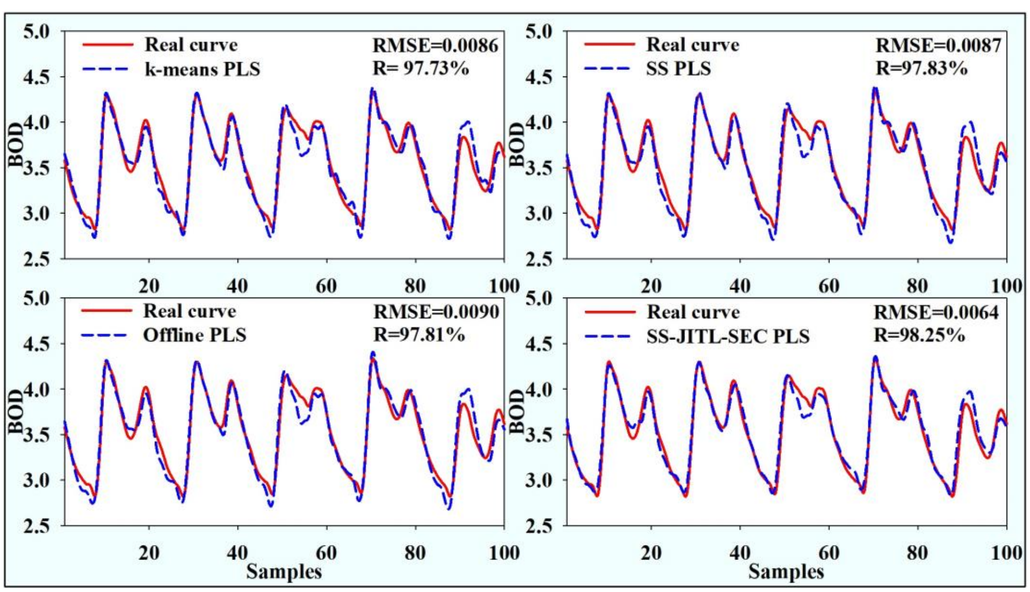

Fig. 7. The real and predictive curves in the second case study

prediction results of SS-JITL-SEC with PLS and MLR are compared with k-means clustering, standard SS labeling method and offline counterparts. In all the soft-sensors, the SS-JITL-SEC PLS soft-sensor achieves the best prediction with RMSE and R being 0.0064 and 98.26%, respectively. Compared with using k-means clustering, standard SS labeling method and offline counterparts with PLS, RMSE value decreases by 34.38%, 35.94% and 43.33%. Firstly, this proves the resulted clustering sub-dataset with the SEC method can

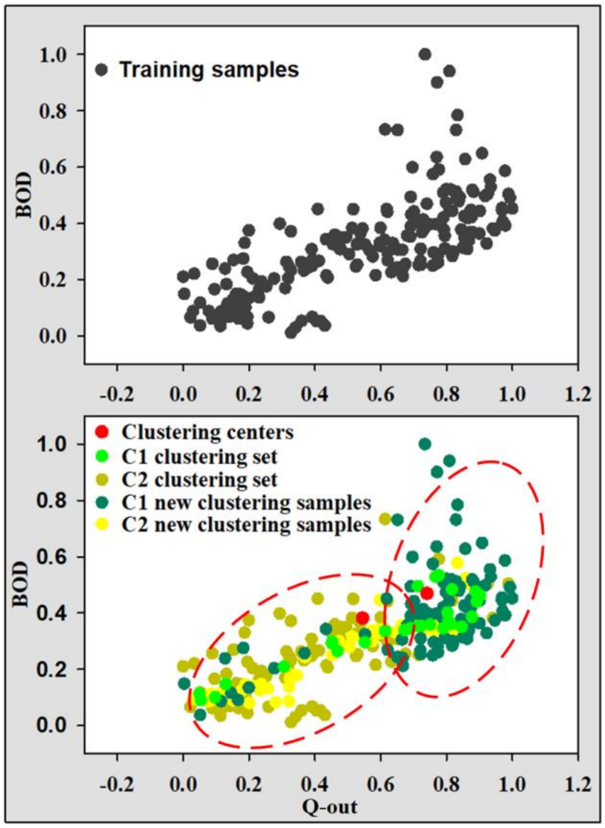

Fig. 8. The result of clustering and extending in the second case study

acquire more similar data than k-means clustering. Then, the mixed SS labeling method extends the original training dataset by labeling all the unlabeled data with more proper approaches than the standard SS labeling method. Finally, the dissimilarity-based JITL can enhance the adaptive capability and sequentially reduce performance deterioration during prediction processes.

Fig. 7 shows the fits between real and predictive curves with respect to the prediction results of soft-sensors using PLS. As demonstrated in Fig. 7, the SS-JITL-SEC PLS soft-sensor fits the real curve better than using k-means clustering, standard SS labeling method and offline counterparts with PLS, especially the peaks and valleys. This is because the SS-JITL-SEC PLS soft-sensor fully mines the information of peaks and valleys in the unlabeled data to enrich the original training samples. At the same time, it selects the sub-dataset with the highest correlation with peaks and valleys in the monitoring process, and then builds a local adaptive prediction model with dissimilarity-based JITL. Thus, the SS-JITL-SEC PLS soft-sensor can trace these abnormal data better in this case study.

The detail change of the training dataset for the SS-JITL-SEC PLS soft-sensor are shown in Fig. 8. As we can see, the original training dataset is divided into two high-quality clustering sub-datasets, C1 and C2 with the SEC method. Each clustering sub-dataset shares the best similarity among data. Then, they are extended by the new labeled data which are produced by the mixed SS labeling method and share highly similar to original labeled data. Finally, a local adaptive prediction model is built for the test data with the most proper clustering sub-dataset online. Therefore, the SS-JITL-SEC PLS soft-sensor can achieve the best prediction results than others due to its adaptive ability.

3) Process Monitoring: The Shewhart control chart is shown in Fig. 9. Similar to the first case study, the blue line represents the residuals between the real and the predictive values. According to different attentions for process monitoring, three red lines, σ, 2σ and 3σ, are used as the upper control limits of residual to recognize the deviations. The residuals obey the normal distribution, so they represent the percentage of qualified data, 68.27%, 95.45% and 99.73%. As shown in Fig. 9, if samples come across the upper control limits, we must diagnose and deal with these problems in time. This information will be utilized to avoid serious industrial processes. In this case study, the larger the BOD values are, the worse the effluent quality. This is because the output variable, BOD, represents the oxidation reaction efficiency of microorganisms.

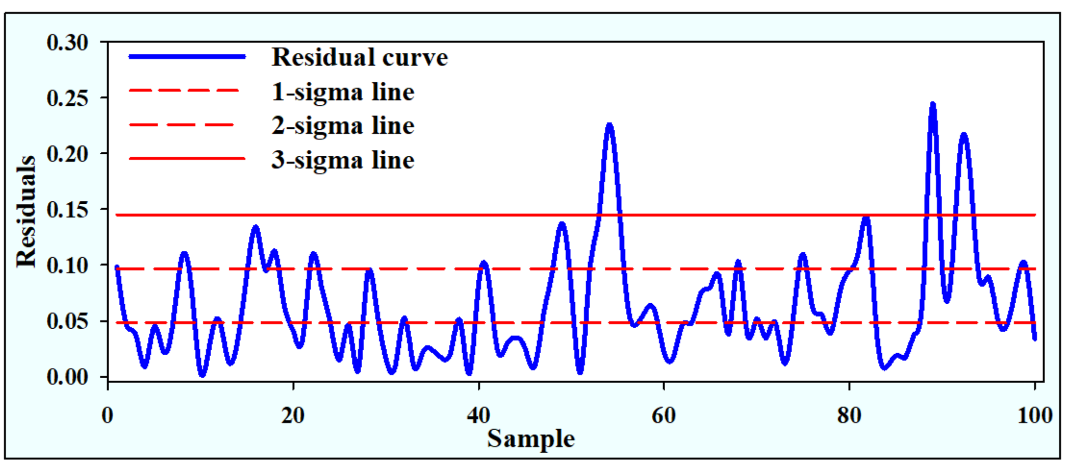

Fig. 9. The Shewhart control chart in the second case study

Therefore, process monitoring can be achieved by predicting and monitoring the BOD value as well as the residuals between the predictive and real values.

V. Discussions

This paper focuses on enhancing the prediction performance of standard soft-sensors for industrial processes monitoring. With industrial processes becoming more and more complex, high-precision soft-sensors cannot achieve easily for some difficult-to-measured variables. Clustering methods can provide a powerful alternative to address the high-nonlinear and high-dimension data issues. In this paper, unlike the traditional clustering methods, the dissimilarity-based SEC method is able to explore the similarity between high-dimension samples, acquiring resulted clustering sub-datasets and simplifying the modeling problems. Then, the mixed SS labeling method is proposed to deal with the imbalance of labeled and unlabeled data. This method considers the specific distribution scenarios of each unlabeled data, and then estimates the output by the surrounding labeled data or training models. Finally, a local adaptive prediction model is constructed with the dissimilarity-based JITL to adapt to the change of test data. To verify the effectiveness of the proposed soft-sensors, two case studies are provided. In the real full-scale activated sludge-based wastewater plant, the sampling period is so long that the collected data are dramatically dynamic. However, in the OD-based wastewater plant, 20 samples are collected in a day, thus they are easier to predict and trace. In two case studies, the SS-JITL-SEC soft-sensor has better prediction performance than other soft-sensors, such as using the k-means clustering, standard SS labeling method and offline standard counterparts. Finally, we further validated the monitoring performance of the SS-JITL-SEC soft-sensor though Shewhart control chart, and obtained the intuitive monitoring results.

However, there are still some disadvantages to the proposed soft-sensors. Firstly, the prediction accuracy of peaks and valleys is still terrible and requires further improvement. Since they have important significance in industrial processes, it is necessary to achieve more accurate predictions [44]. We can use the time difference (TD) method to improve the prediction accuracy for abrupt noises and outliers in the future. Secondly, even though the difficulty of modeling is reduced and the prediction performance is improved, the SS-JITL-SEC soft-sensor breaks the continuity of samples. When facing a larger continuous sample set, errors may be accumulated and even lead to the failure of process monitoring [45]. Therefore, establishing an appropriate prediction model with continuous samples and ensuring prediction performance of the peaks and valleys in the time series and large dataset can be an important direction of future research.

VI. Conclusions

This paper proposes a novel semi-supervised soft-sensor, which is learned by a Just-in-time method together with structure entropy clustering (SS-JITL-SEC). The proposed soft-sensor can achieve the best performance compared with standard soft-sensors in terms of RMSE and R, even further to use for monitoring the high-nonlinear, high-dimension and imbalance data issues. The studies showed that the SS-JITL-SEC soft-sensor can simplify the complex and nonlinear modeling problems with a divide and conquer strategy. Also, the entire unlabeled data are fully used to enrich original training samples and to better semi-supervised soft-sensor learning. Finally, a local adaptive model has been constructed with the resulted clustering sub-dataset to ensure prediction quality and adapt to significant changes in the test dataset. However, since the auto-correlativity and cross- correlativity in the time-series samples is broken by the SEC method, the proposed soft-sensors could be inappropriate for the time series and large datasets modeling. Therefore, future research will concentrate on addressing the time series and large dataset prediction issues of industrial processes and further using for process monitoring and diagnosis.

Appendix

Table S I
Process variables in the second case study

No	Variable comments	Variables
1	Oxygen uptake rate-2	OUT-2
2	Oxygen uptake rate-4	OUT-4
3	Dissolved oxygen-sludge	DO-s
4	Heterotrophic biomass-influent	XBH-in
5	Inert particulate material- influent	XI-in
6	Biodegradable substrate-influent	XS-in
7	Ammonium-sludge	NH4-s
8	Dissolved oxygen-out	DO-out
9	Ammonium-out	NH4-out
10	Flow rate-out	Q-out
11	Total kjeldahl nitrogen-out	TKN-out
12	Total suspended solids-out	TSS-out
13	Volatile suspended solids-out	VSS-out
14	**Biochemical oxygen demand**	**BOD**

References

[1] C. Xu, D. Huang, D. Li and Y. Liu "Novel Process Monitoring Approach Enhanced by a Complex Independent Component Analysis Algorithm with Applications for Wastewater Treatment," Industrial & Engineering Chemistry Research, vol. 60, no. 38, pp. 13914-13926, 2021.

[2] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," Chemometrics and Intelligent Laboratory Systems, vol. 171, pp. 16-25, 2017.

[3] H. Chen, B. Jiang, S. Ding and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," IEEE Transactions on Intelligent Transportation Systems, 2020.

[4] H. Luo, K. Li, O. Kaynak, S. Yin, M. Huo, and H. Zhao, "A robust data-driven fault detection approach for rolling mills with unknown roll eccentricity," IEEE Transactions on Control Systems Technology, vol. 28, no. 6, pp. 2641-2648, 2019.

[5] W. Yan, D. Tang, and Y. Lin, "A data-driven soft sensor modeling method based on deep learning and its application," IEEE Transactions on Industrial Electronics, vol. 64, no. 5, pp. 4237-4245, 2016.

[6] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor

modeling with variable-wise weighted SAE," IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3235-3243, 2018.

[7] H. Luo, S. Yin, T. Liu, and A. Khan, "A data-driven realization of the control-performance-oriented process monitoring system," IEEE Transactions on Industrial Electronics, vol. 67, no. 1, pp. 521-530, 2019.

[8] L. Yao, and Z. Ge, "Refining data-driven soft sensor modeling framework with variable time reconstruction," Journal of Process Control, vol. 87, pp. 91-107, 2020.

[9] D. Li, D. Huang and Y. Liu, "A novel two-step adaptive multioutput semisupervised soft sensor with applications in wastewater treatment," Environmental Science and Pollution Research, pp. 1-15, 2021.

[10] L. Yao, and Z. Ge, "Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application," IEEE Transactions on Industrial Electronics ,vol. 65, no. 2, pp. 1490-1498, 2017.

[11] J.-C. Ren, D. Liu, and Y. Wan, "Model-Free Adaptive Iterative Learning Control Method for the Czochralski Silicon Monocrystalline Batch Process," IEEE Transactions on Semiconductor Manufacturing, 2021.

[12] M. Xu, M. Li, W. Xu, Z. Deng, Y. Yang, and K. Zhou, "Interactive mechanism modeling from multi-view images," ACM Transactions on Graphics (TOG), vol. 35, no. 6, pp. 1-13, 2016.

[13] Z. Zeng, X. Zhang, and X. Li, "Layout-dominated dynamic current imbalance in multichip power module: Mechanism modeling and comparative evaluation," IEEE Transactions on Power Electronics, vol. 34, no. 11, pp. 11199-11214, 2019.

[14] H. Chen, and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 2, pp. 450-465, 2019.

[15] Y. Liu, and M. Xie, "Rebooting data-driven soft-sensors in process industries: A review of kernel methods," Journal of Process Control, vol. 89, pp. 58-73, 2020.

[16] K. B. Newhart, R. W. Holloway, A. S. Hering, and T. Cath, "Data-driven performance analyses of wastewater treatment plants: A review," Water research, vol. 157, pp. 498-513, 2019.

[17] M. Huo, H. Luo, X. Wang, Z. Yang, and O. Kaynak, "Real-time Implementation of Plug-and-Play Process Monitoring and Control on an Experimental Three-Tank System," IEEE Transactions on Industrial Informatics, 2020.

[18] H. Chen, Z. Chen, Z. Chai, B. Jiang, and B. Huang, "A Single-Side Neural Network-Aided Canonical Correlation Analysis With Applications to Fault Diagnosis," IEEE Transactions on Cybernetics, 2021.

[19] Y. Liu, Y. Pan and D. Huang, "Development of a novel adaptive soft-sensor using variational Bayesian PLS with accounting for online identification of key variables," Industrial & Engineering Chemistry Research, vol. 54, no. 1, pp. 338-350, 2015.

[20] B. Bidar, J. Sadeghi, F. Shahraki, M. M. Khalilipour, "Data-driven soft sensor approach for online quality prediction using state dependent parameter models," Chemometrics and Intelligent Laboratory Systems, vol. 162, pp. 130-141, 2017.

[21] S. Shin, R. H. Osgouei, K.-D. Kim, and S. Choi, "Data-driven modeling of isotropic haptic textures using frequency-decomposed neural networks." IEEE World Haptics Conference (WHC), pp. 131-138., 2015

[22] Y. Liu, D. Huang, B. Liu, Q. Feng, and B. Cai, "Adaptive ranking based ensemble learning of Gaussian process regression models for quality-related variable prediction in process industries," Applied Soft Computing, vol. 101, pp. 107060, 2021.

[23] D. Li, Y. Liu and D. Huang, "Development of semi-supervised multiple-output soft-sensors with Co-training and tri-training MPLS and MRVM," Chemometrics and Intelligent Laboratory Systems, vol. 199, pp. 103970, 2020.

[24] H. Mohamed, A. Negm, M. Zahran and O. Saavedra, "Bathymetry determination from high resolution satellite imagery using ensemble learning algorithms in Shallow Lakes: Case study El-Burullus Lake," International Journal of Environmental Science and Development, vol. 7, no. 4, pp. 295, 2016.

[25] X. Sun, S. Wang, and M. Liu, "A Novel Coding Architecture for Multi-Line LiDAR Point Clouds Based on Clustering and Convolutional LSTM Network," IEEE Transactions on Intelligent Transportation Systems, 2020.

[26] Y. J. Chun, M. O. Hasna, and A. Ghrayeb, "Modeling heterogeneous cellular networks interference using Poisson cluster processes," IEEE Journal on Selected Areas in Communications, vol. 33, no. 10, pp. 2182-2195, 2015.

[27] O. Oyelade, O. O. Oladipupo, and I. Obagbuwa, "Application of k Means Clustering algorithm for prediction of Students Academic Performance," arXiv preprint arXiv:1002.2425, 2010.

[28] Y. Lu, Z. Tian, P. Peng, J. Niu, W. Li and H. Zhang, "GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system," Energy and Buildings, vol. 190, pp. 49-60, 2019.

[29] Y. Zhang, J. Wang, and B. Chen, "Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach," IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 623-634, 2020.

[30] T. Liu, Y. Yang, G.-B. Huang, Y. K. Yeo, and Z. Lin, "Driver distraction detection using semi-supervised machine learning," IEEE transactions on intelligent transportation systems, vol. 17, no. 4, pp. 1108-1120, 2015.

[31] D. Li, D. Huang, G. Yu, and Y. Liu, "Learning adaptive semi-supervised multi-output soft-sensors with co-training of heterogeneous models," IEEE Access, vol. 8, pp. 46493-46504, 2020.

[32] X. Li, Q. Sun, Y. Liu, Q. Zhou, S. Zheng, T.-S. Chua, and B. Schiele, "Learning to self-train for semi-supervised few-shot classification," Advances in Neural Information Processing Systems, vol. 32, pp. 10276-10286, 2019.

[33] H. Tao, C. Hou, F. Nie, J. Zhu, and D. Yi, "Scalable multi-view semi-supervised classification via adaptive regression," IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4283-4296, 2017.

[34] D. Zhou, F. Gao, E. Breaz, A. Ravey, and A. Miraoui, "Degradation prediction of PEM fuel cell using a moving window based hybrid prognostic approach," Energy, vol. 138, pp. 1175-1186, 2017.

[35] J. Wu, H. Cheng, Y. Liu, B. Liu, and D. Huang, "Modeling of adaptive multi-output soft-sensors with applications in wastewater treatments," IEEE Access, vol. 7, pp. 161887-161898, 2019.

[36] X. Yuan, J. Zhou, Y. Wang, and C. J. J. o. C. Yang, "Multi-similarity measurement driven ensemble just-in-time learning for soft sensing of industrial processes," Journal of Chemometrics, vol. 32, no. 9, pp. e3040, 2018.

[37] B. Bein, "Entropy," Best Practice & Research Clinical Anaesthesiology, vol. 20, no. 1, pp. 101-109, 2006.

[38] A. Li, and Y. Pan, "Structural information and dynamical complexity of networks," IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3290-3339, 2016.

[39] Y. Liu, B. Liu, X. Zhao, and M. Xie, "Development of RVM-based multiple-output soft sensors with serial and parallel stacking strategies," IEEE Transactions on Control Systems Technology, vol. 27, no. 6, pp. 2727-2734, 2018.

[40] X. Yuan, Z. Ge, B. Huang, Z. Song, and Y. Wang, "Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR," IEEE Transactions on Industrial Informatics, vol. 13, no. 2, pp. 532-541, 2016.

[41] F. N. Mim, M. B. C. Khoo and S. Saha, "A side-sensitive group runs median control chart with measurement errors," Communications in Statistics-Simulation and Computation, pp. 1-19, 2021.

[42] H. D. Nguyen, Q. T. Nguyen and K. P. Tran, "On the performance of VSI Shewhart control chart for monitoring the coefficient of variation in the presence of measurement errors," The International Journal of Advanced Manufacturing Technology, vol. 104, no. 1, pp. 211-243, 2019.

[43] Y. Liu, B. Liu, X. Zhao, and M. Xie, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," IEEE Transactions on Industrial Electronics, vol. 65, no. 8, pp. 6478-6486, 2017.

[44] W. Shao, Z. Ge, Z. Song, and J. Wang, "Semisupervised robust modeling of multimode industrial processes for quality variable prediction based on Student's t mixture model," IEEE Transactions on Industrial Informatics, vol. 16, no. 5, pp. 2965-2976, 2019.

[45] T. Wang, C. Wang, Z. Xu, C. Cui, X. Wang, Z. Demitrack, Z. Dai, A. Bagtzoglou, M. D. Stuber, and B. Li, "Precise control of water and wastewater treatment systems with non-ideal heterogeneous mixing models and high-fidelity sensing," Industrial & Engineering Chemistry Research, vol. 60, no. 38, pp. 13914-13926 132819, 2021.

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2022.3217028

Dong Li was born in Kaifeng, Henan, China. He received the B.E. degree in computational mathematics from He'nan Normal University, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering with the Department of Automation, South China University of Technology, Guangzhou, China. His research interests include statistical theory research, machine learning, data-driven soft sensors, and semi-supervised learning in wastewater treatment.

Yiqi Liu was born in Haikou, China in 1983. He received B.S. and M.S. degrees in control engineering from the Chemical University of Technology, Beijing, in 2009 and a Ph.D. degree in control engineering from the South China University of Technology, Guangzhou, China in 2013. From 2013 to 2016, he was a lecturer at the South China University of Technology. Since 2016, he has been an Associate Professor in the Department of Automation at the South China University of Technology. He is Scenior IEEE member and Associate Editor of IEEE ACCESS, ISA Transactions. He is the author of more than 60 articles. His research interests include soft sensors, fault diagnosis, and wastewater treatment, and he holds three patents. He was a recipient of EU H2020 Marie Skłodowska-Curie Individual Fellowships in 2020, the Hong Kong Scholar Award in 2016, the Chinese Scholarship Council Award in 2011 and the Deutscher Akademischer Austausch Dienst Award in 2015.

Daoping Huang received a B.Eng. degree in chemical automation and instruments and M.Eng. and Ph.D. degrees in automatic control theory and applications from the South China University of Technology (SCUT), Guangzhou, China in 1982, 1986, and 1998, respectively. From 1995 to 1996, he worked at the University of Gent in Belgium as a government sponsored visiting scholar. In 1982, he started his teaching and research career with the Department of Automation at SCUT. Now, as a Full Professor of the School of Automation Science and Engineering, he has authored three academic books and more than 200 conference and journal papers. His research interests cover intelligent detection, control, soft-sensing technology as well as fault diagnosis and accident prediction for industrial processes.

Dr. Huang serves as the Vice Director of the Education Committee and a member of the Process Control and Application Committees at the Chinese Association of Automation. He has directed over 10 research projects. He was granted a Third Prize from the National Education Committee in 1992 and a Second Prize from the Guangdong Provincial Government in 2005 for his contributions to science and technology.

Chong xu received a B.E degree in measurement-control technology and instrument specialty from Jiangxi University of Science and Technology (JXUST), in 2010. He then got a M.E degree in control theory and engineering from South China University of Technology (SCUT), in 2013. Form 2013 till then, Xu serves as a lecturer of the School of Physics& Electronic Information at Gannan Normal University (GNNU). Now, He is pursuing a Ph.D degree in control theory and engineering from school of Automation Science and Engineering, SCUT. His research interest includes process monitoring, making learning, fault detection& diagnosis in wastewater treatment.