IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 22, 2020, accepted April 15, 2020, date of publication April 20, 2020, date of current version May 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2988668

LSTM Soft Sensor Development of Batch Processes With Multivariate Trajectory-Based Ensemble Just-in-Time Learning

FEIFAN SHEN[1], JIAQI ZHENG[2], LINGJIAN YE[1], (Member, IEEE), AND XIUSHUI MA[1]

[1]School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China
[2]School of Mechanical Engineering and Automation, College of Science and Technology, Ningbo University, Ningbo 315300, China

Corresponding author: Feifan Shen (ffshen@nit.net.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61673349, in part by the Natural Science Foundation of Zhejiang Province under Grant LQ19F030004, in part by the Scientific Research Fund of Zhejiang Provincial Education Department under Grant Y201839765, and in part by the Foundation of Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, under Grant APCLI1804.

ABSTRACT To implement the quality prediction scheme for batch processes, long short-term memory (LSTM) neural network is a feasible tool to handle with the process dynamics and nonlinearity. However, a global LSTM soft sensor suffers a decline in performance facing batch-to-batch variations. To overcome the batch diversity problem and take advantage of LSTM model, a multivariate trajectory based ensemble just-in-time learning strategy is proposed in this paper. Different trajectory based similarity measurements are designed to extract historical batch trajectories which share similar spatial positions and trends. For each selected trajectory, an online local LSTM soft sensing model is constructed and the real-time quality prediction result for each local model can be obtained. Then, a weighting parameter is determined for each model by cross validation. Bringing together quality prediction results from different local models, the ensemble prediction result can be finally figured out. Two case studies are carried out to prove the effectiveness of the proposed methodology including a fed-batch reactor and the fed-batch penicillin fermentation process.

INDEX TERMS Batch production systems, Ensemble just-in-time learning, long short-term memory, multivariate trajectory analysis, soft sensor, quality prediction.

I. INTRODUCTION

Nowadays, the proportion of batch processes in modern industry is increasing rapidly due to the growing demand of high-value-added products (e.g., food, pharmaceuticals, semiconductors, polymers, etc.) [1]–[5]. It is noted that the quality of batch process products should be paid extreme attention to, although online quality variables are usually difficult to be measured. Hence, the quality prediction of batch processes becomes a significant task of process industry. Meanwhile, the real-time property of quality prediction is also required to avoid the control failure and quality corruption during productions [6]–[8]. To achieve the quality prediction scheme of batch processes, data-driven soft sensor techniques are developed by making use of the process data [9]–[14]. Mathematical models are established with historical data to extract the latent relationship between ordinary process variables and quality variables. For online quality prediction purpose, real-time measurement of the easy-to-measure process variables is implemented to estimate the critical quality variables through the established prediction model. There are many kinds of data-driven soft sensors of batch processes, such as partial least squares (PLS) [15], support vector regression (SVR) [16], Gaussian process regression (GPR) [17] and artificial neural network (ANN) methods [18], [19]. PLS is developed to extract the correlations between process input and output with a linear model. To solve the problem of data nonlinearity, SVR and ANN based soft sensors are designed with the idea of kernel function and nonlinear mapping, respectively. Considering the limited process data, GPR is used to implement adaptive modeling and quality prediction scheme. In summary, these existing soft sensing models mainly focus on the within-batch characteristics of complicated batch processes.

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

However, the aforementioned soft sensors are under the assumption that the data of different batches used for modeling are identical, which means these batches share the similar trajectories without prominent batch-to-batch variations. In light of this framework, the global monitoring strategy can be carried out to generate an overall integrated soft sensor model. When faced with practical targets such as parallel-running processes, the gap between the ideal model and the run-to-run batch variations become a formidable challenge [20], [21]. As a result, model updating is necessary referring to different types of process data and the modeling step is no longer restricted to the offline stage. Phase/stage based data-driven modeling techniques are developed to handle with this issue where the range of modeling data is limited to the local samples. Besides, online modeling strategies are also suggested to follow the changing trajectory of the real-time batch data. To implement online local modeling, just-in-time learning (JITL) algorithm is proposed as a useful online tool to extract local similar samples, which is also known as locally weighted learning and lazy learning [22]–[25]. The main idea of JITL is conducting the real-time similarity measurement between the current query sample and historical samples and sorting by similarity indices. To take the advantage of JITL, may relevant soft senor methods are proposed by scholars to deal with quality prediction and monitoring tasks. Three steps are designed for a typical JITL based soft sensor, which consist of similar sample extraction, online local modeling and quality prediction.

It is worth mentioning that the criterion of the JITL similar measurement varies in terms of different literatures. One of the most widely used strategy is to measure both the Euclidean distance and angle between two samples, where the weighted sum of these two similarity values are regarded as the similarity index of the current sample. However, there is no acknowledged similarity measurement performing effectively under various process conditions.

To make JITL adapt to different circumstances, the ensemble strategy is adopted to integrate multiple similarity measurements. Recently, an ensemble JITL (EJITL) technique is proposed, which preserves the advantage of the traditional JITL and improves the robustness of similar sample extractions [26]. In the EJITL framework, several similarity measurements are carried out simultaneously to obtain individual online local prediction models. Hence, the prediction results can be integrated with ensemble learning strategies. EJITL has been successfully applied on the online local modeling and quality prediction of continuous processes.

Different from continuous processes, batch processes behave more complicated than continuous processes, particularly due to the strong process dynamics and nonlinearity. As a result, the performance of the JITL/EJITL strategy used in continuous processes is limited since it only takes one single query sample for similarity measurements, where the feature of the current batch trajectory is ignored. Besides, a nonlinear dynamic soft sensor is necessary for the quality prediction of complex batch processes.

To address the aforementioned issues, a multivariate trajectory based ensemble JITL (TEJITL) technique is developed in this study. Meanwhile, long short-term memory (LSTM) is introduced as the nonlinear dynamic soft sensing method [27]–[29]. Firstly, multiple consecutive real-time samples of a batch process are used as the query samples. During the similarity calculation stage, three similarity measurements are adopted including the information of the distance, angle and trend between batch trajectories. For each individual measurement, historical trajectories with larger similarity measurements are collected as the online modeling samples. Hence, several LSTM soft sensor models can be constructed with the extracted batch trajectories and used for the quality prediction of online query samples. To integrate the quality prediction results of different sub-models, weighting parameters of different similarity measurements are defined and calculated based on the cross validation strategy. Finally, the weighted sum of each prediction result is judged as the ensemble result of the real-time batch trajectory. By the use of TEJITL-LSTM soft sensor, the issue of within-batch process nonlinearity and dynamics is firstly resolved owing to the nonlinear dynamic modeling framework. Meanwhile, the batch-to-batch variations are also taken into consideration with online local modeling strategy. A novel TEJITL framework is developed to implement online local modeling for batch processes with run-to-run variation. Combined with the nonlinear dynamic LSTM soft sensing model, ensembled quality prediction results can be obtained and the average prediction error is expected to be smaller than a single JITL technique based LSTM for batch processes.

The remainder of this paper is organized as follows. Section II offers a brief view of the basic approaches. The next section demonstrates the detailed methodology of the proposed soft sensor framework for batch processes, followed by two case studies as the verification of the proposed method. In the final section, concluding remarks are drawn.

II. PRELIMINNARIES

In this section, some preliminaries of the basic models used in this paper are demonstrated, which mainly consists of the fundamental knowledge of JITL technique and LSTM network.

A. JITL ALGORITHM

In general, traditional data-driven methods are mostly designed for global modeling, which is difficult to handle with multiple batch trajectories within a single global model. To this end, JITL is designed to provide an online local modeling strategy, which proves to be an effective model updating technique.

Assume a historical dataset $X_H \in \mathbb{R}^{m \times n}$ is given, where m is the variable index and n is the sample index. For an online query sample $x_q \in \mathbb{R}^m$, JITL aims at the extraction of the most similar samples in X_H compared with x_q. A fixed similarity measurement is conducted as $s_i = f\left(x_q, x_i\right)$, where $i = 1, 2, \ldots, n$. For most similarity measurements, a larger simi-

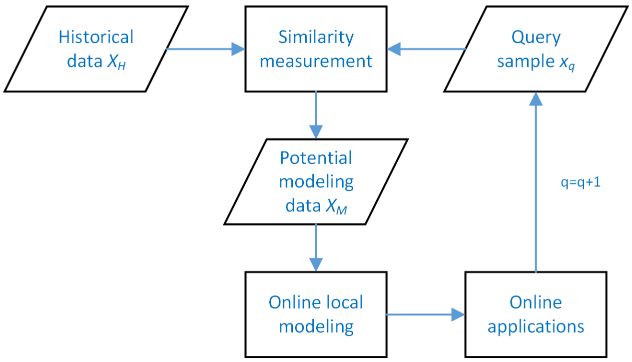


FIGURE 1. The flow diagram of the typical JITL procedure.

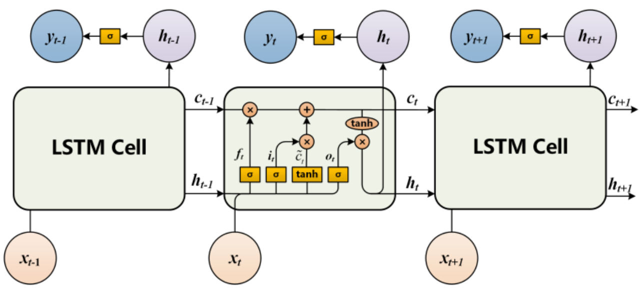


FIGURE 2. The structure of LSTM neural network.

larity value indicates the higher degree of similarity between two samples. Hence, the historical samples can be sorted according to the similarity measurement in the descending order. Therefore, a predefined number of historical samples with the largest similarity measurements are collected as the potential modeling data X_M. Then, a specific online local model can be constructed with the selected modeling data and further utilized to accomplish a practical task (e.g., process monitoring, process control, quality prediction, etc.). The typical procedure of JITL is presented in Fig.1.

B. LONG SHORT-TERM MEMORY NEURAL NETWORK

As an improved neural network based on the recurrent neural network (RNN), LSTM provides a modified structure by designing several "gates" in its basic unit which is named as "cell". The purpose of these gates is to capture both the short-term information and the long-term memory along the time index. LSTM preserves the advantage of RNN and reveals a better performance when handling with time-series data [30]. Therefore, LSTM has been adopted for nonlinear dynamic modeling on plentiful applications. The detailed structure of the LSTM structure is illustrated in Fig.2.

For each time instant t, an LSTM cell is constructed to establish the link along the time index between the sample of time instant $t-1$ and $t+1$. Moreover, at the variable layer, the cell input x_t is connected to a latent variable h_t. Then, a nonlinear mapping between h_t and the cell output y_t are carried out. It can be easily inferred that the crucial issue in training the LSTM model is to clarify the inner structure and parameters of the cell. In other words, the major task of LSTM is to obtain h_t on the basis of x_t and deliver useful dynamic information to the next cell. To resolve these problems, the following gates are developed. The input gate i_t of the LSTM cell can be illustrated as

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{1}$$

where σ in (1) represents the sigmoid activation function that $\sigma(x) = 1\big/(1+e^{-x})$; W_{xi}, W_{hi} are weighting parameters and b_i is the bias. Then, a tanh activation function is designed as $\tilde{c}_t$ to capture the necessary part of the system input, which is defined as:

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{2}$$

where W_{xc}, W_{hc} are weighting parameters and b_c denotes the bias. To determine whether the long-term memory should be remained from the previous cells, the forget gate is defined as

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{3}$$

where W_{xf}, W_{hf} are weighting parameters and b_f is the bias. Hence, the cell state c_t can be developed by combing the weighted cell input and the remaining information of the long-term memory, which can be described as

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{4}$$

where $\odot$ indicates the pointwise multiplication. To create the connection between the cell state and the hidden latent state, the output gate is defined as

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{5}$$

where W_{xo}, W_{ho} are weighting parameters and b_o denotes the bias. Based on c_t and o_t, h_t can be calculated as

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

Finally, the current value of system output can be estimated according to the latent cell state as

$$\hat{y}_t = \sigma\left(W_y h_t + b_y\right) \tag{7}$$

where W_y denotes the weighting parameter and b_y is the output bias.

In summary, the complete forward pass network structure of LSTM has been demonstrated referring to (1) to (7). To train a LSTM neural network, the back propagation through time (BPTT) can be used, which is demonstrated in ***Appendix A***.

III. METHODOLOGY

In this section, the detailed methodology of the soft sensor development is demonstrated and discussions are held.

A. MULTIVARIATE TRAJECTORY BASED EJITL

1) DATA UNFOLDING

According to Fig. 1, the historical dataset should firstly be determined before the implementation of EJITL. For continuous processes, the raw data $X_C \in \mathbb{R}^{M\times N}$ are two-dimensional and convenient for similarity measurements. On the contrary,

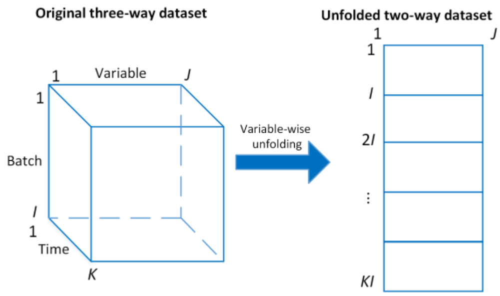


FIGURE 3. The unfolding way of batch dataset.

the original data of batch processes are three-way data, which can be denoted as $X_B \in \mathbb{R}^{I\times J\times K}$ including the batch index I, variable index J, and time index K. Therefore, a data unfolding procedure is necessary to convert the original data to the two-way form for similarity measurements. The unfolding way is illustrated in Fig. 3.

According to Fig. 3, the three-way data is unfolded with the variable-wise strategy and an unfolded two-way dataset can be obtained as $X_H \in \mathbb{R}^{J\times IK}$. Then, data normalization can be conducted along the variable index for further EJITL procedures.

2) SIMILARITY MEASUREMENTS

Given a fixed number of consecutive online query samples $x_q, x_{q+1}, \ldots x_{q+n-1}$ which have been normalized, the key task of the TEJITL is to extract the most similar historical trajectories for online local modeling. Firstly, the historical data are synchronized with the query sample length. For example, the h-th historical batch trajectory can be represented as $x_h, x_{h+1}, \ldots x_{h+n-1}$. Thus, totally $IK-n+1$ historical batch trajectories are stored for similarity measurements. In this study, three different similarity measurements are designed to evaluate the similarity degree of two batch trajectories.

The first JITL strategy is the Euclidean distance measurement. The distance based measurement can be denoted as

$$d_{h,i} = \sqrt{(x_h - x_i)^T \Delta^{-1} (x_h - x_i)}, \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-1) \tag{8}$$

$$s_{h,i} = e^{-d_{h,i}/\omega}, \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-1) \tag{9}$$

$$s_h = \sum_{i=q}^{q+n-1} s_{h,i} \Big/ n \tag{10}$$

where x_i denotes the normalized online sample; x_h denotes the normalized historical sample; Δ is a diagonal matrix with variances of each variable of the original historical input; $d_{h,i}$ is the normalized distance between the two samples. Therefore, the similarity measurement between every two samples can be defined as $s_{h,i}$, where ω is an adjustment parameter of the exponential function. Finally, the similarity measurement s_h between the online trajectory and the h-th historical batch trajectory can be calculated as the mean value of the sample similarity measurements.

Note that the distance is not the only explanation of the sample similarity. Therefore, the second JITL strategy takes the information of the angle into consideration as a supplement of the distance measurement. Given the same dataset as the first measurement, an additional angle measurement is defined based on the distance measurement as [23]

$$d_{h,i} = \|x_h - x_i\|_2 \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-1) \tag{11}$$

$$\cos(\theta_{h,i}) = \frac{\langle x_h, x_i \rangle}{\|x_h\|_2 \, \|x_i\|_2} \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-1) \tag{12}$$

$$s_{h,i} = \begin{cases} \lambda e^{-d_{h,i}/\omega} + (1-\lambda)\cos(\theta_{h,i}), & \cos(\theta_{h,i}) \geq 0 \\ 0, & \cos(\theta_{h,i}) < 0 \end{cases} \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-1) \tag{13}$$

$$s_h = \sum_{i=q}^{q+n-1} s_{h,i} \Big/ n \tag{14}$$

The value of $\cos(\theta_{h,i})$ indicates the angle between the online sample and the historical sample. In addition, samples with an extremely large angle will result in a negative result of $\cos(\theta_{h,i})$. Under such a situation, the similarity value is automatically adjusted to 0. Compared to the first measurement, it defines a trade-off parameter λ to evaluate the weight of the angle measurement. In general, the angle measurement is useful when facing several equal-distance trajectories. Historical trajectories with smaller angle compared to the online trajectory will have the priority for local modeling.

Apart from the information of the distance and angle, the trend of the batch trajectory should be evaluated as well. The previous two measurements mainly judge the similarity degree in terms of the static spatial coordinates of batch trajectories and neglect the information of the time-varying change to some extent. To find the solution of the problem, the third similarity measurement is proposed to compare the trend of two batch trajectories, which is defined as [23]

$$d_{h,i} = \|(x_{h+s} - x_h) - (x_{i+s} - x_i)\|_2, \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-s-1) \tag{15}$$

$$s_{h,i} = e^{-d_{h,i}/\omega}, \quad (h = 1, 2, \ldots, IK-n+1; \; i = q, q+1, \ldots, q+n-s-1) \tag{16}$$

$$s_h = \sum_{i=q}^{q+n-s-1} s_{h,i} \Big/ (n-s-1) \tag{17}$$

By the calculation of the difference between the nearby samples with a fixed step size s, the trajectory trend can be described and compared as a novel similarity measurement. Thus, the time-varying feature of batch trajectories is able to be captured for similarity measurements.

3) ENSEMBLE QUALITY PREDICTION

To implement ensemble quality prediction, several individual soft sensing models should be firstly constructed based on the results of similarity measurements. Several historical batch trajectories with the largest similarity measurements s_h are extracted as the modeling samples of the online local LSTM soft sensor for each strategy. Thus, there are a number of separated soft sensing models, which is determined by the types of similarity measurements and the number of selected trajectories of different strategies.

Next, the easy-to-measure process variables of a selected batch trajectory at time instant t are extracted as the LSTM input x_t referring to (1)-(5), while the quality variables are regarded as the LSTM output y_t. Therefore, the modeling dataset can be denoted as $X_{H,b} = \{x_{t,b}, x_{t+1,b}, \ldots, x_{t+n-1,b}\}$ and $Y_{H,b} = \{y_{t,b}, y_{t+1,b}, \ldots, y_{t+n-1,b}\}$, where t is the first moment of the time index of the modeling samples and $b = 1, 2, \ldots, B$ is the batch index of all the extracted historical trajectories. Hence, the LSTM structure can be determined as well as the values of system input and output. According to the BPTT algorithm illustrated in ***Appendix A***, all the model parameters (W_*, b_*) can be estimated after the root mean squared error (RMSE) presented in (18) converges.

$$RMSE = \sqrt{\sum_{k=t}^{t+n-1} (\hat{y}_k - y_k) \Big/ s} \tag{18}$$

where $\hat{y}_k$ and y_k are the predicted value and the real value of the quality variable, respectively.

With the constructed LSTM prediction model of the b-th historical trajectory, the predicted output of the online samples can be directly calculated according to the forward pass network structure as $\hat{Y}_{q,b} = \{\hat{y}_{q,b}, \hat{y}_{q+1,b}, \ldots, \hat{y}_{q+n-1,b}\}$. As a result, totally B distinct predicting results can be figured out for a single online trajectory.

As an ensemble soft sensing method, the most important issue is how to integrate different prediction results and improve the performance of quality prediction. To this extent, the cross validation strategy is carried out for each established LSTM prediction model. The effectiveness of one model will be verified according to the testing results of all the $B-1$ batch trajectories. The RMSE through the cross validation can be denoted as

$$RMSE_{b,r} = \sqrt{\sum_{k=1}^{n} (\hat{y}_{k,b,r} - y_{k,b,r}) \Big/ s} \tag{19}$$

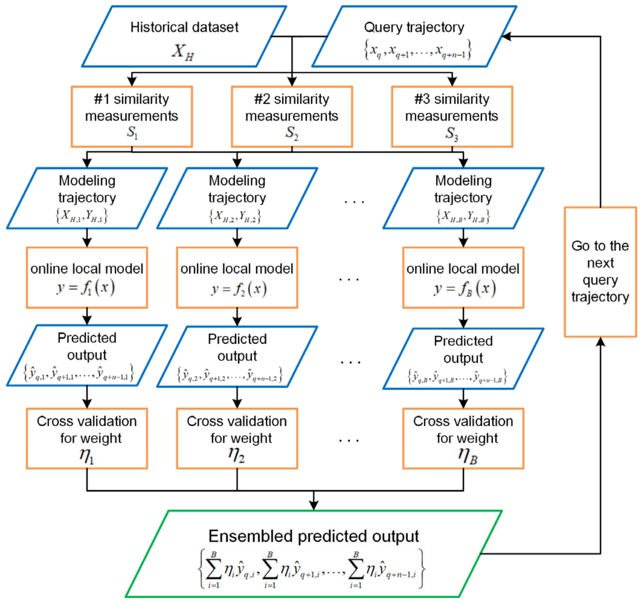


FIGURE 4. The flow diagram of the proposed methodology.

where $r = 1, 2, \ldots, B-1$ denotes other testing historical trajectories. Then, the average RMSE of the b-th historical modeling trajectory can be calculated as

$$RMSE_b = \frac{\sum_{r=1}^{B-1} RMSE_{b,r}}{B-1} \tag{20}$$

The value of the average RMSE of the b-th modeling trajectory is able to evaluate the quality prediction performance in a directly way. Obviously, a larger average RMSE indicates the corresponding large prediction error. To reflect the results of cross validation, a weighting parameter is designed for each modeling result, which can be defined as

$$\eta_i = \frac{e^{-RMSE_i^2}}{\sum_{b=1}^{B} e^{-RMSE_b^2}} \tag{21}$$

By the use of the exponential function, a larger RMSE value will result in the smaller weight with the negative correlation. Finally, since the quality prediction result of the current query samples have been acquired as $\hat{Y}_{q,b} = \{\hat{y}_{q,b}, \hat{y}_{q+1,b}, \ldots, \hat{y}_{q+n-1,b}\}$, the ensemble result can be calculated based on the weighting parameters as

$$\hat{y}_q = \sum_{i=1}^{B} \eta_i \hat{y}_{q,i} \tag{22}$$

where $\hat{y}_q$ is the q-th ensemble prediction result and $\hat{Y}_q = \{\hat{y}_q, \hat{y}_{q+1}, \ldots, \hat{y}_{q+n-1}\}$ becomes the integrated prediction result of the query trajectory. It can be inferred that soft sensing models with larger weights will contribute more to the final prediction output, which is expected to improve the robustness of the quality prediction through ensemble learning strategy.

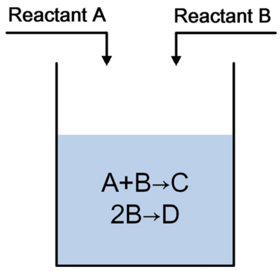


FIGURE 5. The fed-batch reactor.

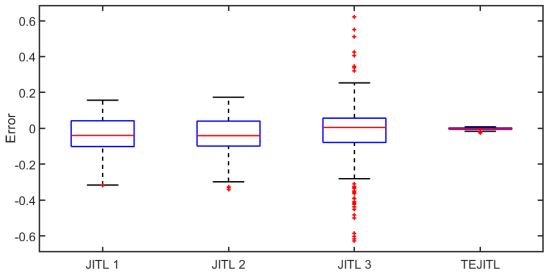


FIGURE 6. Error distributions of the batch reactor process.

In summary, the detailed procedure of the proposed methodology is demonstrated in Fig. 4 as

1) Collecting historical dataset X_H and implementing data normalization;
2) Collecting the real-time query samples as the batch trajectory $\{x_q, x_{q+1}, \ldots, x_{q+n-1}\}$;
3) Calculating different similarity measurements and extracting the modeling trajectories for each strategy;
4) Constructing online local soft sensing models $y = f_b(x)$ for each modeling trajectory;
5) Collecting the predicted results of different models $\{\hat{y}_{q,b}, \hat{y}_{q+1,b}, \ldots, \hat{y}_{q+n-1,b}\}$;
6) Determine the weighting parameters η_b of local models by cross validation strategy;
7) Obtain the ensembled prediction result $\{\hat{y}_q, \hat{y}_{q+1}, \ldots, \hat{y}_{q+n-1}\}$ by the weighted sum of local models;
8) Go to the next query trajectory.

B. DISCUSSIONS

There are two major concerns in this work. The first point is the development of the online local LSTM soft sensor, which is expected to make progress in modeling accuracy. For the traditional LSTM soft sensor for batch processes, the global modeling strategy is utilized and all historical samples are regarded as modeling data. However, the dynamic characteristic of batch data varies during different operating period. Although LSTM is able to extract the long-term time-varying feature, the network structure still retains the idea of the recurrent neural network. Hence, it is not proper to describe the entire batch process within a single LSTM model. This is the reason of using the online local LSTM soft sensor in this study.

Another innovation is the design of the multivariate trajectory based EJITL strategy. When using one type of similarity measurement as the conventional JITL strategy, the modeling effect may vary from one dataset to another due to batch-to-batch variations. By the use of EJITL, the performance of online local modeling can be significantly improved. Besides, different from the single-sample similarity measurements of conventional EJITL methods, batch trajectories are made full use of to create a compatible modeling dataset with the online local LSTM soft sensor.

By the combination of TEJITL strategy and online local LSTM soft sensing model, the within-batch dynamics and nonlinearity problem can be dealt with as well as the batch-to-batch variations. It is noted that the length of the batch trajectory for modeling should be determined firstly. Meanwhile, it is doubtful whether the ensemble quality prediction can provide a more effective result than any individual prediction mode. To verify these problems, the following two case studies are presented.

IV. CASE STUDIES

A. A FED-BATCH REACTOR PROCESS

In this subsection, a fed-batch reactor process is used to verify the proposed method as presented in Fig. 5 [31]. In this process, two chemical reactions occur as $A + B \rightarrow C$ and $2B \rightarrow D$(A and B are reactants, C is the final product, D is the byproduct). As a fed-batch process, reactant A is fed at the beginning of the batch, while reactant B is continuously fed during the batch operation. Thus, the only manipulated variable is the feed rate of reactant B. The detailed mathematical model of the process is presented in Appendix B [1].

TABLE 1. Process variables used for soft sensor in fed-batch reactor.

Number	Description
1	reactant A concentration (mol/L)
2	reactant B concentration (mol/L)
3	reactor holdup (L)
4	feed rate of reactant B (L/min)

In this case, the concentration of the final product C is regarded as the quality variable. The relevant process variables for quality prediction are listed in Table. 1.

There are totally 20 historical batches and each batch has 220 samples. Batch-to-batch variations are generated by setting diverse trajectories of the manipulated variable (feed rate of reactant B) and adding random noise. All the variables for the similarity measurement and soft sensor modeling are normalized to the range of -1 to 1. During the similarity measurement stage, the trade-off parameter λ between distance and angle of the second measurement is set as 0.5 and the step size s of the third measurement is set as 3. Given an online batch trajectory with 5 consecutive samples, 5 historical trajectories with the largest values of s_h for each

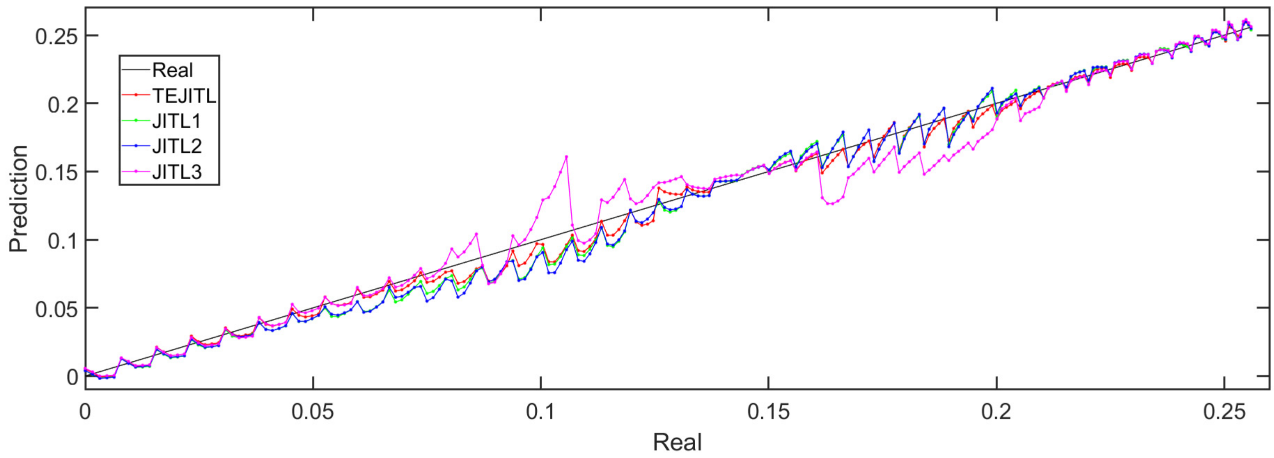


FIGURE 7. Comparisons of the predicted result with the reference of real value for the batch reactor process.

TABLE 2. Quality prediction results of fed-batch reactor.

Method	JITL1-LSTM	JITL2-LSTM	JITL3-LSTM	TEJITL-LSTM
RMSE	0.0085	0.0091	0.0137	**0.0067**
R^2	0.9873	0.9854	0.9666	**0.9921**

similarity measurement are selected as the modeling data. Hence, totally 15 historical datasets are collected for online local quality prediction.

For the LSTM modeling, the dimensions of the cell input and output are 4 and 1, respectively. The number of neurons in a hidden layer is set as 25. Thus, 15 individual LSTM models can be constructed and the corresponding predicted output can be calculated directly. After cross validation, the weights of different models can be determined and the final ensemble quality prediction result is revealed. The RMSE and the coefficient determination (R^2) of different JITL strategies and the proposed TEJITL framework are listed in Table. 2.

It can be inferred from the prediction results that the proposed TEJITL-LSTM soft sensor provides a smaller RMSE value and a larger R^2 value that any individual JITL-LSTM strategy, which means the quality prediction performance can be significantly improved by the use of the multivariate trajectory based ensemble strategy.

The boxplot with the error information of different strategies is listed in Fig. 6. As shown in Fig. 6, the boxes in blue indicate the ranges between the upper and lower quartiles, which include 50% results of the prediction error each. It can be observed clearly that the proposed TEJITL strategy offers the narrowest error range, which proves that it provides a better quality prediction result than other methods. Besides, the comparisons of the predicted output between the proposed method and other strategies are also presented in Fig. 7.

B. THE PENICILLIN FREMENTATION PROCESS

In this subsection, the effectiveness of the proposed method is further verified through the fed-batch penicillin fermentation process. The structure of the process is shown in Fig. 8.

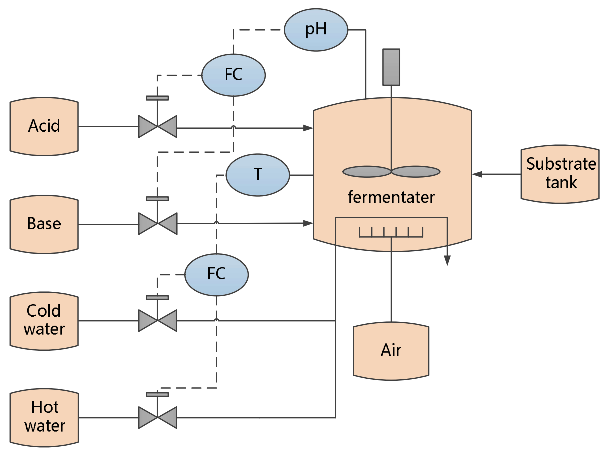


FIGURE 8. Flow chart of the fed-batch penicillin fermentation process.

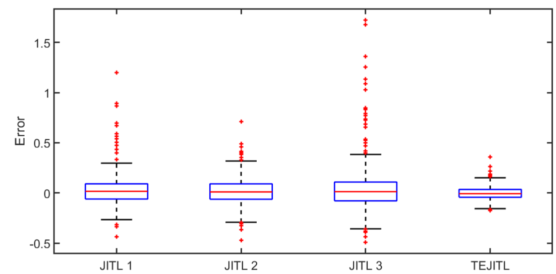


FIGURE 9. Error distributions of the penicillin process.

The process consists of two major phases. The first one is the pre-culture phase as the biomass growing and fed-batch stage. Next, the penicillin occurs at the second phase as the final product. In most literatures, the PenSim v2.0 benchmark software is used to generate process data [32], [33]. In this work, an improved penicillin simulation platform based on the PenSim model proposed by one of the authors in this paper is utilized for data generation [34]. Customized trajectories of manipulated variables are feasible in the framework, which is more flexible than the original PenSim benchmark.

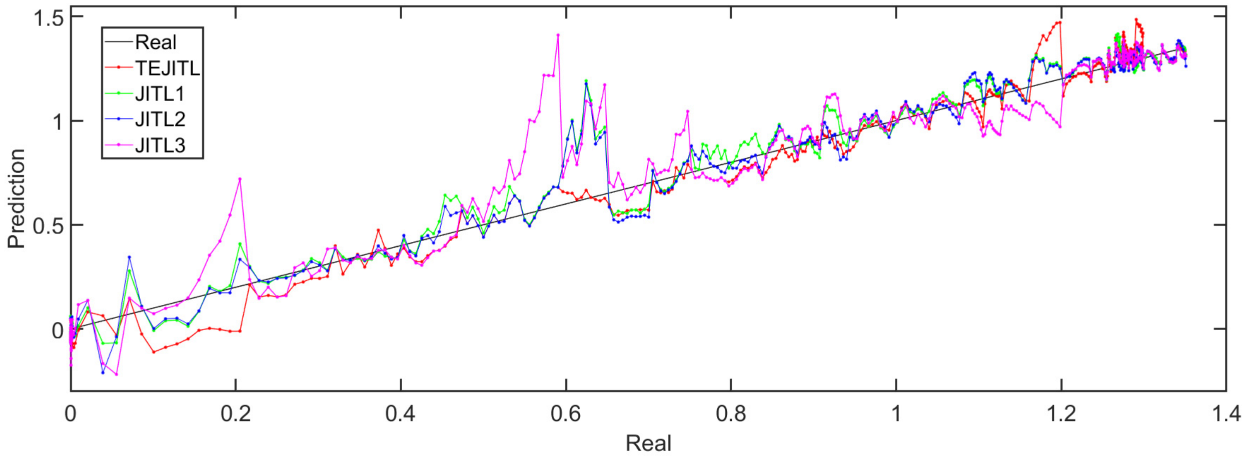


FIGURE 10. Comparisons of the predicted result with the reference of real value for the penicillin process.

TABLE 3. Process variables used for soft sensor in penicillin process.

Number	Description
1	aeration rate(L/h)
2	agitator power(W)
3	substrate feed rate(L/h)
4	substrate temperature(K)
5	substrate concentration(g/L)
6	dissolved oxygen concentration(g/L)
7	biomass concentration(g/L)
8	culture volume(L)
9	CO_2 concentration(g/L)
10	pH
11	generated heat(KCal)
12	cooling water flow rate(mL/h)

In this case, 10 historical batches are generated with 400 hours in each batch operation. The sampling interval is set as 1 hour. The penicillin concentration is regarded as the quality variable and other 12 process variables listed in Table. 3 are used as the input variables of LSTM soft sensor.

All the variables are normalized to the range between −1 and 1. The trade-off parameter λ of JITL 2 between distance and angle is set as 0.5, while the step size s of JITL 3 is set as 5. Consider a real-time batch trajectory with 10 consecutive samples, 5 historical trajectories with the largest values of similarity measurements are collected for modeling. Then, totally 15 historical datasets are extracted.

The dimensions of the LSTM cell input and output are 12 and 1, respectively. The number of neurons in a hidden layer is set as 70. Hence, 15 individual quality predicted results can be obtained according to the established LSTM models. Finally, the ensemble quality prediction result can be calculated based on the model weights. The RMSE and R^2 of each strategy are listed in Table. 4.

The result indicates that the prediction error of the proposed TEJITL based strategy is smaller than other JITL strategies. Meanwhile, the boxplot is shown in Fig. 9, where

TABLE 4. Quality prediction results of penicillin process.

Method	JITL1-LSTM	JITL2-LSTM	JITL3-LSTM	TEJITL-LSTM
RMSE	0.1787	0.1709	0.2804	**0.1549**
R^2	0.9643	0.9675	0.9121	**0.9732**

the proposed method performs better with a tight error range. The comparisons of the predicted result are also shown in Fig. 10. Although some outliers of TEJITL can be found with larger prediction errors, the boxplot in Fig. 9 indicates that the total outlier number of TEJITL is less than other methods. Hence, the overall performance of TEJITL is proved to be more effective. Compared with the reactor process, the penicillin process is more complicated with stronger process dynamics. As a result, a single JITL similarity measurement has a greater opportunity to match wrong historical trajectories even the searching range is limited to the same phase.

V. CONCLUSION

In this study, a novel TEJITL strategy is developed for quality prediction of batch processes combined with the online local LSTM soft sensor. Two major issues are concerned including the ensemble similarity measurements for batch trajectories and the online local LSTM soft sensor development. Therefore, the proposed framework can handle with both the problem of within-batch process nonlinearity, dynamics and run-to-run variations for batch processes. For simpler cases that both batch-to-batch variations and within-batch dynamics are not obvious, an individual global JITL-LSTM model may perform as well as the proposed method.

Meanwhile, it may consume heavier computation load for the proposed method due to the online modeling strategy.

The case studies prove that the proposed TEJITL-LSTM soft sensor is able to provide a reliable and robust quality prediction result. In future work, the adaptive parameter selection scheme will be focused on, aiming at making auto determination for model parameters such as the length of the modeling strategy and the neuron number of the LSTM hidden layer. Besides, the proposed method will be further testified on practical industrial processes with uneven batch length, which requires an improved trajectory based similarity measurements.

APPENDIXES

APPENDIX A
THE BPTT ALGORITHM

According to the network structure of LSTM, the back propagation through time (BPTT) algorithm is described as follows

$$L = \sum_{t=1}^{T} \left(\hat{y}_t - y_t\right)^2 \Big/ T \tag{23}$$

$$\delta h_t = \frac{\partial L}{\partial h_t} + \delta i_{t+1} W_{hi}^T + \delta i_{f+1} W_{hf}^T + \delta o_{t+1} W_{ho}^T + \delta \tilde{c}_{t+1} W_{h\tilde{c}}^T \tag{24}$$

$$\delta i_t = (\delta h_t)^T \odot o_t \odot \left(1 - \tanh^2(c_t)\right) \odot c_{t-1} \odot (i_t \cdot (1 - i_t)) \tag{25}$$

$$\delta f_t = (\delta h_t)^T \odot o_t \odot \left(1 - \tanh^2(c_t)\right) \odot \tilde{c}_t \odot (f_t \cdot (1 - f_t)) \tag{26}$$

$$\delta o_t = (\delta h_t)^T \odot \tanh(c_t) \odot (o_t \cdot (1 - o_t)) \tag{27}$$

$$\delta \tilde{c}_t = (\delta h_t)^T \odot o_t \odot \left(1 - \tanh^2(c_t)\right) \odot i_t \tilde{\odot} \left(1 - \tanh^2(\tilde{c}_t)\right) \tag{28}$$

$$\delta W_{x*} = \sum_{t=1}^{T} \langle \delta *_t, x_t \rangle \tag{29}$$

$$\delta W_{h*} = \sum_{t=1}^{T} \langle \delta *_t, h_{t-1} \rangle \tag{30}$$

$$\delta b* = \sum_{t=1}^{T} \delta *_t \tag{31}$$

where L represents the loss function; T is the total number of the data samples for model training; $\delta W_{h^*} = \partial E / \partial W_{h^*}$ are gradients of weighting parameters; $\delta b^* = \partial E / \partial b^*$ are gradients of biases; * represents different gates which can be $i, f, o, \tilde{c}$; α defines the step size of the gradient descending rate. Thus, the values of model parameters W_{h^*}, W_{x^*}, b^* can be estimated by gradient descent algorithms.

APPENDIX B
MATHEMATICAL MODEL OF THE BATCH REACTOR

The first principle model of the fed-batch reactor is based on the material balance law described as follows

$$\dot{c}_A = -k_1 c_A c_B - c_A - \frac{c_A u}{V}, c_A(0) = c_{A0} \tag{32}$$

$$\dot{c}_B = -k_1 c_A c_B - 2k_2 c_B^2 - \frac{\left(c_A - c_B^{in}\right) u}{V}, c_B(0) = c_{B0} \tag{33}$$

$$\dot{V} = u, V(0) = V_0 \tag{34}$$

$$\dot{c}_C = k_1 c_A c_B - \frac{c_C u}{V}, c_C(0) = c_{C0} \tag{35}$$

$$\dot{c}_D = k_2 c_B^2 - \frac{c_D u}{V}, c_D(0) = c_{D0} \tag{36}$$

where c_A, c_B, c_C, c_D are concentrations of reactants A, B, C, D, respectively; V represents the reactor holdup; k_1, k_2 are kinetic coefficients; c_B^{in} is the inlet concentration of reactant B; u is the manipulated variable.

ACKNOWLEDGMENT

The authors are thankful to the editors and reviewers for their kind assistance and useful comments, which is very helpful to improve this work.

REFERENCES

[1] L. Ye, H. Guan, X. Yuan, and X. Ma, "Run-to-run optimization of batch processes with self-optimizing control strategy," *Can. J. Chem. Eng.*, vol. 95, no. 4, pp. 724–736, Apr. 2017.

[2] M. Onel, C. A. Kieslich, Y. A. Guzman, C. A. Floudas, and E. N. Pistikopoulos, "Big data approach to batch process monitoring: Simultaneous fault detection and diagnosis using nonlinear support vector machine-based feature selection," *Comput. Chem. Eng.*, vol. 115, pp. 46–63, Jul. 2018.

[3] I. Vergili, U. Golebatmaz, Y. Kaya, Z. B. Gönder, H. Hasar, and G. Yilmaz, "Performance and microbial shift during acidification of a real pharmaceutical wastewater by using an anaerobic sequencing batch reactor (AnSBR)," *J. Environ. Manage.*, vol. 212, pp. 186–197, Apr. 2018.

[4] K. Peng, Q. Li, K. Zhang, and J. Dong, "Quality-related process monitoring for dynamic non-Gaussian batch process with multi-phase using a new data-driven method," *Neurocomputing*, vol. 214, pp. 317–328, Nov. 2016.

[5] H. Zhang, X. Tian, X. Deng, and Y. Cao, "Multiphase batch process with transitions monitoring based on global preserving statistics slow feature analysis," *Neurocomputing*, vol. 293, no. 7, pp. 64–86, Jun. 2018.

[6] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometric Intell. Lab. Syst.*, vol. 171, pp. 16–25, Dec. 2017.

[7] F. Shen, J. Zheng, L. Ye, and D. Gu, "Quality-relevant monitoring of batch processes based on stochastic programming with multiple output modes," *Processes*, vol. 8, no. 2, p. 164, Feb. 2020.

[8] L. Luo, "Trajectory-based phase partition and multiphase multilinear models for monitoring and quality prediction of multiphase batch processes," *J. Chemometrics*, vol. 32, no. 7, p. e3013, Jul. 2018.

[9] X. Yuan, Y. Wang, C. Yang, Z. Ge, Z. Song, and W. Gui, "Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1508–1517, Feb. 2018.

[10] X. Yuan, Z. Ge, and Z. Song, "Spatio-temporal adaptive soft sensor for nonlinear time-varying and variable drifting processes based on moving window LWPLS and time difference model," *Asia–Pacific J. Chem. Eng.*, vol. 11, no. 2, pp. 209–219, Mar. 2016.

[11] Z. Ge, B. Huang, and Z. Song, "Nonlinear semisupervised principal component regression for soft sensor modeling and its mixture form," *J. Chemometrics*, vol. 28, no. 11, pp. 793–804, Nov. 2014.

[12] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[13] Q. Jiang, X. Yan, and B. Huang, "Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes," *Ind. Eng. Chem. Res.*, vol. 58, no. 29, pp. 12899–12912, Jul. 2019.

[14] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.

[15] L. Luo, S. Bao, J. Mao, and D. Tang, "Quality prediction and quality-relevant monitoring with multilinear PLS for batch processes," *Chemometric Intell. Lab. Syst.*, vol. 150, pp. 9–22, Jan. 2016.

[16] V. Gopakumar, S. Tiwari, and I. Rahman, "A deep learning based data driven soft sensor for bioprocesses," *Biochem. Eng. J.*, vol. 136, pp. 28–39, Aug. 2018.

[17] X. Yuan, Z. Ge, and Z. Song, "Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression," *Chemometric Intell. Lab. Syst.*, vol. 138, pp. 97–109, Nov. 2014.

[18] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *J. Process Control*, vol. 24, no. 3, pp. 223–233, Mar. 2014.

[19] Y. Liu, C. Yang, Z. Gao, and Y. Yao, "Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes," *Chemometric Intell. Lab. Syst.*, vol. 174, pp. 15–21, Mar. 2018.

[20] Y. Wang, Q. Jiang, B. Li, and L. Cui, "Joint-individual monitoring of parallel-running batch processes based on MCCA," *IEEE Access*, vol. 6, pp. 13005–13014, 2018.

[21] D. B. Nickel, M. N. Cruz-Bournazou, T. Wilms, P. Neubauer, and A. Knepper, "Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale," *Eng. Life Sci.*, vol. 17, no. 11, pp. 1195–1201, Nov. 2017.

[22] Z. Lv, X. Yan, and Q. Jiang, "Batch process monitoring based on just-in-time learning and multiple-subspace principal component analysis," *Chemometric Intell. Lab. Syst.*, vol. 137, no. 20, pp. 128–139, Oct. 2014.

[23] F. Shen, Z. Ge, and Z. Song, "Multivariate trajectory-based local monitoring method for multiphase batch processes," *Ind. Eng. Chem. Res.*, vol. 54, no. 4, pp. 1313–1325, Feb. 2015.

[24] C. Cheng and M.-S. Chiu, "Nonlinear process monitoring using JITL-PCA," *Chemometric Intell. Lab. Syst.*, vol. 76, no. 1, pp. 1–13, Mar. 2005.

[25] 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 Trans. Ind. Informat.*, vol. 13, no. 2, pp. 532–541, Apr. 2017.

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

[27] H. Zhao, S. Sun, and B. Jin, "Sequential fault diagnosis based on LSTM neural network," *IEEE Access*, vol. 6, pp. 12929–12939, 2018.

[28] M. Han, W. Chen, and A. D. Moges, "Fast image captioning using LSTM," *Cluster Comput.*, vol. 22, no. S3, pp. 6143–6155, May 2019.

[29] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018.

[30] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3168–3176, May 2020.

[31] B. Chachuat, B. Srinivasan, and D. Bonvin, "Adaptation strategies for real-time optimization," *Comput. Chem. Eng.*, vol. 33, no. 10, pp. 1557–1567, Oct. 2009.

[32] W. Sun, Y. Meng, A. Palazoglu, J. Zhao, H. Zhang, and J. Zhang, "A method for multiphase batch process monitoring based on auto phase identification," *J. Process Control*, vol. 21, no. 4, pp. 627–638, Apr. 2011.

[33] G. Birol, C. Ündey, and A. Çinar, "A modular simulation package for fed-batch fermentation: Penicillin production," *Comput. Chem. Eng.*, vol. 26, no. 11, pp. 1553–1565, Nov. 2002.

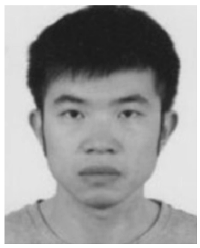
[34] L. Ye and J. Cheng, "Simulator of penicillin fermentation process in MATLAB/simulink environment," *J. Syst. Simul.*, vol. 27, no. 3, pp. 515–520, Mar. 2015.

FEIFAN SHEN received the B.Eng. degree in electronic information technology and instruments and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2011 and 2016, respectively.

Since 2017, he has been with the School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo, China, where he is currently a Lecturer. His research interests include data-driven modeling, batch process control, process monitoring, and soft sensor and their applications in process systems.

JIAQI ZHENG received the B.Eng. degree in automation from the Beijing University of Chemical Technology, Beijing, China, in 2015, and the M.S. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2018.

Since 2019, she has been with the School of Mechanical Engineering and automation, College of Science and Technology, Ningbo University, Ningbo, China, where she is currently an Assistant Researcher. Her research interests include statistical process control and dynamic process monitoring, and their applications in process systems.



LINGJIAN YE (Member, IEEE) received the B.Eng. degree in biochemical engineering and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2006 and 2011, respectively.

Since 2011, he has been with the School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo, China, where he is currently a Professor. His research interests include control structure design, plant-wide control, and real-time optimization and their applications in process systems.

XIUSHUI MA received the B.S. degree in industrial electrical automation from the Anhui College of Mechanical and Electrical Engineering, Wuhu, China, in 1985, the M.S. degree in automation engineering and the Ph.D. degree in precision machinery and instruments from the Hefei University of Technology, Hefei, China, in 1995 and 2002, respectively.

Since 2011, he has been with the School of Information Science and Engineering, Zhejiang University Ningbo Institute of Technology, Ningbo, China, where he is currently a Professor. His research interests include detection technology and automation equipment, and precision instruments and machinery.