Downloaded from orbit.dtu.dk on: Apr 26, 2024

DTU Library

A Supervised Bidirectional Long Short-Term Memory Network for Data-driven Dynamic Soft Sensor Modeling

Fai Lui, Chun ; Liu, Yiqi; Xie, Min

Published in:
IEEE Transactions on Instrumentation and Measurement

Link to article, DOI:
10.1109/TIM.2022.3152856

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Fai Lui, C., Liu, Y., & Xie, M. (2022). A Supervised Bidirectional Long Short-Term Memory Network for Data-driven Dynamic Soft Sensor Modeling. *IEEE Transactions on Instrumentation and Measurement*, *71*, Article 2504713. https://doi.org/10.1109/TIM.2022.3152856

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A Supervised Bidirectional Long Short-Term Memory Network for Data-Driven Dynamic Soft Sensor Modeling

Chun Fai Lui, Yiqi Liu, *Member, IEEE*, and Min Xie, *Fellow, IEEE*

***Abstract*— Data-driven soft sensors have been widely adopted in industrial processes to learn hidden knowledge automatically from process data, then to monitor difficult-to-measure quality variables. However, to extract and utilize useful dynamic latent features accurately for efficient quality estimations remains one of the most important research issues in soft sensor modeling. In this article, a supervised bidirectional long short-term memory (SBiLSTM) is proposed for data-driven dynamic soft sensor modeling. The SBiLSTM incorporates extended quality information with a moving window up to k time steps and enhances learning efficiency by bidirectional architecture. With this novel structure, the SBiLSTM can extract and utilize nonlinear dynamic latent information from both process variables and quality variables, then further improve the prediction performance significantly. The effectiveness of the proposed SBiLSTM network-based soft sensor model is demonstrated through two case studies on a debutanizer column process and an industrial wastewater treatment process. Results show that the SBiLSTM outperforms state-of-the-art and traditional deep learning-based soft sensor models.**

***Index Terms*— Data-driven, deep learning, long short-term memory (LSTM), quality prediction, soft sensor, supervised bidirectional long short-term memory (SBiLSTM).**

NOTATION

x_t	Input process variable vector at time t.
y_t	Quality output variable vector at time t.
y_{t-k}	Past quality output variable vector lagged by k time steps.
$\hat{y}_t$	Predicted quality output variable vector at time t.
i_t	State variable of input gate at time t.
f_t	State variable of forget gate at time t.
o_t	State variable of output gate at time t.
g_t	Cell update state variable vector at time t.
c_t	Cell state variable vector at time t.
h_t	Hidden state variable vector at time t.
$\tilde{h}_t$	Output variable vector of neuron unit at time t.

Manuscript received November 23, 2021; revised January 17, 2022; accepted February 10, 2022. Date of publication February 22, 2022; date of current version March 9, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 71971181 and Grant 72032005, in part by the Research Grant Council of Hong Kong under Grant 11203519 and Grant 11200621, in part by the Hong Kong Innovation and Technology Commission (InnoHK Project Centre for Intelligent Multidimensional Data Analysis (CIMDA)), and in part by the Hong Kong Institute of Data Science under Project 9360163. The work of Yiqi Liu was supported by the EU Horizon 2020 Research and Innovation Program through the Marie Skłodowska-Curie Grant under Agreement 891627. The Associate Editor coordinating the review process was Dr. Min Xia. *(Corresponding author: Chun Fai Lui.)*

Chun Fai Lui is with the Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong (e-mail: chunflui5-c@my.cityu.edu.hk).

Yiqi Liu is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China, and also with the Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Lyngby, Denmark (e-mail: aulyq@scut.edu.cn).

Min Xie is with the Department of Advanced Design and Systems Engineering and the School of Data Science, City University of Hong Kong, Hong Kong (e-mail: minxie@cityu.edu.hk).

Digital Object Identifier 10.1109/TIM.2022.3152856

I. INTRODUCTION

IN MANY industrial processes, it is essential to monitor quality variables for real-time process monitoring, effective operation control, and optimal management. However, using physical sensors to measure quality variables can be challenging in some industrial processes due to complex process environment, economic costs, measurement delay, and requirement of reliable measuring instruments [1]–[3]. Soft sensor, a type of virtual model that inputs measurable variables, learns latent variable information, and outputs quality characteristics prediction, has therefore gained more and more attention in process monitoring, especially in quality prediction as well as fault detection and diagnosis [4], [5].

In general, there are two categories of soft sensors: model-based soft sensors and data-driven soft sensors [6]. Model-based soft sensors are based on physical background of processes, therefore it requires extensive system knowledge about industrial process mechanism. In most industrial processes, however, accurate and complete understanding of process mechanism are time-consuming, costly, and often impractical due to multistage, highly complex, and dynamic process environment [7]. In contrast to model-based soft sensors, data-driven soft sensors use data-driven methods for modeling, analysis, and estimations without requiring much system knowledge [8], [9]. Because of this advantage, many data-driven soft sensors have been developed and successfully applied in various advanced industrial processes in the past two decades. Some typical data-driven methods applied to soft sensors include principal component analysis (PCA) [10], partial least square (PLS) [11], canonical correlation analysis (CCA) [12], relevant vector machine (RVM) [13], and artificial neural network (ANN) [14].

Traditionally, quality predictions are made by directly applying linear regression models on the relationship between measurement data and quality output. However, industrial process data are often high-dimensional and dynamic. To extract and utilize useful hidden variable representation

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

accurately and automatically from process data for effective quality estimations remains one of the most important research issues in soft sensor modeling. In recent years, deep learning has been under the spotlight of data-driven soft sensor modeling due to its ability to learn abstract latent features from data and applicability in soft sensors. For instance, Yuan *et al.* [15], [16] proposed a hybrid variable-wise weighted stacked autoencoder (HVW-SAE) and stacked quality-driven autoencoder (SQAE) successively to perform feature extraction for soft sensing modeling. Although the proposed model outperformed SAE by incorporating linear and nonlinear correlations, difficulty in extracting dynamic process information renders it inadequate for widely use. To construct soft sensing models that are able to capture dynamic information from process variables, various machine learning methods have been studied. Ge and Chen [17] developed a supervised linear dynamic system (LDS) model to detect process faults, and Shen and Ge [6] extended the supervised LDS model to a weighted nonlinear dynamic system (WNDS) model using variational autoencoder (VAE) whose weights were determined by intercorrelations. Also, Wang *et al.* [18] suggested an extension of deep belief networks (DBN) to construct dynamic extended DBN for feature extraction and fault classification. However, these deep learning-based soft sensor models focused on extracting dynamics of process variables but took existing dynamic information of past quality variables being solved for granted. To utilize the dynamic information from quality variables, Yuan *et al.* [19] proposed a supervised long short-term memory (SLSTM) network to learn quality-related latent dynamics as a nonlinear dynamic soft sensor model for accurate quality prediction. Despite the fact that some deep learning methods have successfully exploited quality information in soft sensor modeling, only the most recent past quality information has been considered when constructing the soft sensor model, thus neglecting extensive dynamic information latent in the dynamic quality variables. Moreover, some deep-leaning-based soft sensor models can be unstable and inefficient since the extracted dynamic information is stored using a large number of hidden neurons. In addition, many deep learning-based soft sensors are limited to offline analysis instead of online process monitoring. These limitations have hindered the development of deep learning-based soft sensor methods.

In this article, we propose a supervised bidirectional long short-term memory (SBiLSTM) network model for dynamic soft sensor modeling. The SBiLSTM network comprises a novel dynamic quality-guided supervised LSTM with a bidirectional structure, such that nonlinear dynamic latent information can be captured effectively. In the proposed algorithm, we incorporate an extended utilization of the dynamic quality information for a period of moving dynamic quality information window up to k time steps. Furthermore, dropout optimization is adopted to eliminate elementwise sum of bidirectional hidden variables with a determined probability, by which a relatively small number of neurons can then be achieved to assure model stability and efficiency. Results show that the proposed SBiLSTM network demonstrates superiority over other deep learning-based soft sensor modeling methods with higher prediction accuracy and very short computation time. The effectiveness of the proposed method is validated on a debutanizer column process and an industrial wastewater treatment process.

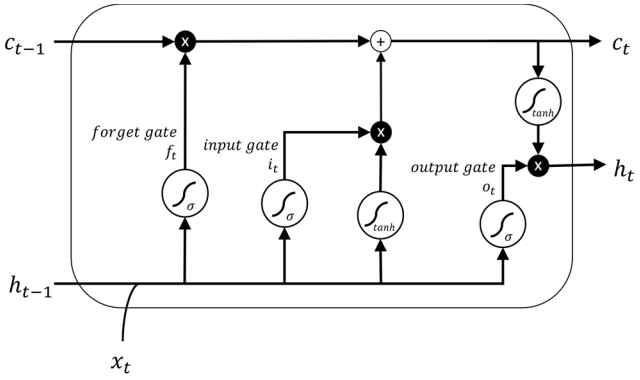

Fig. 1. Structure of LSTM unit.

The remaining sections of this article are organized as follows. Section II reviews the related works of deep learning-based soft sensors. In Section III, the proposed SBiLSTM and its soft sensor application are introduced. In Section IV, we demonstrate the effectiveness of the SBiLSTM network by two case studies on a debutanizer column process and an industrial wastewater treatment process. Finally, Section V concludes this article.

II. Related Works

Deep learning has gained tremendous attention in dynamic soft sensor modeling in recent years. Apart from dynamic soft sensors based on autoencoders, LDS, and DBN, a more commonly recognized deep learning model to exploit dynamic process information is the recurrent neural network (RNN), which has been widely adopted in industrial processes for dynamic soft sensor modeling because of its ability to model and process sequential data with memory. In particular, one popular, efficacious variant of RNN is the long short-term memory (LSTM), which can effectively mitigate the gradient vanishing or explosion problem when the model learns iteratively and propagates over time [20]. Compared to basic RNN, the LSTM added three nonlinear gates to collect nonlinear activations from model input, namely input gate i_t, forget gate f_t and output gate o_t. In addition, the cell state c_t in LSTM unit enables long-term memory of dynamic information. The structure of an LSTM unit is shown in Fig. 1.

Some examples of soft sensor application of these deep learning methods include the RNN-based soft sensor proposed by [21], the LSTM-based soft sensor proposed by [22] and [23], and soft sensors based on other variants of RNN [24]. A detailed review of RNN, LSTM, and other deep learning methods for data-driven soft sensor modeling can be found in [25].

With nonlinear gates and activation functions, LSTM-based soft sensors have shown excellent performance for quality prediction. However, the LSTM can only predict accurately

in limited applications where the industrial processes are less complex and the dynamic relations are explicit. Because of this, LSTM-based soft sensors can fail when handling process data that contain highly dynamic, nonlinear, and latent process information. To address this challenge, Yuan *et al.* [19] developed a SLSTM network to utilize quality information in the construction of LSTM-based soft sensor model. Also, Zhou *et al.* [23] proposed a difference LSTM (DLSTM)-based soft sensor model to capture the impact of sequential differences in the process and quality variables. The SLSTM and DLSTM provided effective, simple solutions to exploit quality information in LSTM for dynamic nonlinear soft sensor modeling. However, they only consider the most recent past quality information but not a moving period of historical quality information for every process sample, which essentially neglects the higher-order dynamic information in the historical quality variables that is readily available. Moreover, the SLSTM and DLSTM require storage of extracted nonlinear information with considerable number of hidden neurons, which can induce potential model instability and inefficiency. In addition, the SLSTM uses output variables y_t for training and replaces y_t with y_{t-1} for testing, which not only brings about potential model non-homogeneity that adds difficulty to model interpretation but also practical limitations on implementation.

In many complex industrial processes, dynamic correlations exist not only between quality and process variables but also within quality variables. When higher-order dynamic information within quality variables is neglected in the deep learning model, the deep learning-based soft sensor can deteriorate and fail to predict quality output accurately in large-scale industrial processes where quality variables are highly inner correlated with high-order dynamics and nonlinearity. Besides, it is also important to note that random selection of the number of hidden neurons can cause overfitting or underfitting issues [26], and the deep learning model can be unstable and inefficient if the number of hidden neurons is too large than required [27]. For industrial application of soft sensors, it is vital to ensure adequate efficiency and stability of the model such that the algorithm is not behind the incoming data stream and the prediction result is sufficiently reliable. However, even though some soft sensor models have shown excellent performance with substantial hidden neurons, the stability and efficiency of the model are often not mentioned. Apart from model effectiveness and efficiency, an online soft sensor that can predict quality output continually and automatically during the industrial process is always desirable than an offline one such that predictive and preventive decisions can be made with prognostics instead of diagnostics. Assuming no time lag in the measurement of process and quality variables is often impractical, which essentially limits the deep learning-based soft sensor to offline analysis instead of online process monitoring.

Consequently, it is vital to consider the dynamic quality information, model stability, and efficiency, as well as information usage in supervision of deep learning-based soft sensor model.

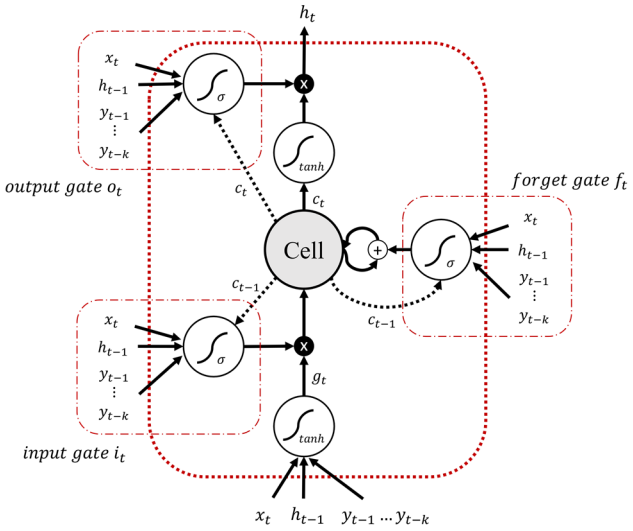

Fig. 2. Structure of SBiLSTM unit.

III. SBiLSTM

To address the above issues and to further enhance the utilization of quality information in soft sensors, this article proposes a SBiLSTM network for data-driven dynamic soft sensor modeling. The SBiLSTM network offsets the limitations and improves the SLSTM network by: 1) extending quality information utilization from the most recent past quality information to a period of quality information window up to k time steps; 2) employing dynamic quality-guided supervision in bidirectional LSTM to learn hidden variable information in both forward and backward directions in order to capture extra latent information; 3) reducing model instability and inefficiency with a reduced number of hidden neurons enabled by the bidirectional architecture of the SBiLSTM; and 4) avoiding model overfitting and improve model stability using dropout optimization.

A. SBiLSTM Unit

The structure of an SBiLSTM unit is shown in Fig. 2. Let P be the dimension of input vector x_t and Q be the dimension of output vector y_t. Further assume that there are n units of SBiLSTM neurons in an SBiLSTM layer, the weights of each node in an SBiLSTM unit can be denoted as

1) Input weights

$$W_{xi}, W_{xf}, W_{xc}, W_{xo} \in \mathbb{R}^{n \times P}.$$

2) Recurrent weights

$$W_{hi}, W_{hf}, W_{hc}, W_{ho} \in \mathbb{R}^{n \times n}.$$

3) Output weights

$$\left.\begin{matrix} W_{y_{t-1}i}, W_{y_{t-1}f}, W_{y_{t-1}c}, W_{y_{t-1}o} \\ \vdots \\ W_{y_{t-k}i}, W_{y_{t-k}f}, W_{y_{t-k}c}, W_{y_{t-k}o} \end{matrix}\right\} \in \mathbb{R}^{n \times Q}.$$

4) Biases

$$b_i, b_o, b_c, b_f \in \mathbb{R}^n.$$

Then, the input gate i_t, forget gate f_t, output gate o_t, cell update state g_t, cell state c_t and the hidden state h_t for an SBiLSTM unit can be expressed as

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{y_{t-1}i}y_{t-1} + \cdots + W_{y_{t-k}i}y_{t-k} + b_i\right) \tag{1}$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{y_{t-1}f}y_{t-1} + \cdots + W_{y_{t-k}f}y_{t-k} + b_f\right) \tag{2}$$

$$g_t = \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + W_{y_{t-1}c}y_{t-1} + \cdots + W_{y_{t-k}c}y_{t-k} + b_g\right) \tag{3}$$

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

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{y_{t-1}o}y_{t-1} + \cdots + W_{y_{t-k}o}y_{t-k} + b_o\right) \tag{5}$$

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

where $\odot$ denotes the pointwise multiplication, $\sigma(\cdot)$ represents the sigmoid nonlinear activation function and $\tanh(\cdot)$ represents the hyperbolic tangent function as a nonlinear activation function.

And the output of the SBiLSTM unit at time t can be written as

$$\tilde{h}_t = \left[\vec{h}_t \oplus \overleftarrow{h}_t\right]. \tag{7}$$

Here, $\oplus$ denotes the elementwise summation, $\vec{h}_t$ represents the hidden state of the forward pass and $\overleftarrow{h}_t$ represents the hidden state of the backward pass of the SBiLSTM unit at time t. The forward pass and the backward pass are combined by the elementwise summation to generate the final output $\tilde{h}_t$ of the SBiLSTM unit.

In an SBiLSTM network that consists of n SBiLSTM units, the input vector x_t and the delayed output vector $\{y_{t-k}, \ldots, y_{t-1}\}$ of window size k are continually fed into the input layer as a combined matrix. After entering the SBiLSTM network, the matrix is consequently transformed into nonlinear features by the nonlinear activation functions $\sigma(\cdot)$ and $\tanh(\cdot)$. With the gate design shown in Fig. 2 and (1)–(6), the sequential dynamic information of the input matrices is iteratively extracted and learned by the SBiLSTM unit. The network of n SBiLSTM units finally generate the cell state c_t and hidden state h_t for each SBiLSTM unit at each sampling time point t. It is particularly important to note that the cell state c_t, which is continually dismissed by the forget gate f_t and iteratively updated by the cell update state g_t, is responsible for storing long-term memory, while the hidden state h_t is responsible for storing short-term memory.

One significant contribution of the SBiLSTM model is the extension of the quality variable utilization with a moving window up to k time steps. Dissimilar to simple BiLSTM method, it can be seen in Figs. 2 and 3 that the proposed SBiLSTM employs a window of past quality variables $\{y_{t-k}, \ldots, y_{t-1}\}$ as part of the activation input in input gate i_t, forget gate f_t, output gate o_t and cell update state g_t. Through incorporating the moving vector $\{y_{t-k}, \ldots, y_{t-1}\}$ of window size k, the SBiLSTM considers the autoregression within quality information to be transformed by activation functions, hence enables automatic learning of nonlinear dynamic quality information by the SBiLSTM cell. By introducing the historical quality variables for an extended period of time, the SBiLSTM model exploits and learns extended quality information as further guidance to the learning of latent dynamics and sequential correlations in hidden state and cell states. Besides capturing dynamic features within the constructed moving window, such inclusion also permits the SBiLSTM network to investigate temporal relationship among process variables and quality variables. After iterative training procedures of deep learning, the determined weights and biases of the SBiLSTM network ultimately provides a complex function of gates and nonlinear functions to model the relationship between process input and quality output. On this basis, when extended dynamic quality variables are introduced to the process input for the model training, we enable extended dynamic feedback of historical quality information on the process variables for process monitoring. As quality data are usually constantly updated and past quality variables are readily available in industrial processes, such extension furthers the mining and utilization of quality information for more accurate quality prediction.

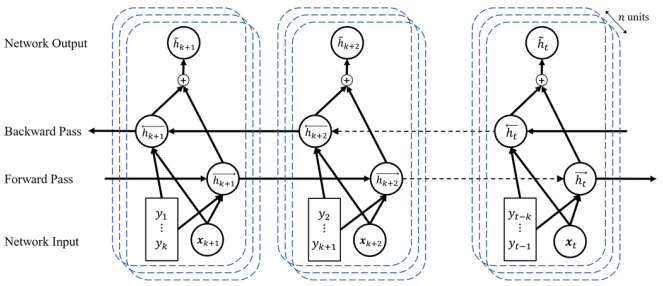

Fig. 3. Bidirectional structure of SBiLSTM network.

Fig. 3 shows the bidirectional structure of the SBiLSTM network at $(k+1, k+2, \ldots, t)$ time points. With both previous hidden state variable h_{t-1} and next hidden state variable h_{t+1} passed by the SBiLSTM network in their corresponding propagation direction, the hidden states are learned in bi-directions instead of uni-direction. The main difference between unidirectional feedforward and bidirectional networks is that a bidirectional architecture enables simultaneous backward and forward information flow for efficient modeling. With the nonlinear gates inherent in LSTM, the bidirectional structure of the SBiLSTM network assembles and collects further hidden features from the input data. Especially when the input data involve highly latent nonlinear dynamics, the bidirectional structure allows the model to take extra consideration and extraction of dynamic variable information iteratively without requiring extra training time. On the other hand, the bidirectional hidden states, $\vec{h}_t$ and $\overleftarrow{h}_t$, can be linked and connected with customized function to learn more complex bidirectional dynamic patterns that are latent in process data. In this article, we consider the most common way to connect the bidirectional hidden states for simplicity. Experiments on the improvements of (7) could serve another interesting future research work.

B. SBiLSTM-Based Soft Sensor

As the SBiLSTM network is capable of learning extended quality-guided latent dynamic nonlinear features, it is very applicable for soft sensor modeling. Fig. 4 provides a

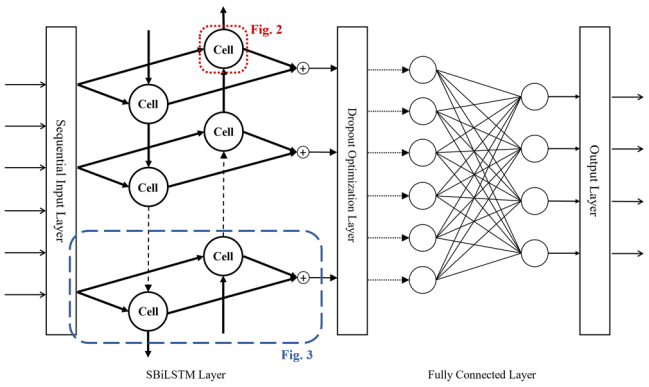

Fig. 4. Schematic of SBiLSTM network-based soft sensor.

schematic illustrating the SBiLSTM network-based soft sensor modeling framework, which includes five main layers: sequential input layer, SBiLSTM layer, dropout optimization layer, fully connected layer, and output layer. Given a set of time series data, the sequential input layer feeds the data into the neural network in sequential order. The sequential input data are then passed to the SBiLSTM layer comprises a network of SBiLSTM units. In each SBiLSTM unit, the sequential input data are transformed into nonlinear dynamic features by a cell update state and three nonlinear gates. Through learning the nonlinear dynamic features, the memory cell c_t in the SBiLSTM unit stores nonlinear dynamic information as long-term memory and the hidden state variable h_t represents short-term memory for each sampling time t. The resulted SBiLSTM network output $\tilde{h}_t$ is then passed to the dropout optimization layer that randomly removes output of each SBiLSTM unit with a determined probability to prevent model overfitting. Finally, the fully connected layer connects the features to predicted quality output $\hat{y}_t$ by weights and biases.

Considering a SBiLSTM network without dropout optimization, the hidden variables are fully connected into quality outputs by

$$\tilde{y}_j = \sum_i W_{ij}\tilde{h}_i + b_j. \tag{8}$$

The objective of the network training is to minimize the loss function, which is given by

$$\tilde{L} = \text{argmin} \sum_j \left(y_j - \tilde{y}_j\right)^2. \tag{9}$$

When dropout optimization is employed, the hidden nodes are randomly eliminated with a predetermined probability p. Subsequently, (8) can be rewritten as

$$\hat{y}_j = \sum_i W_{ij} d_i \tilde{h}_i + b_j \tag{10}$$

where $d_i \sim \text{Bernoulli}(p)$.

Then, the loss function can be expressed as

$$\begin{aligned} L &= \text{argmin} \sum_j \left(y_j - \hat{y}_j\right)^2 \\ &= \text{argmin} \sum_j \left[y_j - d\left(\tilde{y}_j - b_j\right) + b_j\right]^2. \end{aligned} \tag{11}$$

Notice that the loss function is regularized by the dropout optimization with a dropout probability of p during every epoch of the training process. By eliminating the elementwise sum of bidirectional hidden variables, the dropout optimization acts as a regularizer to avoid the co-adaptation of neurons during the training process of the SBiLSTM network. Detailed discussion on the usage of the dropout optimization in preventing overfitting, improving generalization capability, and allowing more gradient information flow can be found in [28].

For effective modeling of SBiLSTM-based soft sensor, the determination of k is exceptionally crucial. In general, the value of k is associated with the dynamic order of the industrial process. When we consider a dynamic order $k = 0$, we assume a simple process that involve trivial dynamics between quality variables. In this sense, a complex industrial process involving sophisticated dynamics distributed throughout the system should very likely associate with a higher value of k. Although there is no universal method to systematically calculate an optimal value of k as industrial processes differ from one to another, the value of k can be determined by a trial-and-error of increasing k values. From a theoretical perspective, if there exist sufficiently significant dynamics among quality variables of a process, a dynamic order $k = 1$ should yield a better-performed SBiLSTM-based soft sensor than one with $k = 0$. Therefore, the SBiLSTM-based soft sensor will eventually stops improving at a certain value of k. Except special cases, the value of k normally ranges from 1 to 5.

To construct an efficient soft sensor, mini-batch gradient descent is adopted to achieve stable and faster convergence in the training process. For mini-batch training, the training set is divided into subsets called mini-batches which help to update weights quicker and evaluate the gradient of the loss function with metrics. To evaluate and select the best SBiLSTM network over iterations in the training stage, mini-batch root-mean-squared-error (RMSE) is employed to evaluate model performance over iterations of mini-batch training. Compared to other error metrics, such as mean absolute error (MAE) and mean absolute percentage error (MAPE), RMSE has excellent sensitivity to modeling errors and presents the errors in a relatively comparable scale. In the training process, the model with the least mini-batch RMSE is selected as it indicates a model with the most accurate quality prediction in the training stage. Likewise, RMSE is used to compare and evaluate the overall model effectiveness of the SBiLSTM network model in both training and test stages. Let M be the mini-batch size, T be the number of samples in the training set, and N be the number of samples in the test set, the RMSE for mini-batch, training stage and test stage can be written as

$$\text{RMSE}_{\text{mb}} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(y_m - \hat{y}_m)^2} \tag{12}$$

$$\text{RMSE}_{\text{train}} = \sqrt{\frac{1}{T-k}\sum_{t=k+1}^{T}(y_t - \hat{y}_t)^2} \tag{13}$$

$$\text{RMSE}_{\text{test}} = \sqrt{\frac{1}{N}\sum_{t=T+1}^{T+N}(y_t - \hat{y}_t)^2}. \tag{14}$$

TABLE I
PROCEDURE OF SOFT SENSOR MODELING BASED ON SBiLSTM

Training stage:	
1:	Collect measurable process variables $\{x_1, \ldots, x_T\}$ and quality variables $\{y_1, \ldots, y_{T-1}\}$ from the industrial process. Clean the data by dropping null samples in the data.
2:	Split the data into training set and test set by a predetermined ratio.
3:	Determine the dynamic order k of the quality variables.
4:	Determine the hyperparameters in the model, including mini-batch size m, maximum number of epochs, and initial learning rate η and number of hidden neurons.
5:	Initialize the input weights $W_{xi}, W_{xf}, W_{xc}, W_{xo}$ with Glorot initializer and the recurrent weights $W_{hi}, W_{hf}, W_{hc}, W_{ho}$ with orthogonal matrix from QR decomposition.
6:	Train the SBiLSTM network with x_t and dynamic quality variable window $\{y_{t-k}, \ldots, y_{t-1}\}$ for all integer $t \in [k+1, T]$ by mini-batch training. Select the model with the lowest mini-batch RMSE.
Test stage:	
7:	**for** $i \geq 1$,
8:	Combine new input variable vector x_{t+i} and the historical quality variable window $\{y_{t+i-k}, \ldots, y_{t+i-1}\}$ for sequential input of the SBiLSTM network.
9:	Substitute the sequential input developed in Step 8 into the SBiLSTM network trained in Step 6 to compute the bidirectional hidden states $\overrightarrow{h}_{t+i}$, $\overleftarrow{h}_{t+i}$, the cell state c_{t+i}, and the network output $\tilde{h}_{t+i}$ at time $t+i$.
10:	Predict the quality output $\hat{y}_{t+i}$ by passing $\tilde{h}_{t+i}$ to dropout optimization layer, fully connected layer and output layer.
11:	**end for**

Furthermore, R-squared values are also computed to provide a metric for more intuitive comparisons on training and test performance of the model. Different from RMSE, the R-squared is a scale-independent error metric that ranges from 0% to 100%. Such properties allow intuitive and direct comparison of prediction ability of different soft sensor modeling methods. The R-squared is defined as

$$R^2 = 1 - \frac{\sum_{t=k}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=k}^{T}(y_t - \bar{y})^2}. \tag{15}$$

Note that only available information is adopted at each sampling time point t as input variables throughout the training and test stages such that an online soft sensor is enabled for continual process monitoring. The training and test procedure for soft sensor modeling based on SBiLSTM network are shown in Table I.

IV. CASE STUDY

A. Debutanizer Column Process

The debutanizer column process is a significant part of the petrochemical refinery process designed for desulfurization and naphtha split. The main goal of the debutanizer column process is to remove butane from the continuous stream of hydrocarbon mixture, therefore the quality variable of the debutanizer column process is defined as the butane concentration. Fig. 5 shows the flowchart of the debutanizer column process. It can be seen from the flowchart the locations

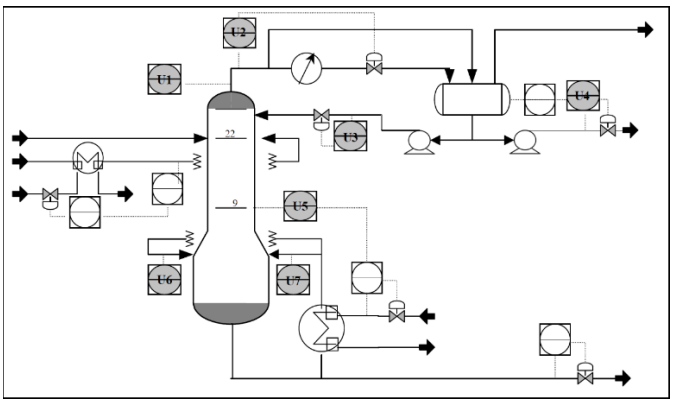

Fig. 5. Flowchart of the debutanizer column process [28].

TABLE II
DESCRIPTION OF VARIABLES OF THE DEBUTANIZER COLUMN PROCESS

Variables	Description
u_1	Top temperature
u_2	Top pressure
u_3	Reflux flow
u_4	Flow to next process
u_5	6th tray temperature
u_6	Bottom temperature
u_7	Bottom temperature
Quality variable y	Butane (C4) content

of seven sensors installed on the petrochemical plant for product quality monitoring, each collects separate measured process variables during the debutanizer column process. The detailed description of the seven process variables is given in Table II. It is important to note that the quality variable y, i.e., the butane content, is not indicated in Fig. 5 because it is usually measured on the overheads of the deisopentanizer column which is another fractional distillation column after the debutanizer column. The reason being that it is difficult to directly measure butane content at the bottom flow of the debutanizer column due to equipment and operational environment limitations. Because of this, accurate quality prediction by soft sensors is critical in the debutanizer column process for effective process control and monitoring.

In this case study, we use the debutanizer column dataset shared by Fortuna *et al.* [29], which is a popular benchmark for data-driven soft sensor modeling. The data were measured by sensors installed in a debutanizer column and a measuring device on the overhead of a deisopentanizer column with a measuring cycle of 15 min [29]. There are a total number of 2394 valid samples, seven process variables, and one process output in the dataset. We use the first 65% percent of the data (1556 samples) for model training and the remaining 35% percent (838 samples) for testing. The software and hardware configurations used for this case study are as follows: OS: Windows 10 Home (64-bit), CPU: Intel(R) Core(TM) i7-8565U (1.80GHz), GPU: NVIDIA GeForce MX150, RAM: 16GB.

Before training the SBiLSTM network, it is important to determine the hyperparameters and the dynamic order k of the quality variable. To select the best hyperparameters and dynamic order, the hyperparameters are tuned based on grid

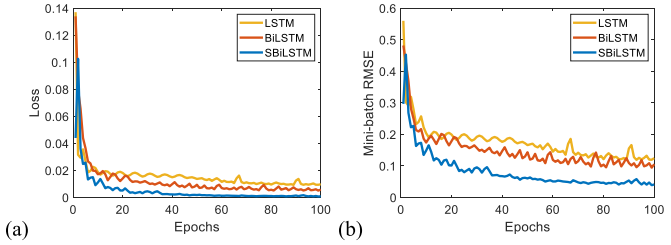

Fig. 6. (a) Training loss of LSTM, BiLSTM and SBiLSTM network over epoch. (b) Mini batch RMSE of LSTM, BiLSTM and SBiLSTM network over epoch on the debutanizer column process.

search and the dynamic order is determined by trial-and-error. For the grid search, we search the number of neurons from 1 to 100 with an interval of 5, the mini-batch size from 10 to 100 with an interval of 10, the maximum number of epochs from 50 to 200 with an interval of 10, and the initial learning rate from 10^{-4} to 10^{-1} with a geometric interval of 10. In addition, the dynamic order k is determined after trials on $k = [1, 2, 3, 4, 5]$. It was observed that the performance of the soft sensor model stops improving at $k = 3$. The SBiLSTM network is finally set with number of hidden neurons $n = 15$, mini-batch size $m = 20$, maximum number of epochs $= 100$, initial learning rate $\eta = 0.01$, and dynamic order $k = 3$. Then, we follow the procedures in Section III-B and the SBiLSTM network is trained using the adaptive moment estimation (Adam) optimizer.

The training loss and the mini-batch RMSE of the LSTM, BiLSTM, and SBiLSTM network during the training stage are given in Fig. 6. Comparing to the training loss and mini-batch RMSE of LSTM and BiLSTM, it is obvious that the SBiLSTM network converges faster and stabler with less fluctuation and smoother convergence. For SBiLSTM, the loss function and the RMSE value both drop exponentially over training epoch until convergence to nearly 0. This shows the excellent efficiency and performance of the SBiLSTM network on the debutanizer column data during training phase, and that the hyperparameters are well determined for the model. In addition, it is also important to note that the training process of the SBiLSTM network takes only 16 s using the very affordable setup. As previously mentioned, it is vital to ensure adequate efficiency of model such that the algorithm is not behind the incoming data stream. Since the proposed SBiLSTM network requires very short computation time, it refrains from practical applicability issue of soft sensors and allows efficacious applications in advanced industrial processes.

Fig. 7 compares the quality prediction to the true quality output of the debutanizer column process and shows the prediction errors of the SBiLSTM network model for both training and test phases. From the figure, it can be observed that the predicted quality fits almost perfectly to the true quality output in both the training and test phase. This shows that the capability of the SBiLSTM network model to exploit and learn nonlinear dynamic latent features between process variables and within quality variable. In Fig. 7(a), slight prediction deviation appears near the end of the test phase. Since lower butane content indicates a better quality in the

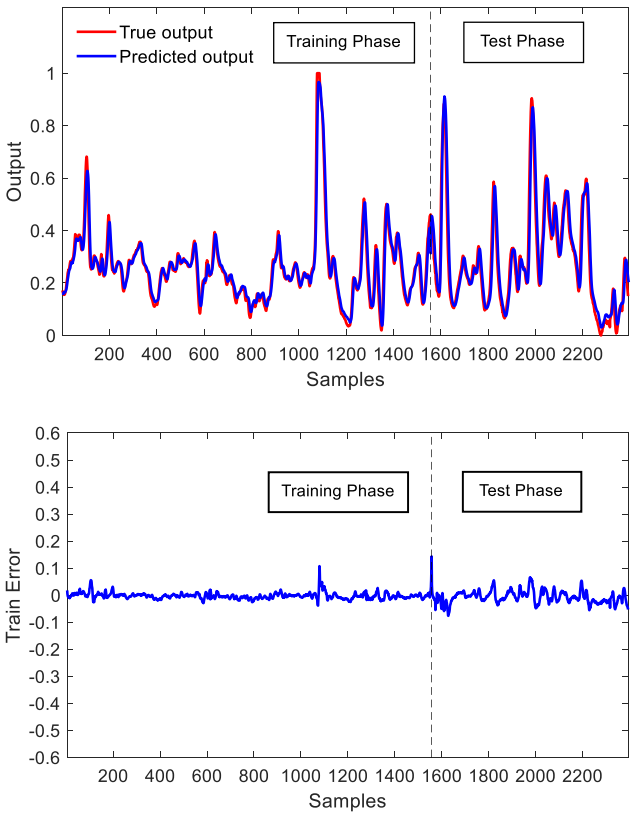

Fig. 7. (a) Comparison of true quality output and predicted quality output of the debutanizer column process by SBiLSTM network model during training phase and test phase. (b) Model prediction errors of SBiLSTM network model on the quality prediction of the debutanizer column process during training phase and test phase.

debutanizer column process, such slight deviation can be due to unexpected process improvement due to manual control or adjustment in which the extracted nonlinear dynamic features fail to capture and explain. Nonetheless, the extremely clear overlapping between the predicted quality and the true quality output illustrates the effectiveness of the proposed SBiLSTM network for soft sensor modeling. Additionally, the prediction errors centers about zero with trivial fluctuations. This further validates the stability and accuracy of the SBiLSTM network model.

The main reason why the SBiLSTM network-based soft sensor performs well on the debutanizer column is that the SBiLSTM network is able to extract extra dynamic latent quality information from historical quality variable and learn the dynamic latent information in a bidirectional manner. Unlike unidirectional RNN models, the SBiLSTM network model captures and utilizes bidirectional information to predict quality variable with additional robustness owing to the inclusion of backward directional information for information compensation. Moreover, the mining and exploitation of dynamic quality information from a moving window of historical quality variable extended the utility of past quality information from data. The obtained result demonstrated the effectiveness of the quality-supervised bidirectional structure in industrial soft sensor modeling.

To compare the performance of the SBiLSTM network with traditional dynamic deep learning-based soft sensor modeling

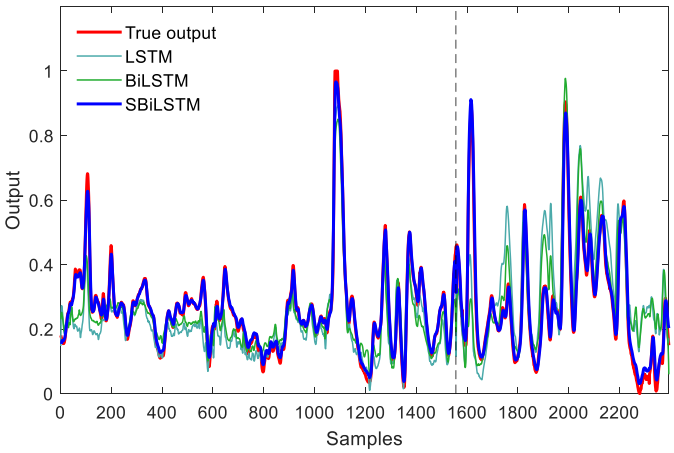

Fig. 8. Comparison of true quality output and predicted quality output of the debutanizer column process by LSTM, BiLSTM, and SBiLSTM during training phase and test phase.

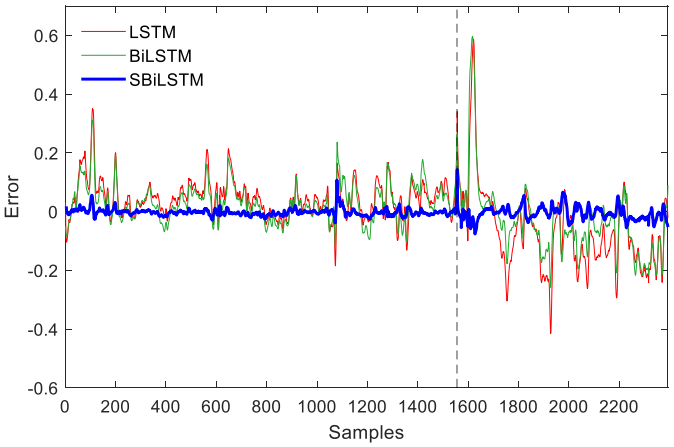

Fig. 9. Comparison of prediction errors of the debutanizer column process by LSTM, BiLSTM, and SBiLSTM during training phase and test phase.

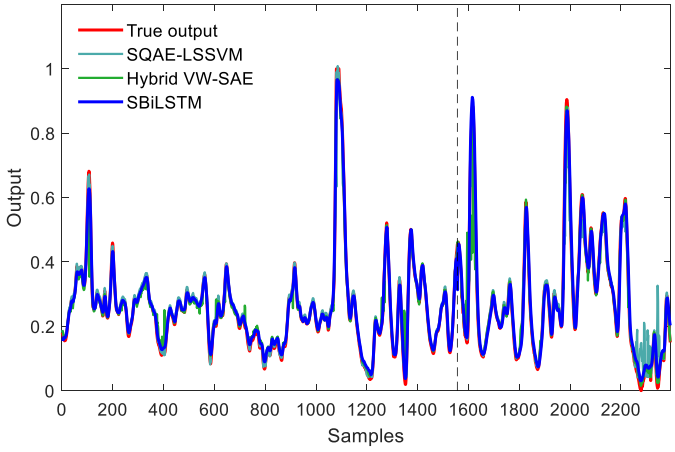

Fig. 10. Comparison of true quality output and predicted quality output of the debutanizer column process by SQAE-LSSVM, Hybrid VW-SAE, and SBiLSTM during training phase and test phase.

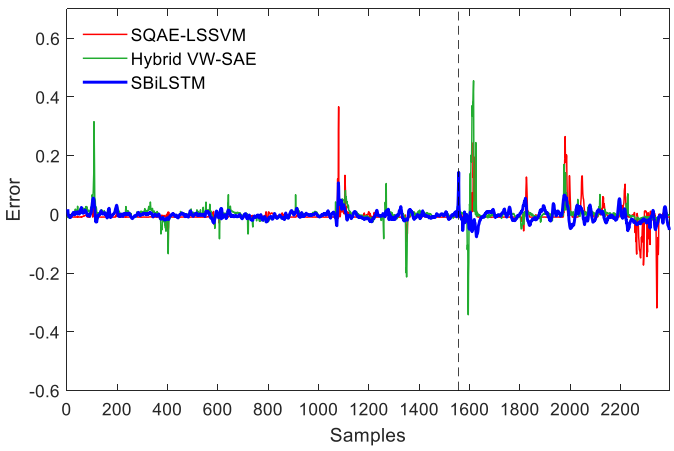

Fig. 11. Comparison of prediction errors of the debutanizer column process by SQAE-LSSVM, Hybrid VW-SAE, and SBiLSTM during training phase and test phase.

methods, Figs. 8 and 9 provide a direct view of the prediction results and the prediction errors on the entire debutanizer column dataset using LSTM, BiLSTM, and SBiLSTM network models, respectively. As can be seen from the figures, the soft sensor models based on LSTM and BiLSTM can only make very rough and inaccurate butane content predictions. Although BiLSTM model provides a slightly better result than LSTM, both LSTM and BiLSTM soft sensor model generate similar predictions that largely deviate from true output. It is obvious from Fig. 9 that the prediction error of the SBiLSTM network is much lower than that of LSTM and BiLSTM as it is much overlapping with the true output, resulting in a prediction error very close to 0. Compared to the LSTM and BiLSTM models, the SBiLSTM network is more effective to investigate nonlinear dynamic features latent in both process variables and quality variables in high-dimensional data for quality prediction, as it incorporates additional consideration on the nonlinear dynamic information in historical quality variables.

In nowadays industrial processes, data are often high-dimensional, nonlinear, and dynamic due to the evolving process complexity and scale. Take the debutanizer column process as an example, the debutanizer column process comprises sophisticated, dynamic distribution of temperature and pressure throughout the debutanizer column where the continuous stream of hydrocarbon mixture flows through. Even though temperature and pressure are known as significant process variables affecting the butane content in the debutanizer column, there exists complex nonlinear dynamic relationship among process variables and quality variables, which traditional deep learning methods fail to extract. In particular, dynamic information that are related to higher-order process dynamics are intrinsic among process variables and quality variables. By considering the dynamics between and within process variables and quality variables, the SBiLSTM network is able to capture quality and process dynamics effectively and predict quality output accurately in large-scale industrial processes even when quality variables are highly inner correlated with dynamics and nonlinearity.

Figs. 10 and 11 further compare the SBiLSTM network model with state-of-the-art data-driven soft sensor models including SQAE-LSSVM [16] and Hybrid VW-SAE [15]. In addition, Table III tabulated the test RMSE and test R-squared values of SBiLSTM and other deep learning-based soft sensor models on the debutanizer column dataset for

TABLE III

TEST RMSE AND R^2 VALUES OF THE SBiLSTM AND OTHER DEEP LEARNING-BASED MODELS ON THE DEBUTANIZER COLUMN

Method	RMSE	R^2
LSTM	0.1616	0.2792
BiLSTM	0.1397	0.4617
SQAE-LSSVM	0.0352	0.9521
Hybrid VW-SAE	0.0308	0.9615
SBiLSTM	0.0175	0.9915

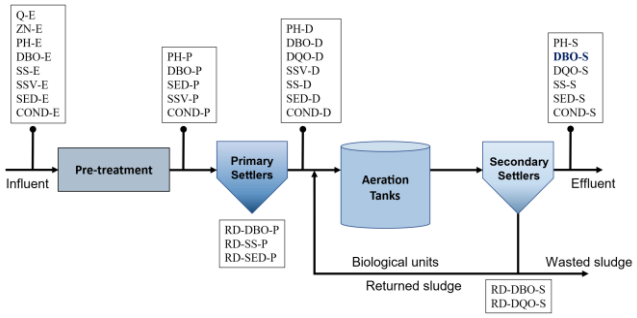

Fig. 12. Flowchart of the wastewater treatment process.

fair comparison. Although SQAE and Hybrid VW-SAE both exhibit exceptional performance on the quality prediction as shown in Fig. 10, occasional large prediction errors still occur throughout the training and test phases. In contrast, the SBiLSTM is much stabler of which the variance of its prediction errors is apparently lower. Comparing the quantitative error metrics in Table III, the SBiLSTM network resulted in an R-squared value of 99.15%, while SQAE-LSSVM and Hybrid VW-SAE obtained 95.21% and 96.15%, respectively. This further validated the improved effectiveness of the SBiLSTM network. As mentioned, many state-of-the-art deep learning-based soft sensor models focus on increasing the exploitation and utility of dynamics of input variables but neglected the past quality information that is readily available. In the proposed SBiLSTM soft sensor model, extended dynamic quality information latent in the historical quality variables are comprised by incorporating a moving window up to k time steps in the network input. By the nonlinear gates in the SBiLSTM units, useful nonlinear features can then be extracted from the dynamic quality information and input variables, and be learned iteratively by the soft sensor model. With dropout optimization, stable, efficient, and effective quality prediction that exploits dynamic quality information is then provided. The ability to extract and learn latent dynamic information from process variables and dynamic quality variable hence lead to the superiority of the proposed SBiLSTM network over other deep learning-based soft sensor model counterparts.

B. Wastewater Treatment Process

The industrial wastewater treatment process is a biochemical process that removes pollutants from industrial wastewater. Many industrial processes, including the debutanizer column process, require the wastewater treatment plant (WWTP) to treat its effluent discharge to comply with regulations [30]. However, controlling and predicting the output quality of treated wastewater in the complex, multistage industrial wastewater treatment process can be challenging. Fig. 12 shows the flowchart of the wastewater treatment process and Table IV gives the description of the process variables. As can be seen, the wastewater treatment process consists of multiple stages where various process variables are measured repeatedly throughout the entire process. This implies inherent dynamics among process variables of the same kind. For example, the measured suspended solids SS-*E*, SS-*D*, and SS-*S* are dynamically related and contain latent information about performance of the two settlers, since they are measured before pretreatment, after primary settlers, and after secondary settlers, respectively. Such dynamic latent information can be of complex structure and with high nonlinearity which model-based soft sensor models and traditional modeling methods fail to capture. Moreover, although repeated measurements of process variables suggest latent information about the process, it generates high-dimensional data that adds difficulty to modeling and analysis. It is also difficult to mine and extract the latent process information from data. Because of these challenges, soft sensor model is desirable to learn nonlinear dynamic latent features for effluent quality prediction in the wastewater treatment process.

TABLE IV

DESCRIPTION OF PROCESS VARIABLES OF THE WASTEWATER TREATMENT PROCESS

Process Variables	Description
Q-E	Input flow to plant
ZN-E	Input Zinc to plant
PH	pH value
DBO	Biological demand of oxygen
DQO	Chemical demand of oxygen
SS	Suspended solids
SSV	Volatile suspended solids
SED	Input sediments
COND	Conductivity

In this case study, we use the urban WWTP dataset in the UCI Machine Learning Repository, which is a set of field data collected from daily sensor measurements in an urban WWTP [31]. The WWTP dataset has been widely used by many researchers to evaluate data-driven models on a practical effluent prediction problem [32]–[35]. Instead of installing sensors at different stages of a process, the WWTP dataset collects data from repeated measurement of a fixed set of process variables throughout the process. Therefore, this case study also allows us to evaluate the robustness of the SBiLSTM model in a different instrumentation and measurement environment. In the WWTP dataset, there are a total number of 400 valid samples, 37 process variables, and one quality variable, where the quality variable y is the output biological demand of oxygen, DBO-*S*, which is an important indicator of wastewater quality. We use the first 70% of the data (280 samples) for model training and the remaining 30% (120 samples) for testing. We follow modeling procedure and grid search setting similar to Section IV-A with the same computational equipment.

After hyperparameter tuning with grid search, the SBiLSTM network is finally set with number of hidden neurons $n = 50$,

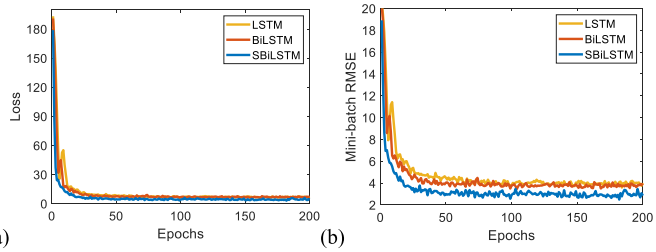

Fig. 13. (a) Training loss of LSTM, BiLSTM and SBiLSTM network over epoch; (b) Mini batch RMSE of LSTM, BiLSTM, and SBiLSTM network over epoch on the wastewater treatment process.

mini-batch size $m = 20$, maximum number of epochs $= 200$, initial learning rate $\eta = 0.01$, and dynamic order $k = 4$. Note that the number of hidden neurons required for this case study is more than that of the debutanizer column because of the increased number of process variables hence an increased input dimension. Therefore, given an input dimension five times, such as in the previous case study, the SBiLSTM network model in fact allows and retains a relatively small number of hidden neurons to perform efficient and stable soft sensor modeling. Moreover, a dynamic order of $k = 4$ indicates a highly dynamic process environment of the wastewater treatment process. From Fig. 13, the training loss and the mini-batch RMSE of the SBiLSTM network both descend and converge smoothly to their minima during the training phase, indicating a stable model training process. Similar to the previous case study, the training loss and mini-batch RMSE of LSTM and BiLSTM converge at a higher level and are more fluctuated than the SBiLSTM. This indicates that the LSTM and BiLSTM exhibit inferior model performance and stability than the SBiLSTM. It is also important to note that the soft sensor modeling algorithm takes only 10 s to run in this case study. This demonstrates the efficiency of the proposed SBiLSTM network even when handling high-dimensional data.

Fig. 14 shows the quality prediction results of the wastewater treatment process. It can be clearly observed that the predicted outputs closely overlap with the true output, indicating effective learning of latent process representations from the data. Even though the process variables and quality variables are highly fluctuating in the wastewater treatment process, the SBiLSTM-based soft sensor model exhibits efficacy to extract useful dynamic information for accurate quality prediction. Furthermore, the prediction results on both the debutanizer column process and the wastewater treatment process have demonstrated stability and applicability of the SBiLSTM in varied types of industrial processes. The major reason why the SBiLSTM model is particularly efficient and stable is the reduced number of hidden neurons enabled by both the bidirectional architecture and the dropout optimization in the SBiLSTM model. On the one hand, the bidirectional structure allows capture of additional dynamic process information that permits reduced network size to achieve desired effectiveness and efficiency. On the other hand, the dropout optimization stabilizes and regularizes the neural network by preventing excessive co-adaptation of neurons such that model overfitting is avoided. The novel model architecture thus resulted in

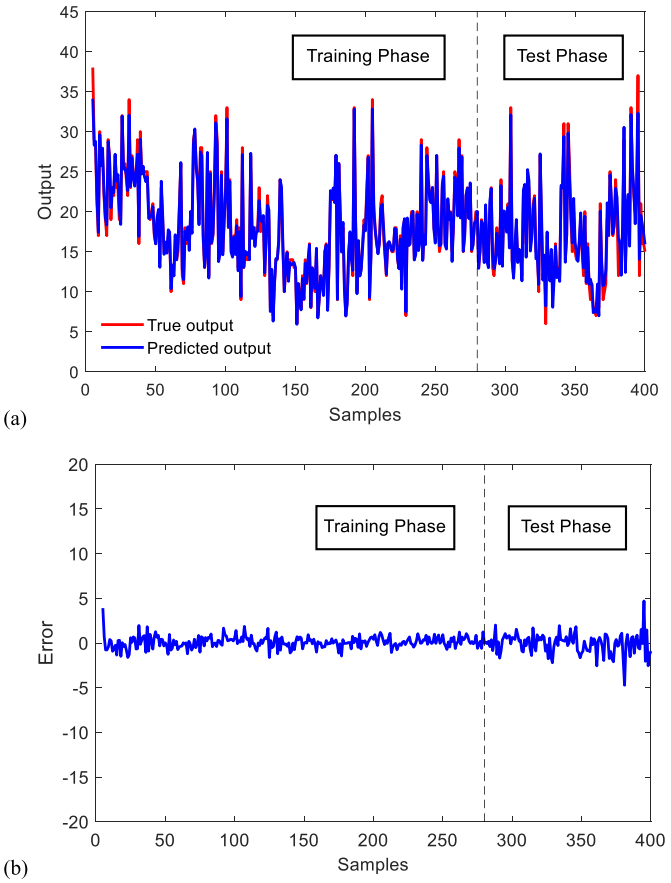

Fig. 14. (a) Comparison of true quality output and predicted quality output of the wastewater treatment process by SBiLSTM network model during training phase and test phase. (b) Model prediction errors of SBiLSTM network model on the quality prediction of the wastewater treatment process during training phase and test phase.

the distinctive efficacy and stability favorable to soft sensor modeling.

The quality prediction results of the two case studies have clearly shown that the SBiLSTM network can exploit additional dynamic quality information adaptively from different types of industrial processes with different instrumentation and measurement settings. In fact, many industrial processes, such as chemical processes and manufacturing processes, involve sophisticated dynamic process environments due to multistage, large-scale process operations [34], [36]. Provided significant process dynamics throughout the process system, our proposed methodology is able to effectively exploit nonlinear dynamic patterns latent in process and quality variables. In this regard, the SBiLSTM network is effective and widely applicable to dynamic industrial processes.

For performance comparison, the original LSTM and BiLSTM are also adopted for soft sensor modeling of the WWTP data. Figs. 15 and 16 shows the prediction results and the prediction errors of the soft sensors based on LSTM, BiLSTM, and SBiLSTM network, respectively. As shown in the figures, although the LSTM and BiLSTM show similar model performance to the SBiLSTM network in the training phase, they fail to make accurate predictions throughout the test phase. This implies that traditional dynamic deep

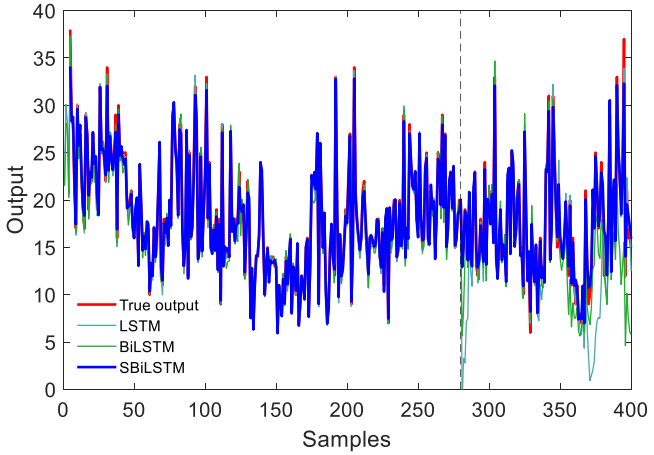

Fig. 15. Comparison of true quality output and predicted quality output of the wastewater treatment process by LSTM, BiLSTM, and SBiLSTM during training phase and test phase.

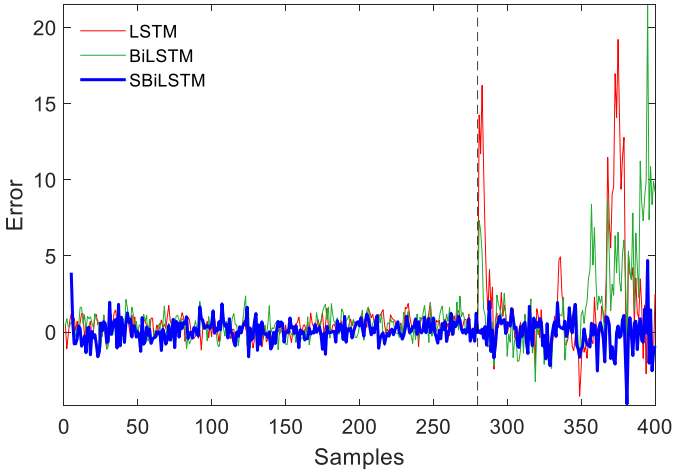

Fig. 16. Comparison of prediction errors of the wastewater treatment process by LSTM, BiLSTM, and SBiLSTM during training phase and test phase.

learning-based soft sensor models can be unstable and unreliable due to overfitting issues and limited ability to capture latent dynamic information among process variables and quality variables. In contrast, it is noticeable that the SBiLSTM network has almost exact predictions with the true output values in both training and test phases. This indicates an improved dynamic nonlinear feature extraction and feature learning ability of the SBiLSTM network due to the employment of extended quality variables and its novel structure.

To compare with state-of-the-art deep learning models, Figs. 17 and 18 plotted the quality prediction and prediction errors of the wastewater treatment process by SQAE-LSSVM, Hybrid VW-SAE, and SBiLSTM during training phase and test phase. From Fig. 18, it is obvious that the SQAE-LSSVM and Hybrid VW-SAE are slightly less accurate when predicting the quality output of the wastewater treatment process. In addition, it is surprising that although both SQAE-LSSVM and Hybrid VE-SAE successfully captured the trend and level of the predictions, both models present relatively conservative predictions throughout the training and test stages as shown in Fig. 17. The difference between the results of SBiLSTM and the two methods can be accounted by the ability of the

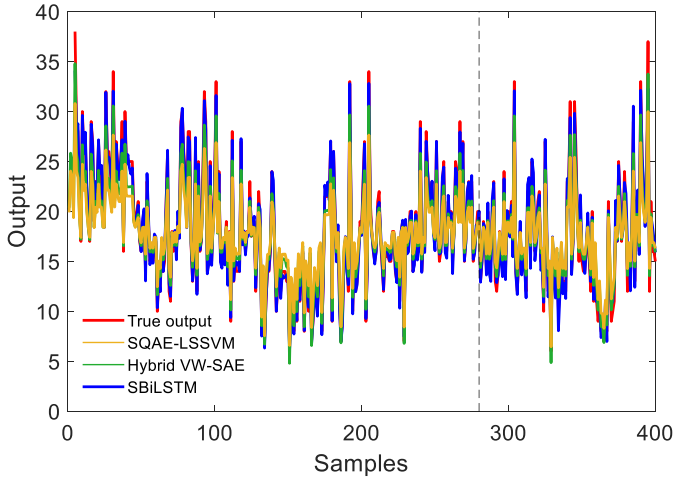

Fig. 17. Comparison of true quality output and predicted quality output of the wastewater treatment process by SQAE-LSSVM, Hybrid VW-SAE, and SBiLSTM during training phase and test phase.

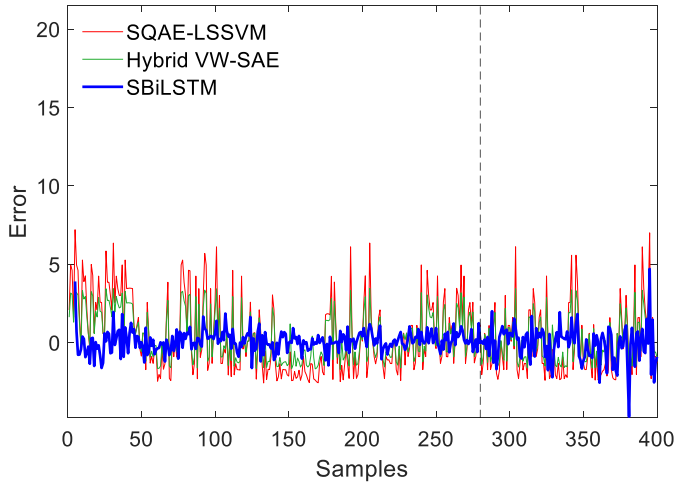

Fig. 18. Comparison of prediction errors of the wastewater treatment process by SQAE-LSSVM, Hybrid VW-SAE, and SBiLSTM during training phase and test phase.

SBiLSTM to learn and comprehend higher-order dynamics in highly dynamic processes. In wastewater treatment process, sudden or abrupt changes constitutes not only short-term influences but also long-term effects underlying in dynamic latent information. The ability of the SBiLSTM to capture and store both short-term and long-term memory effectively hence explains its superior prediction ability over the SQAE-LSSVM and Hybrid VE-SAE methods on wastewater treatment process.

Table V lists the RMSE and R-squared values of the SBiLSTM network and some state-of-the-art deep learning-based soft sensors on the WWTP data during the test phase. From Table V, the SBiLSTM network resulted in an RMSE of 1.1709 and an R-squared value of 95.80%. We can see that the SBiLSTM network model has better prediction performance than LSTM, BiLSTM, SQAE-LSSVM, and Hybrid VW-SAE. The results indicate that the SBiLSTM network model is more efficacious and powerful than both traditional dynamic deep learning-based soft sensor models and state-of-the-art counterparts in extracting and utilizing useful dynamic latent features

TABLE V
TEST RMSE AND R^2 VALUES OF THE SBiLSTM AND OTHER DEEP LEARNING-BASED MODELS ON THE WASTEWATER TREATMENT PROCESS

Method	RMSE	R^2
LSTM	4.9527	0.2453
BiLSTM	4.3368	0.4213
SQAE-LSSVM	2.2173	0.8487
Hybrid VW-SAE	1.5778	0.9289
SBiLSTM	1.1709	0.9580

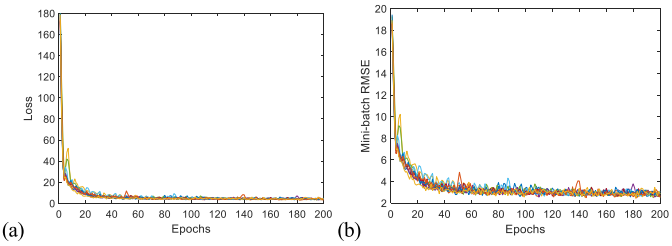

Fig. 19. (a) Training loss and (b) mini batch RMSE of SBiLSTM network on the wastewater treatment process with ten-fold cross validation.

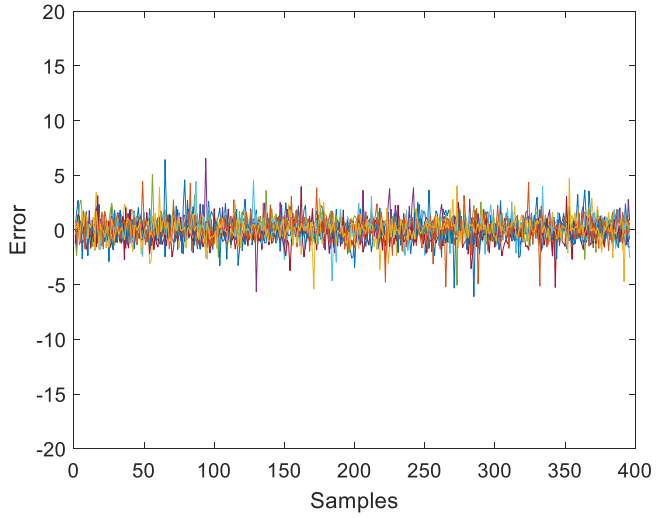

Fig. 20. Prediction errors of the wastewater treatment process by SBiLSTM with ten-fold cross validation.

for quality prediction. Although the network size reduces as the number of hidden neurons reduces, the SBiLSTM network model not only retains but shows better prediction performance with improved stability and efficiency. This further demonstrates the enhancement of knowledge representation learning ability, information utility, and model efficacy after introducing the improved model structure. As a result, the proposed SBiLSTM network exhibits superior quality prediction accuracy and efficiency over other deep learning-based soft sensor model counterparts.

Since there are only 400 samples with 37 process variables in this case study, we further validate our result by a ten-fold cross validation, which constitutes 360 samples for training and 40 samples for testing for each of the ten-fold partitions. The resulted training loss and mini-batch RMSE are plotted in Fig. 19 and the prediction errors of the ten-fold cross validation are plotted in Fig. 20. From Fig. 19, the training loss and training RMSE of the ten-fold cross validations fluctuated very slightly with a clear common trajectory of convergence, indicating a very stable model performance throughout the ten validations. Furthermore, despite existence of some spikes, the prediction errors of the ten-fold cross validations have its majority lie within -3 and 3 with an obvious mean of 0. Both observations from the ten-fold cross validation exemplified the stability and validity of the SBiLSTM network in data-driven dynamic soft sensor modeling.

V. CONCLUSION

In this article, a SBiLSTM network is proposed for data-driven dynamic soft sensor modeling. The SBiLSTM incorporates extended quality information utilization with quality information window up to k time steps. With bidirectional architecture and dynamic quality supervision structure, the SBiLSTM network is capable of extracting and utilizing nonlinear dynamic latent information from both process variables and quality variables in industrial process data. In addition, model instability and inefficiency are avoided by resorting to dropout optimization and by enabling reduced hidden neurons while maintaining model effectiveness. The effectiveness of the proposed SBiLSTM network soft sensor model was demonstrated through two case studies on the debutanizer column process and the wastewater treatment process. Results show that SBiLSTM outperforms LSTM, BiLSTM, and other state-of-the-art deep learning-based soft sensor model counterparts in terms of prediction accuracy. For further research, experiments to study random noise and uncertainties on the proposed SBiLSTM network are worth investigating. In addition, an ablation study can be conducted to understand the impact of the proposed design novelties on model performances. Besides, exploring soft sensors based on other data-driven methods, such as Bayesian learning, fuzzy learning, autoencoders, and kernel methods [4], [25] would be interesting topics, which deserve further investigation in the future.

REFERENCES

[1] B. Shen and Z. Ge, "Supervised nonlinear dynamic system for soft sensor application aided by variational auto-encoder," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6132–6142, Sep. 2020, doi: 10.1109/TIM.2020.2968162.

[2] L. Yao and Z. Ge, "Dynamic features incorporated locally weighted deep learning model for soft sensor development," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2021.3073702.

[3] X. Yuan, L. Li, K. Wang, and Y. Wang, "Sampling-Interval-Aware LSTM for industrial process soft sensing of dynamic time sequences with irregular sampling measurements," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10787–10795, May 2021, doi: 10.1109/JSEN.2021.3056210.

[4] Y. Liu and M. Xie, "Rebooting data-driven soft-sensors in process industries: A review of kernel methods," *J. Process Control*, vol. 89, pp. 58–73, May 2020, doi: 10.1016/j.jprocont.2020.03.012.

[5] M. G. Xibilia, M. Latino, Z. Marinkovic, A. Atanaskovic, and N. Donato, "Soft sensors based on deep neural networks for applications in security and safety," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7869–7876, Oct. 2020, doi: 10.1109/TIM.2020.2984465.

[6] B. Shen and Z. Ge, "Weighted nonlinear dynamic system for deep extraction of nonlinear dynamic latent variables and industrial application," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3090–3098, May 2021, doi: 10.1109/TII.2020.3027746.

[7] X. Jiang and Z. Ge, "Augmented multidimensional convolutional neural network for industrial soft sensing," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3075515.

[8] Y. Fan, B. Tao, Y. Zheng, and S.-S. Jang, "A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 3972–3979, Jul. 2020, doi: 10.1109/TIM.2019.2947126.

[9] X. Yuan, S. Qi, and Y. Wang, "Stacked enhanced auto-encoder for data-driven soft sensing of quality variable," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7953–7961, Oct. 2020, doi: 10.1109/TIM.2020.2985614.

[10] B. Zhang, Y. Han, B. Yu, and Z. Geng, "Novel nonlinear autoregression with external input integrating PCA-WD and its application to a dynamic soft sensor," *Ind. Eng. Chem. Res.*, vol. 59, no. 35, pp. 15697–15706, Sep. 2020, doi: 10.1021/acs.iecr.0c02944.

[11] J. Zheng and Z. Song, "Mixture modeling for industrial soft sensor application based on semi-supervised probabilistic PLS," *J. Process Control*, vol. 84, pp. 46–55, Dec. 2019, doi: 10.1016/j.jprocont.2019.09.007.

[12] Y. Liu, B. Liu, X. Zhao, and M. Xie, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6478–6486, Aug. 2018, doi: 10.1109/TIE.2017.2786253.

[13] Y. Liu, B. Liu, X. Zhao, and M. Xie, "Development of RVM-based multiple-output soft sensors with serial and parallel stacking strategies," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2727–2734, Nov. 2019, doi: 10.1109/TCST.2018.2871934.

[14] I. Pisa, I. Santín, J. Vicario, A. Morell, and R. Vilanova, "ANN-based soft sensor to predict effluent violations in wastewater treatment plants," *Sensors*, vol. 19, no. 6, p. 1280, Mar. 2019, doi: 10.3390/s19061280.

[15] X. Yuan, C. Ou, Y. Wang, C. Yang, and W. Gui, "Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE," *Neurocomputing*, vol. 396, pp. 375–382, Jul. 2020, doi: 10.1016/j.neucom.2018.11.107.

[16] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3721–3730, Jun. 2020, doi: 10.1109/TII.2019.2938890.

[17] Z. Ge and X. Chen, "Supervised linear dynamic system model for quality related fault detection in dynamic processes," *J. Process Control*, vol. 44, pp. 224–235, Aug. 2016, doi: 10.1016/j.jprocont.2016.06.003.

[18] Y. Wang, Z. Pan, X. Yuan, C. Yang, and W. Gui, "A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network," *ISA Trans.*, vol. 96, pp. 457–467, Jan. 2020, doi: 10.1016/j.isatra.2019.07.001.

[19] 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 2019, doi: 10.1109/TII.2019.2902129.

[20] K. Greff, R. K. Srivastava, J. Koutnìk, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.

[21] C. A. Duchanoy, M. A. Moreno-Armendáriz, L. Urbina, C. A. Cruz-Villar, H. Calvo, and J. D. J. Rubio, "A novel recurrent neural network soft sensor via a differential evolution training algorithm for the tire contact patch," *Neurocomputing*, vol. 235, pp. 71–82, Apr. 2017, doi: 10.1016/j.neucom.2016.12.060.

[22] W. Ke, D. Huang, F. Yang, and Y. Jiang, "Soft sensor development and applications based on LSTM in deep neural networks," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–6, doi: 10.1109/SSCI.2017.8280954.

[23] J. Zhou, X. Wang, C. Yang, and W. Xiong, "A novel soft sensor modeling approach based on difference-LSTM for complex industrial process," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 2955–2964, May 2022, doi: 10.1109/TII.2021.3110507.

[24] C.-H. Chou *et al.*, "Physically consistent soft-sensor development using sequence-to-sequence neural networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2829–2838, Apr. 2020, doi: 10.1109/TII.2019.2952429.

[25] Q. Sun and Z. Ge, "A survey on deep learning for data-driven soft sensors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 5853–5866, Sep. 2021, doi: 10.1109/TII.2021.3053128.

[26] K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Math. Problems Eng.*, vol. 2013, pp. 1–11, May 2013, doi: 10.1155/2013/425740.

[27] K. Shibata and Y. Ikeda, "Effect of number of hidden neurons on learning in large-scale layered neural networks," in *Proc. ICCAS-SICE*, 2009, pp. 5008–5013.

[28] S. Hahn and H. Choi, "Understanding dropout as an optimization trick," *Neurocomputing*, vol. 398, pp. 64–70, Jul. 2020, doi: 10.1016/j.neucom.2020.02.067.

[29] L. Fortuna, S. Graziani, A. Rizzo, and M. G. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*. Berlin, Germany: Springer, 2007.

[30] S. Jafarinejad and S. C. Jiang, "Current technologies and future directions for treating petroleum refineries and petrochemical plants (PRPP) wastewaters," *J. Environ. Chem. Eng.*, vol. 7, no. 5, Oct. 2019, Art. no. 103326, doi: 10.1016/j.jece.2019.103326.

[31] D. Dua and C. Graff, "UCI machine learning repository," School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, Tech. Rep., 2019. [Online]. Available: http://archive.ics.uci.edu/ml

[32] 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," *Ind. Eng. Chem. Res.*, vol. 60, no. 38, pp. 13914–13926, Sep. 2021, doi: 10.1021/acs.iecr.1c01990.

[33] W. Li, M. Chu, and J. Qiao, "A pruning feedforward small-world neural network based on Katz centrality for nonlinear system modeling," *Neural Netw.*, vol. 130, pp. 269–285, Oct. 2020, doi: 10.1016/j.neunet.2020.07.017.

[34] H. Cheng, Y. Liu, D. Huang, C. Xu, and J. Wu, "A novel ensemble adaptive sparse Bayesian transfer learning machine for nonlinear large-scale process monitoring," *Sensors*, vol. 20, no. 21, p. 6139, Oct. 2020, doi: 10.3390/s20216139.

[35] Y. Liu, L. Yuan, D. Li, Y. Li, and D. Huang, "Process monitoring of quality-related variables in wastewater treatment using Kalman–Elman neural network-based soft-sensor modeling," *Water*, vol. 13, no. 24, p. 3659, Dec. 2021, doi: 10.3390/w13243659.

[36] D. Zhang, Z. Liu, W. Jia, H. Liu, and J. Tan, "Path enhanced bidirectional graph attention network for quality prediction in multistage manufacturing process," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1018–1027, Feb. 2022, doi: 10.1109/TII.2021.3076803.

Chun Fai Lui received the B.Eng. degree in systems engineering from the City University of Hong Kong, Kowloon, Hong Kong, as the college representative, in 2020, where he is currently pursuing the Ph.D. degree.

His current research interests include data-driven process monitoring and quality improvement in intelligent manufacturing systems.

Mr. Lui was a recipient of the Hong Kong Ph.D. Fellowship awarded by the Hong Kong Government.

Yiqi Liu (Member, IEEE) received the B.S. and M.S. degrees in control engineering from the Chemical University of Technology, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in control engineering from the South China University of Technology, Guangzhou, China, in 2013.

During his Ph.D., he was trained jointly with The University of Queensland, Australia. From 2013 to 2016, he was an Associate Professor with the Department of Automation, South China University of Technology. He is the author of more than 60 articles. His research interests include soft-sensors, fault diagnosis, and wastewater treatment and holds three patents.

Dr. Liu was a recipient of the EU-Horizon 2020 MSCA Individual Fellowship in 2019, the Hong Kong Scholar Award in 2016, and the Deutscher Akademischer Austausch Dienst Visiting Scholar Award in 2015.

Min Xie (Fellow, IEEE) received the Ph.D. degree in quality technology from Linkoping University, Linkoping, Sweden, in 1987.

He is currently a Chair Professor with the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong. He has authored or coauthored about 300 peer-reviewed journal articles and eight books on quality and reliability engineering. Over 50 Ph.D. students have graduated under his supervision.

Prof. Xie was the recipient of the prestigious Lee Kuan Yew (LKY) Research Fellowship in 1991. He has chaired many international conferences and given keynote speeches. He also serves as an Editor and an Associate Editor, and on the Editorial Board of many established international journals.