Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 7617010, 11 pages
https://doi.org/10.1155/2020/7617010

Research Article

Development of a Novel Soft Sensor with Long Short-Term Memory Network and Normalized Mutual Information Feature Selection

Dongfeng Li,[1] Zhirui Li,[2] and Kai Sun [3]

[1]*Shandong Experimental High School, Jinan 250001, China*
[2]*School of Microelectronics, Shandong University, Jinan 250001, China*
[3]*School of Electrical Engineering and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China*

Correspondence should be addressed to Kai Sun; sunkai79@qlu.edu.cn

Received 1 February 2020; Accepted 21 February 2020; Published 25 April 2020

Guest Editor: Ping Zhao

In this paper, a novel soft sensor is developed by combining long short-term memory (LSTM) network with normalized mutual information feature selection (NMIFS). In the proposed algorithm, LSTM is designed to handle time series with high nonlinearity and dynamics of industrial processes. NMIFS is conducted to perform the input variable selection for LSTM to simplify the excessive complexity of the model. The developed soft sensor combines the excellent dynamic modelling of LSTM and precise variable selection of NMIFS. Simulations on two actual production datasets are used to demonstrate the performance of the proposed algorithm. The developed soft sensor could precisely predict the objective variables and has better performance than other methods.

1. Introduction

Due to technological constraints, sensor characteristics, environmental factors, etc., many variables cannot be measured or the measurement frequency is very low in actual industrial processes. Soft measurement provides an excellent solution to construct mathematical models from easily measured variables to hard ones [1–3]. Neural networks (NNs) are advanced methods that can precisely model complex and nonlinear system and therefore have been widely used in soft sensors [4–6]. Heidari et al. [7] developed a new multi-layer perceptron (MLP) network to estimate nanofluid relative viscosity, which are more accurate than other NN structures. Sheela and Deepa [8] designed a synthesized model by combining self-organizing maps (SOMs) with MLP and then applied it to forecast the wind speed of a renewable energy process. He et al. [9] developed an auto-associative hierarchical NN for a soft sensor of chemical processes, and its application to a purified terephthalic acid solvent process demonstrated the effectiveness of the algorithm. Zabadaj et al. [10] proposed an effective soft sensor for the supervisory control of biotransformation production, and the efficiency of the approach was demonstrated. Rehrla et al. [11] developed a soft sensor method for estimating the active pharmaceutical ingredient concentration from the system data and the soft sensor model was tested in the three different continuous production lines. A novel approach of supervised latent factor analysis was proposed based on system data regression modelling, which can effectively predict heterogeneous variances, and soft sensors are established for quality estimate in the two case studies [12].

However, industrial systems are intrinsic complex and have high temporal correlations between dataset samples. That is, process data are time series with strong nonlinearities and dynamics, which increases the difficulty of modelling with conventional NNs. Recently, a powerful type of NN named long short-term memory (LSTM) was

designed to handle sequence dependence [13–15]. An LSTM network is more significant in learning long-term temporal dependencies since its memory cells can maintain its state over a long time and standardize the information moving into and out of the cell. Therefore, LSTM networks have been effectively used in many different fields, such as precipitation nowcasting [16], traffic forecasting [17], human action recognition [18], etc. Due to its advantages, LSTM has also been applied in soft sensor development in industrial processes. Yuan et al. developed a supervised LSTM network for a soft sensor and demonstrated the superiority of the proposed soft sensor by two actual industrial datasets [19]. Sun proposed a new LSTM network by combining unsupervised feature selection and supervised dynamic modelling methods for a soft sensor and validated the network by a practical CO_2 absorption column [20].

The rapid evolution of distributed control systems (DCSs) presents us a lot of data, but also another trouble in nonlinear soft sensing: excessive input variables. If the NN was trained with excessive input variables, the amount of calculations will increase and more computing power is required. In the meantime, the prediction accuracy of the NN is worsened due to extraneous variables creating additional noise in the dataset. Hence, many researchers have focused on the efficiency of variable selection approaches for soft sensors [21–23]. In recent years, mutual information (MI)-based variable selection approaches have been widely studied due to their efficacy and ease of realization [24]. Hanchuan et al. [25] proposed a minimal redundancy maximal correlation criterion to reduce redundancy and apply primitive methods of relevance and redundancy to select significant input variables. Estevez et al. [26] proposed an enhanced version of MIFS and minimal redundancy maximal correlation that imported the normalized MI as an evaluation of redundancy. The developed NMIFS algorithm showed better performance by compensating for the MI partial to multiple features and limiting its setpoint range [0, 1].

This paper develops a new soft sensor algorithm by combining LSTM with NMIFS, in which the NMIFS is used to compress input variables of LSTM. The primary contributions of the paper are summarized as follows:

(1) A novel feature selection approach for LSTM with NMIFS is designed. The developed method can effectively reduce the excessive complexity caused by redundant candidate variables and then improve the modelling performance of LSTM.

(2) The developed soft sensor algorithm is implemented in two practical industrial processes.

(3) Comparative simulation results demonstrate that the developed soft sensor model has better performance and flexibility in performing feedback control.

This paper is arranged as follows: Section 2 presents background theories of the NMIFS and LSTM, and Section 3 describes the development of the presented approach. Section 4 presents the simulation results and an analysis of the developed soft sensor with datasets of actual processes. Some conclusions are given in Section 5.

2. Theoretical Overview

2.1. Input Variable Selection Techniques. The existence of redundant input variables in the training of the NN often complicates the model, deteriorates the accuracy, and even brings about overfitting. The goal of the input variable feature selection (IVFS) is to exactly select n variables from the initial candidate variable set C as the input variable set S in the modelling, where C includes multifarious input variables of the algorithm. During variable selection, variables that have less influence on target variable will be deleted from S.

The IVFS algorithm can be applied in a variety of means: (1) sequential forward selection, selecting an input from C to join S every time until the prediction accuracy of the model is no longer improved; (2) sequential backward selection, where S initially includes all input variables and input variables are deleted one at a time until the model performance is no longer improved; or (3) global optimization, which finds the optimal solution among all the variable selection approaches.

It can be demonstrated that if there are m candidate variables in C, the selection of the m input variables results in $(2^m - 1)$ subsets in total. S is hard to be found with large number of candidate variables. Based on this consideration, a statistical indicator to calculate the extent of dependence between input and output variables is selected, and then the input variable before modelling with NNs is selected. This method of separating variable selection procedures from model calibration procedures can produce a more efficient IVFS algorithm, and the resulting S has wider applicability to different NN algorithms. It is worth noting that the effectiveness of the IVFS approach is based on the statistical standard applied.

MI is considered as an excellent evaluation standard because it is a random measure and does not make assumptions about the structure of the dependencies between variables. MI is also found to be impervious to data transformations and noise.

2.2. Normalized Mutual Information. Suppose that there are two random variables X and Y, where X is the input variable and the output variable Y depends on X. The definition of MI of $I(X;Y)$ for a continuous variable can be shown as follows [27]:

$$I(X;Y) = \iint p(X,Y)\log\frac{p(X,Y)}{p(X)p(Y)}\,\mathrm{d}X\mathrm{d}Y, \tag{1}$$

where $p(X,Y)$ is the joint probability density function (PDF) of two variables and $p(X)$ and $p(Y)$ are the marginal PDFs of X and Y. Figure 1 shows the entropy of X and Y and its relationship to their MI, in which $H(X)$ and $H(Y)$ are entropies, and $H(X \mid Y)$ and $H(Y \mid X)$ are conditional entropies, respectively.

Generally speaking, MI has three basic attributes:

(1) Symmetry: $I(X;Y) = I(Y;X)$. The quantity of information abstracted from Y about X is equal to that

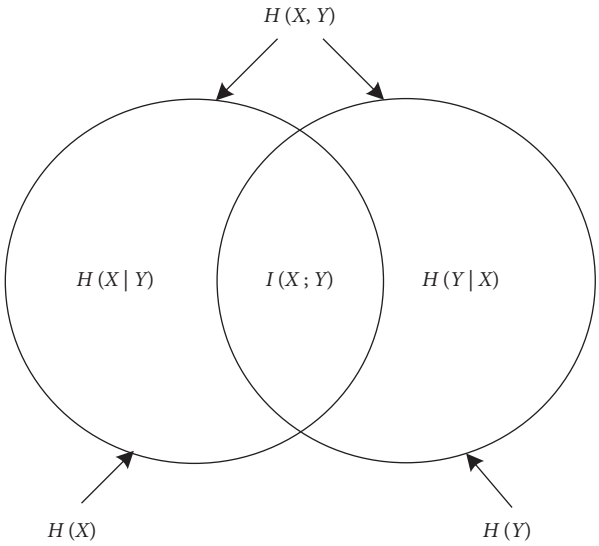

Figure 1: Structure diagram of entropy and MI.

from X about Y. The only difference is the angle of the observer.

(2) Positive: $I(X;Y) \geq 0$. Extracting information about one event from another, the worst case is zero information ($I(X;Y) = 0$). Being aware of one event does not strengthen the uncertainty of another.

(3) Extremum: $I(X;Y) \leq H(X)$, $I(Y;X) \leq H(Y)$. The quantity of information abstracted from one case about another is at maximum same as the entropy of the other case, rather than exceeding the amount of information contained by the other event itself.

The MI $I(X;Y)$ provides dependencies for X and Y measurements and provides reference information for variable selection algorithms, which makes the computation of MI a crucial procedure in MI-based input variable selection approaches [28, 29]. However, the mathematical expression form of the PDF in equation (1) is unconscious in practical problems. A variety of approximate prediction algorithms of MI have been extensively researched to analyze PDFs. For example, kernel density estimation (KDE) is an advanced technique that superposes a basis function on each point of the feature data, usually a Gaussian function. The PDF approximation can then be obtained by adopting an envelope of all the basic functions superimposed on each point. Although these kinds of algorithms bring superior approximation results, the computation load is very high, especially in large-scale problems. Histogram methods provide another competitive method, with admissible precision and significantly more computational performance than KDE methods.

When MI is applied to practical cases, the calculation results fluctuate greatly, and it is difficult to directly compare the similarity between several variables and the target variables used as indicators [30]. This paper introduces a method to normalize MI. There are several methods of doing so. The general idea is to use entropy as the denominator to regulate the value of MI to between 0 and 1. One common implementation is the following formula:

$$N(X,Y) = 2\frac{I(X,Y)}{H(X) + H(Y)}. \tag{2}$$

Then, NMI can be used to evaluate the resemblance between candidate and target input variables.

2.3. Long Short-Term Memory. The LSTM network is applied to predicting target variables with relatively long intervals and postponements in the time series. The structure of neurons in LSTM is shown in Figure 2. It includes a cell state and three gate settings: the cell state is used to record neuron status, the input and output gates are used to receive and output parameters, respectively, and the forget gate is used to dominate the degree of forgetting of the previous unit state [31, 32].

The detailed structure and operation mechanism of LSTM are shown in Figure 3. The forgotten part of the memory unit is decided by the input x_t in the forgetting gate together with the state memory unit S_{t-1} and the intermediate output h_{t-1}. The retention vector in the memory unit is determined by the changed x_t in the input gate through the sigmoid and tanh functions. The intermediate output h_t is determined by the updated S_t and output o_t. The calculation formula is as follows:

$$\begin{aligned} f_t &= \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right), \\ i_t &= \sigma\left(W_{ix}x_t + W_{ih}h_{t-1} + b_i\right), \\ g_t &= \varnothing\left(W_{gx}x_t + W_{gh}h_{t-1} + b_g\right), \\ o_t &= \sigma\left(W_{ox}x_t + W_{oh}h_{t-1} + b_o\right), \\ S_t &= g_t \odot i_t + S_{t-1} \odot f_t, \\ h_t &= \varnothing\left(S_t\right) \odot o_t, \end{aligned} \tag{3}$$

where f_t, i_t, g_t, o_t, h_t, and S_t are the states of the forgetting gate, input gate, input node, output gate, intermediate output, and status unit, respectively; W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{gx}, W_{gh}, W_{ox}, and W_{oh} are the matrix weight multiplied by input x_t of the corresponding gate and the intermediate output h_{t-1}, respectively; b_f, b_i, b_g, and b_o are the biases of the corresponding gates; $\odot$ indicates that the elements in the vector are multiplied by bits; and σ and $\varnothing$ represent the transformation of the sigmoid and tanh function, respectively.

3. Development of NMIFS-LSTM

The evaluation function plays a pivotal role in the MI feature selection, which directly affects the final performance of the algorithm. The method of selecting the variable with the most MI of output variable Y and input variable X_i is the most direct solution. The evaluation function is shown in

$$R = I(X_i;Y). \tag{4}$$

The MIFS [33] method introduces penalty terms based on the measure of relevance, which incorporates correlation and redundancy between variables. The evaluation function is shown in

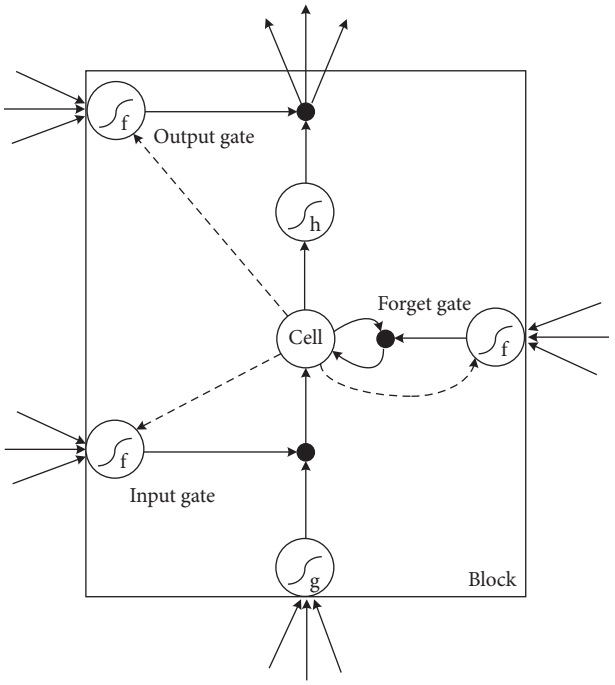

Figure 2: LSTM neuron structure.

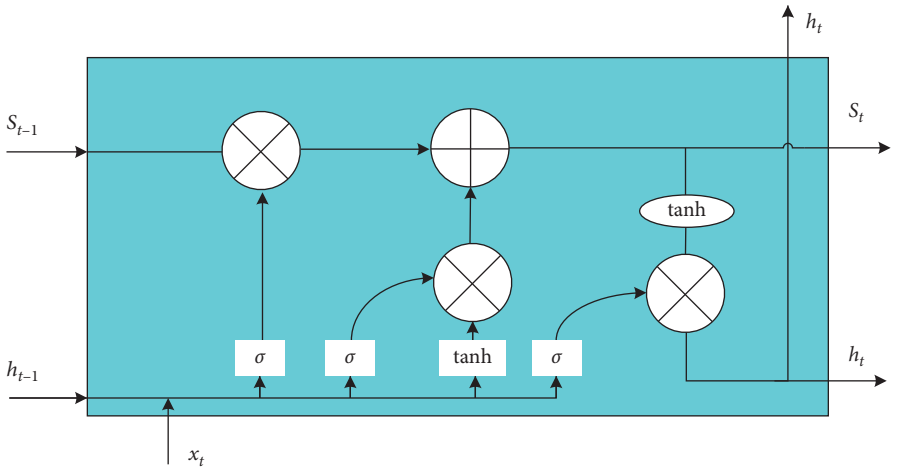

Figure 3: LSTM internal unit structure.

$$R = I(X_i; Y) - \beta \sum_{x_{s\in S}} I(X_i; X_s), \tag{5}$$

where S is the selected feature subset, X_s is the selected feature, and parameter β controls the degree of penalty for redundant items.

In order to reduce the dependence on parameter β, Kwak and Chong-Ho Choi [34] proposed the method of MIFS-U, and the evaluation function expression is exhibited as

$$R = I(X_i; Y) - \beta \sum_{x_{s\in S}} \frac{I(X_s; Y)}{H(X_s)} I(X_i; X_s). \tag{6}$$

Hanchuan et al. [25] enhanced MIFS and developed the minimal redundancy maximal relevance algorithm, which establishes a relationship between sample size and parameter β. The mean value of MI is used as the redundancy evaluation index to avoid the selection of parameter β. The evaluation function can be shown as

$$R = I(X_i; Y) - \frac{1}{|S|} \beta \sum_{x_{s\in S}} I(X_i; X_s). \tag{7}$$

The standardized MI between variables was defined by Estevez et al. [26], and the NMIFS algorithm was proposed. Its evaluation function can be expressed as

$$R = I(X_i; Y) - \frac{1}{|S|} \beta \sum_{x_{s\in S}} NI(X_i; X_s), \tag{8}$$

where the standardized MI is performed as equation (9). The regularized MI compensates for the bias of MI to

Input: dataset imprent MT shadow
Output: predicted value
Begin algorithm
Initialize
LSTM is trained to determine network hyperparameters and network structure;
Set $F = n$; $S =$ empty set ($n =$ number of input variables);
Computation of NMI with LSTM;
For $i = 1{:}j$ (j is frequency of the stop criterion)
$\forall f_i \in F$ compute $I(L; f_i)$;
Find a first variable f_i that maximizes $I(\mathrm{L}f_i)$ and obtain RMSE;
Set $F \longleftarrow F/\{f_i\}$, set $S \longleftarrow \{\{f_i\}\}$, set $i = 1$;
Choose the next variable f_i = argmax, $f_i \in F - S(\min f_s \in S(I(f_i, f_s; L)))$ and obtain new RMSE;
set $F \longleftarrow F/\{f_i\}$; set $S \longleftarrow S \cup \{f_i\}$, $j = j + 1$;
if new RMSE > RMSE
Break
Else
RMSE = newRMSE, return and select the next variable;
End if
Repeat until $|\mathrm{S}| = j$;
End for
Retrain with selected subset
Calculate predicted value
End algorithm

Algorithm 1: Pseudocode of NMIFS-LSTM.

multivalued variables, and the regularized MI value is strictly restricted to the interval of [0, 1].

$$NI(X_i; X_s) = \frac{I(X_i; X_s)}{\min\{H(X_i), H(X_s)\}}. \tag{9}$$

In this paper, a novel variable selection method of NMIFS-LSTM is developed. This method combines NMIFS and LSTM, and after that, the root mean square error (RMSE) of the LSTM network is used as the evaluation standard. The proposed algorithm aims to eliminate redundant variables and improve model accuracy. The pseudocode of NMIFS-LSTM can be shown in Algorithm 1.

The operating mechanism of the NMIFS-LSTM algorithm is mainly divided into two parts. In the algorithm, we build a model for prediction by LSTM with NMI for variable selection. The LSTM NN is trained to determine network hyperparameters and structure. Parameter F is set to "initial set of n variables" and parameter S set to "empty set." The calculation method is NMI with the LSTM of RMSE and the first variable is chosen. We continue to choose next variables every step until the model gets worse or meets the stop criterion. Finally, the selected subset is modeled and the predicted value is obtained. The flowchart of the developed NMIFS-LSTM-based soft sensor model is shown in Figure 4.

4. Simulation Results and Discussion

In this paper, all algorithms use a common dataset with the same variable selection method after several trials in the same simulation environment setting. All established models were simulated in the same experimental environment. The program for algorithm simulation was coded in MATLAB 2019 and run under a Windows 8.1 operating system. The simulation results are recorded with the following standards:

(1) Model size (MS) means the number of candidate variables selected in the ultimate algorithm

(2) PMSE means the mean square error (MSE) is a measure that reflects the difference between the actual and value predicted value and can be calculated as follows:

$$\mathrm{PMSE} = \frac{1}{n_t} \sum_{i=1}^{n_t} (y_i - \hat{y}_i)^2, \tag{10}$$

where y_i and $\hat{y}_i$ are the actual value and predicted value in the algorithm model of the output variable, respectively, and n_t represents the number of datasets in the testing samples

(3) Coefficient of determination (R^2) denotes the square of sample correlation coefficients between the real value and prediction value

4.1. Application to a Debutanizer. To verify the efficacy of the developed soft sensor model, it was applied to a real debutanizer column. The flow diagram of actual debutanizer column unit is given in Figure 5. In the refining industry, the main function of the process is to separate butane from natural gas. At first, the entering liquid is heated into hot steam and then sent into the main tower (T102). The hot vapour condenses into liquid and is separated into a set of fractions with different boiling points. Butane and propane

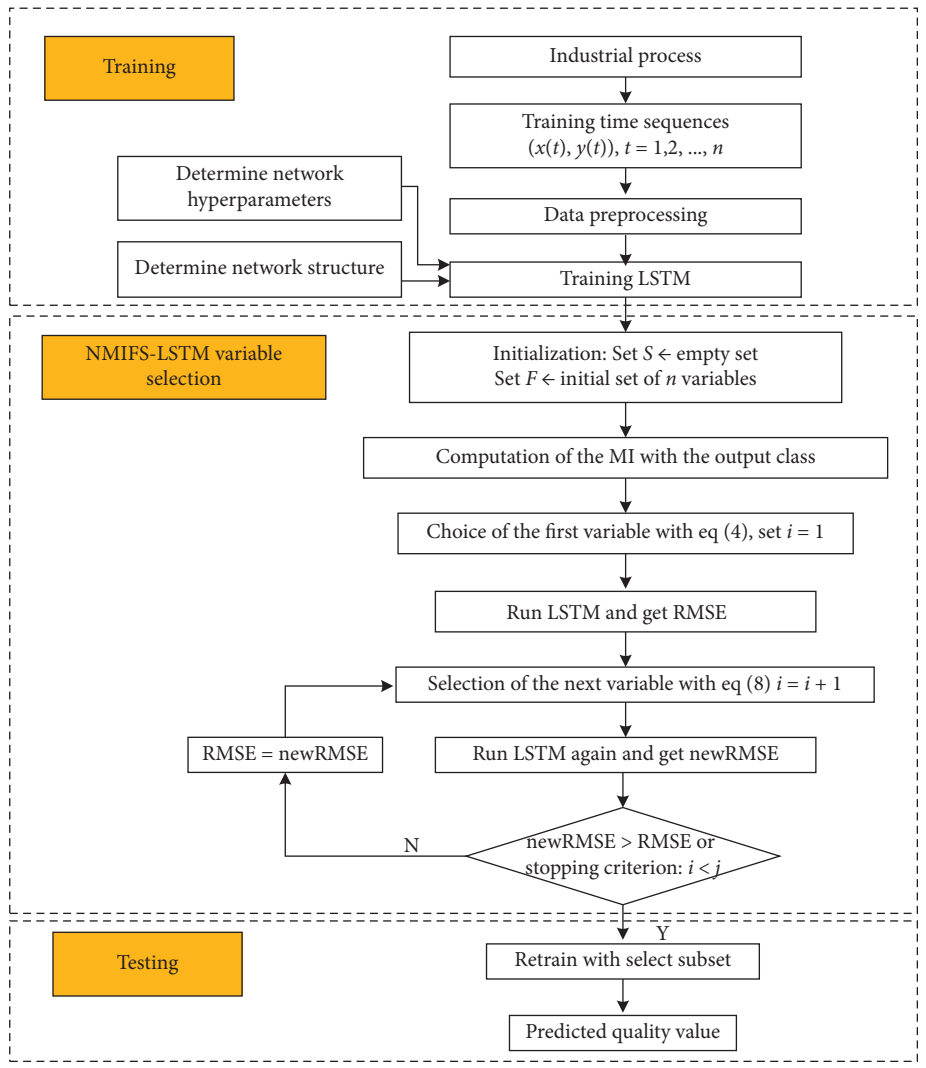

Figure 4: Flowchart of the proposed NMIFS-LSTM-based soft sensor model.

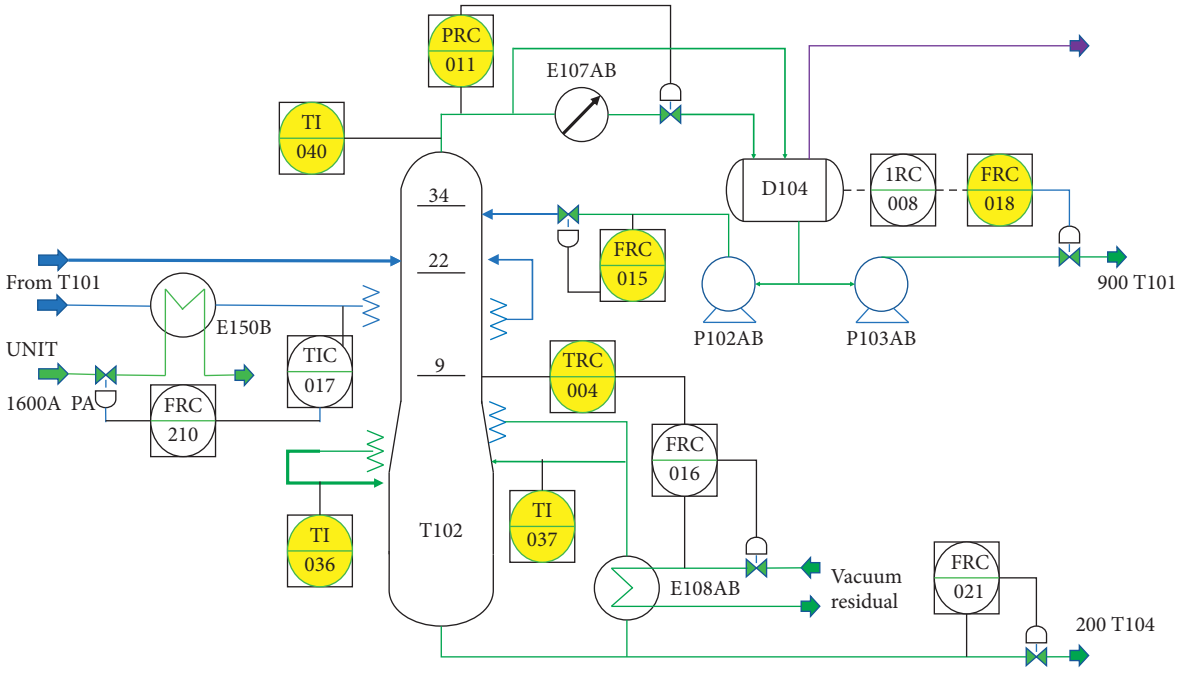

Figure 5: Brief diagram of debutanizer column.

Table 1: Candidate input variables of debutanizer column.

Input Variable	ID	Description
x_1	TI-040	Top temperature
x_2	PRC-011	Top pressure
x_3	FRC-015	Reflux flow
x_4	FRC-018	Flow to next system
x_5	TRC-004	6th tray temperature
x_6	TI-036	Bottom temperature 1
x_7	TI-037	Bottom temperature 2

Table 2: Statistical results for butane content prediction.

	MLP	NMI-MLP	LSTM	NMI-LSTM
MS	40	12	40	8
PMSE	0.0720	0.0673	0.0683	0.0421
R^2	0.9379	0.9524	0.9502	0.9801

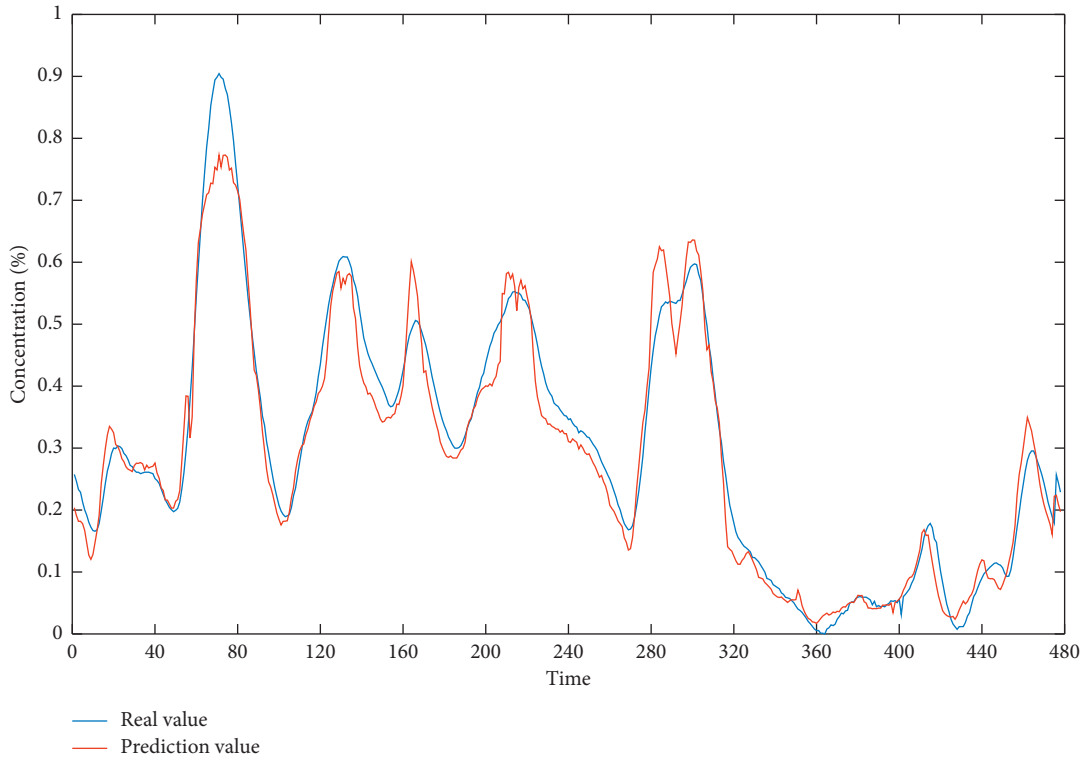

Figure 6: Measured and predicted content of butane.

are detached in the column after the treatment under normal circumstances, which makes the natural gas almost pure methane.

In this case, the content of butane is very important to ensure the product quality during the process. However, this variable is very hard to measure in real time. Hence, a compatible online soft sensing model was proposed to forecast the content. Seven practical sensors were installed in the process, marked as yellow circles in the brief diagram [35], as displayed in Figure 5. All of these candidate variables are listed in Table 1.

2394 data samples were presented at intervals of 15 minutes. The dataset was separated into two parts: the dataset of the first 80% applied for training and the others for testing. On the basis of plant experts guidance [22], the time delay of the process was probably 20–60 minutes. Based on this advice, we extended the input variables to $x = \{x_1(t), ..., x_p(t), x_1(t-1), ..., x_p(t-1), \ldots, x_1(t-4), ..., x_p(t-4)\}$, in which $x_i(t-j)$ means the value of x_i at time $t-j$. In addition, we added a guided value of $(y\text{–}t)$ in each group to enhance the accuracy of modelling. The number of candidate variables was raised from 7 to 40, which resulted in additional complexity of the process.

Table 2 presents the experimental results with these four algorithms. The table shows that the NMI-LSTM algorithm has obvious advantage over others in model

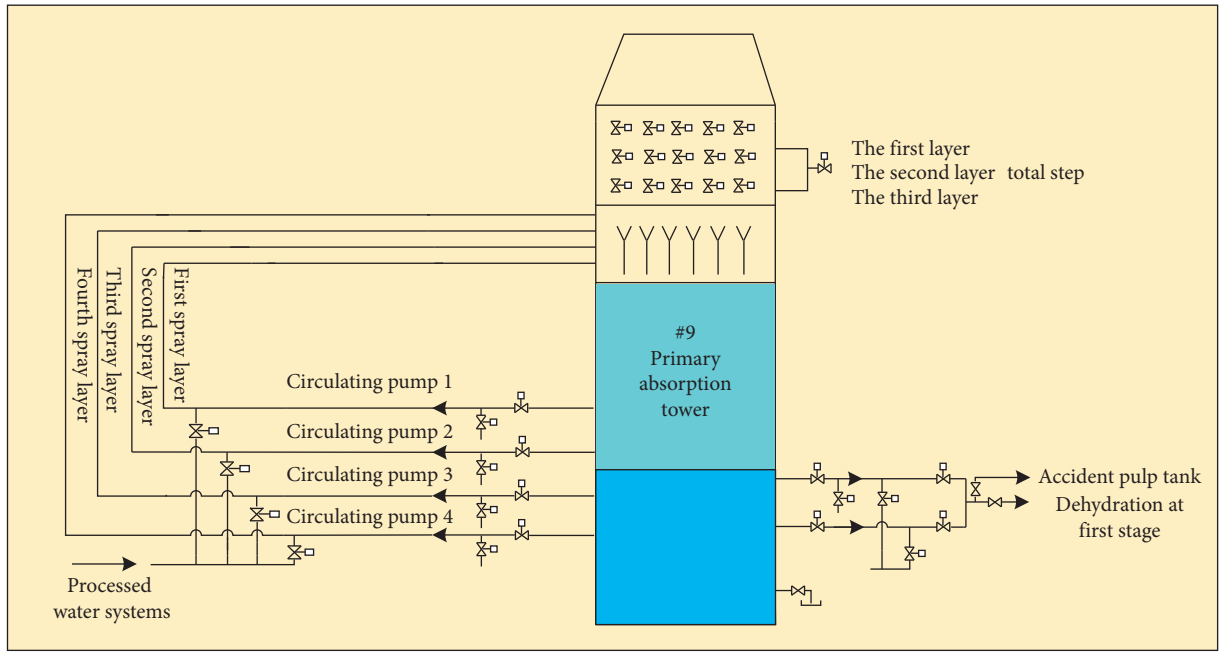

Figure 7: Process flow diagram of primary absorption tower.

Table 3: Candidate input variables of flue gas desulfurization system.

Variable	Description	Unit
1	Generator power	kW
2	#9 furnace original flue gas SO_2	mg/m^3
3	#9-1 absorption tower export flue gas SO_2	mg/m^3
4	#9 absorption tower gypsum slurry pH	
5	Limestone slurry to #9 absorption tower stream	m^3/h
6	#9-2 absorption tower pH 1	
7	#9-2 absorption tower feed flow	m^3/h
8	#9-1 circulating slurry pump current	A
9	#9-3 circulating slurry pump current	A
10	#9 absorption tower entrance flue gas temperature	˚C
11	The sum of air flow	CMH
12	The sum of coal	t
13	#9 absorption tower export flue gas temperature 1	˚C
14	#9 absorption tower export flue gas temperature 2	˚C
15	#9 absorption tower export flue gas temperature 3	˚C
16	#9 furnace chimney entrance net flue gas flow	Nm^3/h
17	#9 furnace chimney entrance net flue gas pressure	Pa
18	#9 furnace raw flue gas NOx	mg/m^3
19	#9 furnace raw flue gas O_2	mg/m^3
20	Median value of desulfurization productivity of unit 9	%
21	#9-2 absorption tower export temperature 1	˚C
22	#9-2 absorption tower export pressure	Pa
23	#9-1 absorption tower export flue gas O_2	mg/m^3
24	#9 absorption tower level value	m
25	#9-2 absorber tower level	m
26	Water circulation vacuum pump	kW
27	Gypsum slurry outflow pump a current	A
28	#9 gypsum slurry cyclone entrance pressure	Pa
29	Dewatering machine power	kW
30	Gypsum filter cake height	m

accuracy. The simulation performance presents that NMI-LSTM shows a tighter and higher accurate model than other methods.

Figure 6 shows the real values and prediction values of target variable by applying the NMIFS-LSTM model. The fitted graph clearly illustrates that our approach can follow

Table 4: Comparison of prediction results of flue gas SO_2 with different algorithms.

	MLP	NMI-MLP	LSTM	NMI-LSTM
MS	30	11	30	8
PMSE	5.9569	3.3295	3.2419	2.1233
R^2	0.3392	0.7765	0.7524	0.9023

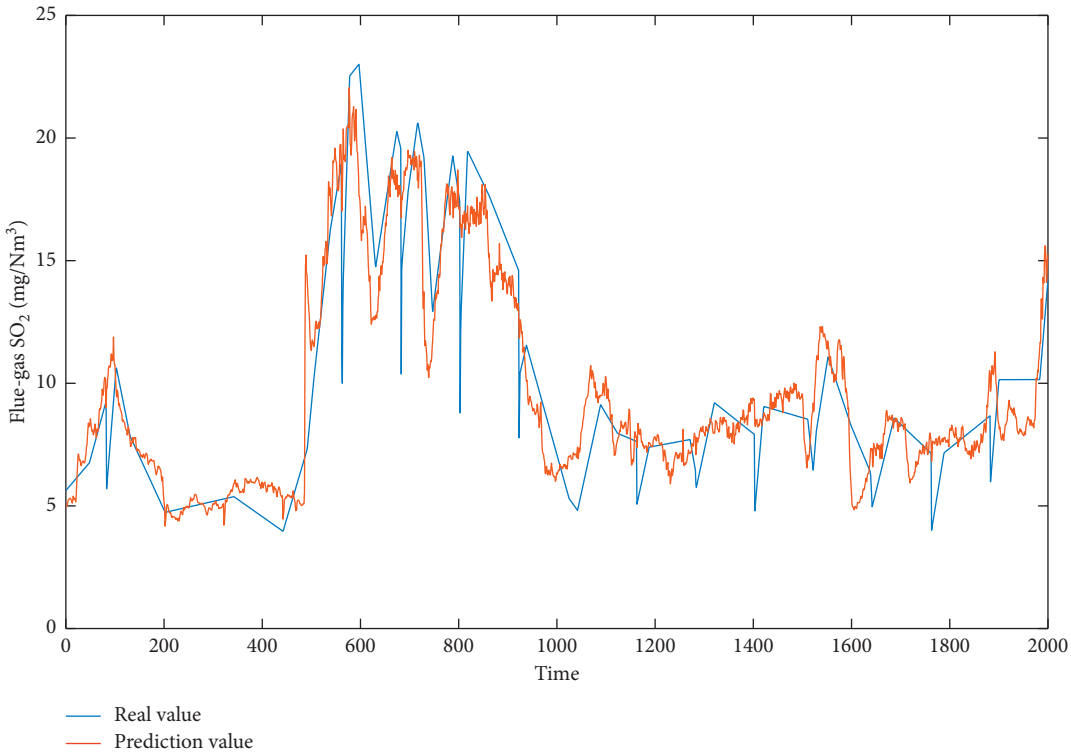

Figure 8: Prediction curve of flue gas SO_2 concentration by NMI-LSTM algorithm.

the variations of the butane content successfully, which further verifies its efficacy.

4.2. Application to Power Plant Desulfurization Technology. The flue gas desulfurization system and industrial process parameters are basically collected from unit 9 of a thermal power plant, which achieves limestone-gypsum wet flue gas desulfurization technology with twin towers. The system's SO_2 is absorbed by lime or limestone with chemical reaction. Compared to the single tower, these twin towers can carry out secondary reaction of transmitted flue gas and eliminate SO_2 in the flue gas more successfully. The flue gas desulfurization process includes SO_2 absorption system, flue gas system, mist eliminator system, absorption tower overflow device, slurry mixing system of absorption tower, oxidizing blower, etc. These twin towers have an absorption area of 12 meters in diameter and a height of 32.6 meters. The flue gas containing SO_2 moves from bottom to top where the bottom of the primary absorption tower (PAT) and encounters a liquid suspension from the spray layer. SO_2 chemically reacts with the alkaline suspension through the gas film and the liquid film in a molecular diffusion manner. The PAT includes four spray layers that are dominated by circulating pumps, shown in Figure 7.

This paper collects the data sample of desulfurization index parameters of unit 9 of a thermal power plant as the research object. The dataset includes 30 input variables and a target output variable flue gas SO_2 concentration. All candidate variables are given in Table 3. The time span is from July 1, 2019, to July 7, 2019, with a time interval of 1 min and a total of 10000 samples. The first 8000 samples are used as the training data and the others are used as the testing data. In the practical simulation experiment, redundant variables in the pool of candidate variables can lead to unsuitable modelling. Consequently, IVFS technique is very important for building a suitable and stable soft measurement model.

Table 4 presents the statistics of data-driven models with different algorithms. Experimental results present that NMI-LSTM has better performance with fewer input variables than other approaches. R^2 of NMI-LSTM is higher than 90%, representing that the proposed soft sensor can precisely forecast the actual values.

Figure 8 shows the prediction curve of SO_2 concentration by NMI-LSTM algorithm. Obviously, NMI-LSTM can

track the dynamic change of target variable effectively, which shows that our algorithm is very effective.

5. Conclusion

In this paper, a novel soft sensor was designed to model complex and dynamic industrial processes with time series characteristics. The LSTM network is trained by datasets taken from actual processes, and NMI is applied to select the variables related to the target variable. The proposed algorithm deletes one irrelevant variable at every step until all the variables are removed. After that, the path of variable selection appears and the algorithm takes the segment with the lowest prediction error. The proposed soft sensor was applied to two practical industrial processes. The simulation and comparison with other algorithms demonstrate the effectiveness and excellence of our approach. The developed soft sensor provides an additional and reliable monitoring tool for pivotal variables and can be further applied to the design of model predictive control systems.

The proposed soft sensor algorithm is easy to implement, and the related program can be preserved as a subroutine in the industrial computer of the DCS. By calling the subroutine, the soft sensor could be periodically retrained and updated with the new production data. The disadvantage of model degradation can be completely eliminated with this technique.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Dongfeng Li and Zhirui Li contributed equally to this article.

Acknowledgments

The authors acknowledge the Yau Mathematical Sciences Center and Huangtai Power Plant for data gathering and experiment. This work was partially supported by the Major Science and Technology Innovation Projects of Shandong Province (Grant no. 2019JZZY010731), the Key Research and Development Program of Shandong Province (Grant no. 2019GGX104037), the S.-T. Yau High School Science Award (Computer), and the National Natural Science Foundation of China (Grant no. 51874300).

References

[1] Z. Ujevic, I. Mohler, and N. Bolf, "Soft sensors for splitter product property estimation in CDU," *Chemical Engineering Communications*, vol. 198, pp. 1566–1578, 2011.

[2] K. Sun, J. Liu, J.-L. Kang, S.-S. Jang, D. S.-H. Wong, and D.-S. Chen, "Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote," *Journal of Process Control*, vol. 24, pp. 1068–1075, 2014.

[3] S. Gao, J. Yang, and J. Wang, "D-FNN based modeling and BP neural network decoupling control of PVC stripping process," *Mathematical Problems in Engineering*, vol. 2014, Article ID 681259, 13 pages, 2014.

[4] A. Nawaz, A. S. Arora, C. M. Yun, H. Cho, S. You, and M. Lee, "Data authorization and forecasting by a proactive soft sensing tool-anammox based process," *Industrial & Engineering Chemistry Research*, vol. 58, no. 22, pp. 9552–9563, 2019.

[5] S. Zhang, X. Chen, and Y. Yin, "An ELM based online soft sensing approach for alumina concentration detection," *Mathematical Problems in Engineering*, vol. 2015, Article ID 268132, 8 pages, 2015.

[6] K. Sun, X. Wu, J. Xue, and F. Ma, "Development of a new multi-layer perceptron based soft sensor for SO_2 emissions in power plant," *Journal of Process Control*, vol. 84, pp. 182–191, 2019.

[7] E. Heidari, M. A. Sobati, and S. Movahedirad, "Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN)," *Chemometrics and Intelligent Laboratory Systems*, vol. 155, pp. 73–85, 2016.

[8] K. G. Sheela and S. N. Deepa, "Neural network based hybrid computing model for wind speed prediction," *Neurocomputing*, vol. 122, pp. 425–429, 2013.

[9] Y. He, Y. Xu, Z. Geng, and Q. Zhu, "Soft sensor of chemical processes with large numbers of input parameters using auto-associative hierarchical neural network," *Chinese Journal of Chemical Engineering*, vol. 23, no. 1, pp. 138–145, 2015.

[10] M. Zabadaj, K. Chreptowicz, J. Mierzejewska, and P. Ciosek, "Two-dimensional fluorescence as soft sensor in the monitoring of biotransformation performed by yeast," *Biotechnology Progress*, vol. 33, no. 2, pp. 299–307, 2017.

[11] J. Rehrla, A. P. Karttunenb, N. Nicolaïc, T. Hörmanne, and M. Hornd, "Control of three different continuous pharmaceutical manufacturing processes: use of soft sensors," *International Journal of Pharmaceutics*, vol. 543, no. 1-2, pp. 60–72, 2018.

[12] Z. Ge, "Supervised latent factor analysis for process data regression modeling and soft sensor application," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 1004–1011, 2015.

[13] G. Hussain, M. Jabbar, J.-D. Cho, and S. Bae, "Indoor positioning system: a new approach based on LSTM and two stage activity classification," *Electronics*, vol. 8, no. 4, p. 375, 2019.

[14] L. Yang, Y. Li, J. Wang, and Z. Tang, "Post text processing of Chinese speech recognition based on bidirectional LSTM networks and CRF," *Electronics*, vol. 8, no. 11, p. 1248, 2019.

[15] J. Chen, Q. Jin, and J. Chao, "Design of deep belief networks for short-term prediction of drought index using data in the Huaihe river basin," *Mathematical Problems in Engineering*, vol. 2012, Article ID 235929, 16 pages, 2012.

[16] S. Xingjian, Z. Chen, H. Wang et al., "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Proceedings of Advances in Neural Information Processing Systems*, Montreal, Canada, pp. 802–810.

[17] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[18] J. Liu, A. Shahroudy, D. Xu et al., "Spatio-temporal lstm with trust gates for 3d human action recognition," in *Proceedings of*

European Conference on Computer Vision, Amsterdam, Netherlands, pp. 816–833.

[19] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3168–3176, 2019.

[20] Q. Sun and Z. Ge, "Probabilistic sequential network for deep learning of complex process data and soft sensor application," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 2700–2709, 2018.

[21] S. Greenland, "Modeling and variable selection in epidemiologic analysis," *American Journal of Public Health*, vol. 79, no. 3, pp. 340–349, 1989.

[22] Z. Hui and T. A. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society*, vol. 67, p. 768, 2005.

[23] J. Wang, J. Zhang, and X. Wang, "A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system," *IEEE Transactions on Semiconductor Manufacturing*, vol. 31, no. 1, pp. 173–182, 2018.

[24] K. Sun, P. Tian, H. Qi, F. Ma, and G. Yang, "An improved normalized mutual information variable selection algorithm for neural network-based soft sensors," *Sensors*, vol. 19, no. 24, p. 5368, 2019.

[25] P. Hanchuan, L. Fuhui, and D. Chris, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[26] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.

[27] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2007.

[28] M. Bennasar, Y. Hicks, and R. Setchi, "Feature selection using joint mutual information maximisation," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8520–8532, 2015.

[29] J. Novovicová, P. Somol, M. Haindl, and P. Pudil, "Conditional mutual information based feature selection for classification task," in *Proceedings of the Iberoamerican Congress on Pattern Recognition*, pp. 417–426, Valparaíso, Chile, November 2007.

[30] A. Dame and E. Marchand, "Mutual information-based visual servoing," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 958–969, 2011.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[32] Z. Yu, G. Chen, Y. Dong et al., "Highway long short-term memory RNNS for distant speech recognition," in *Proceedings of IEEE International Conference on Acoustics*, Shanghai, China, March 2016.

[33] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.

[34] N. Kwak and C.-H. Chong-Ho Choi, "Input feature selection for classification problems," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 143–159, 2002.

[35] L. Fortuna, S. Graziani, and M. G. Xibilia, "Soft sensors for product quality monitoring in debutanizer distillation columns," *Control Engineering Practice*, vol. 13, no. 4, pp. 499–508, 2005.