Chinese Journal of Chemical Engineering 55 (2023) 266–276

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Full Length Article

A dynamic-inner LSTM prediction method for key alarm variables forecasting in chemical process

Yiming Bai [1], Shuaiyu Xiang [1], Feifan Cheng [3], Jinsong Zhao [1,2,*]

[1] Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
[2] Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China
[3] Sinopec Engineering Incorporation, Beijing 100084, China

ARTICLE INFO

Article history:
Received 19 December 2021
Received in revised form 11 August 2022
Accepted 12 August 2022
Available online 28 October 2022

Keywords:
Fault prognosis
Process systems
Safety
Prediction
Principal component analysis
Long short term memory

ABSTRACT

With the increase in the complexity of industrial system, simply detecting and diagnosing a fault may be insufficient in some cases, and prognosing the fault ahead of time could have a certain necessity. Accurate prediction of key alarm variables in chemical process can indicate the possible change to reduce the probability of abnormal conditions. According to the characteristics of chemical process data, this work proposed a key alarm variables prediction model in chemical process based on dynamic-inner principal component analysis (DiPCA) and long short-term memory (LSTM). DiPCA is used to extract the most dynamic components for prediction. While LSTM is used to learn the relationship and predict the key alarm variables. This work used a simulation data set and a real hydrogenation process data set for applications and explained the model validity from the essential characteristics. Comparison of results with different models shows that our model has better prediction accuracy and performance, which can provide the basis for fault prognosis and health management.

1. Introduction

Due to the dangers of chemical materials and operation processes, the chemical industry has many hidden safety hazards [1]. To alleviate the severity of chemical accidents or even prevent the accidents, chemical processes usually have layers of protective measures [2]. Advances in process safety management have highlighted the importance of intelligent process monitoring [3,4]. A number of important advances have been made in the field of fault detection and fault diagnosis (FDD) in the midst of various challenges [5–9]. Simply detecting and diagnosing a fault may be insufficient in some cases, and prognosing the fault ahead of time could have a certain necessity [10,11]. Even when there is an automated control system, a human operator may need to handle possible abnormal conditions if the controller is unable to resist a disturbance effect [12].

In chemical plants, distributed control system (DCS) could provide real-time monitoring of process variables and databases of process variables. In process monitoring, DCS would not focus on the all process variables at the same time. Process variables, which are located at key logical positions in the process and are frequently alarmed, would be selected as key alarm variables. DCS could monitor process variables and perform subsequent FDD based on their changes. Moreover, the warning system is a guidance tool for the operator, and it would greatly reduce the probability of abnormal conditions by achieving the key alarm variables prediction [13].

Current fault prognosis methods can be generally classified as three types: model-based methods, knowledge-based methods and data-based methods [14]. The model-based methods are approaches that build accurate mathematical description of systems using physics or first-principle, which tend to be better than other models when sufficient knowledge are available. However, it is generally unavailable to exact mathematical knowledge as industrial systems become more and more complex. The knowledge-based methods are based on engineering experience and historical events, which could be efficient when the process knowledge has been already accumulated. But the prediction accuracy and application scope are largely limited due to the combinational explosion problem. And the data-based methods determine the status of the system within a certain period of time by analyzing the previously observed data, which is the concentration of this work.

In the early stage, process monitoring and warming were performed on account of the statistical trends measured variables.

* Corresponding author.
E-mail address: jinsongzhao@tsinghua.edu.cn (J. Zhao).

https://doi.org/10.1016/j.cjche.2022.08.024
1004-9541/

The system would issue the warning when the monitoring variables were out of the control limits [12]. Then multivariate statistics such as principal component analysis (PCA) and partial least square (PLS), were widely used since a multivariate monitoring procedure was developed to efficiently monitor the performance of large processes [15]. Afterwards predictive monitoring for abnormal situation was proposed which used predictions from a dynamic model to forecast whether process variables would violate the limit in the future [16]. The accuracy of chemical process data prediction could affect the effectiveness of predicting monitoring system.

Chemical process data is essentially a set of multi-dimensional time series, and the research of time series prediction has gone through a long time of development [17]. Box-Jenkins model provides a comprehensive time-series data forecasting method, which focuses on discreet and linear stochastic processes [18]. Autoregressive (AR), moving average (MA), autoregressive-moving average (ARMA), autoregressive integrated moving average (ARIMA) were proposed based on Box-Jenkins model [19].

The approximation of linear models to complex real-world problems is not always satisfactory, because traditional linear models cannot describe the characteristics of nonlinear problems. In order to predict complex data like chemical process data more accurately, non-linear prediction methods are proposed, such as artificial neural networks (ANNs), autoencoder (AE), support vector machine (SVM), and radial basis function (RBF) network [20].

The above nonlinear models predicted process data better, but they ignored the time sequence nature of process data. Based on ANN, recurrent neural network (RNN) was designed to seize the time-series nature of process data, and to be used to forecast more accurately [21]. The training process of RNN had the natural characteristics of gradient disappearance and gradient explosion. In order to solve the problem, Hochreiter and Schmidhube proposed a long short term memory (LSTM) structure [22]. Based on LSTM, a series of models were well applied in chemical process predictions [16–25].

High-dimensional chemical process data may bring more features and more complex models, which might lead to model overfitting and poor extensiveness. Dimensionality reduction probably abate the impact of high-dimensional data by extracting important features [26]. Ji *et al.* [27] predicted the remaining useful life of airplane engine based on PCA-BLSTM(bi-directional long short-term memory). The principal components extracted by PCA were fed in BLSTM and the hybrid model manifested better prediction accuracy and performance. A major problem of PCA is that it does not pay attention to time series and autocorrelation of the data [28]. In order to seize the autocorrelation of time series, Ku *et al.* [29] proposed dynamic PCA method and Dong *et al.* [30,31] raised a novel dynamic PCA (DiPCA) inspired by dynamic internal local least squares.

A data forecast model of key alarm variables in chemical process based on DiPCA-LSTM is proposed in this paper. The model grasps the dynamic characteristics of chemical process data to obtain preferable prediction. And this paper verifies the effectiveness of the prediction model through actual process data of a chemical pre-hydrogenation process. The model has the following advantages compared to existing researches: (1) Our model uses LSTM structure to focus on longer-term dependencies and performs better in data predictions; (2) Our model extracts the most predictive variables, reducing the complexity of the original forecast problem; (3) The effectiveness of our model can be explained intuitively.

The remainder of this paper is organized as follows. Section 2 reviews the basic knowledge of the dimensionality reduction method and neural networks. Section 3 includes the diagram of the proposed process data forecast method. The results of the case study experiments are revealed in Section 4. Section 5 gives a summary and outlook.

2. Model

2.1. Recurrent neural network

Since the back-propagation algorithm was proposed, neural networks have been an important research area in the field of machine learning [32]. With the development of computing hardware and software, neural networks have become more significant in the field.

The neural network is composed of multiple neuron structures. The classification, diagnosis and prediction are completed through the connection and information transmission between the neurons. The structure of neuron consists of input, weight, summation, activation function and output. The structure of neuron is consisted of input, weight, summation, activation function and output. The activation function allows the neuron to perform a nonlinear transformation to deal with the nonlinear problem. The final output of the neuron can be expressed as:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{z}) = \boldsymbol{f}\left(\sum_j \boldsymbol{W}_j \boldsymbol{x}_j\right) = f\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}\right) \tag{1}$$

The most common and basic structure of neural network is fully connected neural network. For time series data, fully connected neural network may not be applied well, because previous data shows indispensable importance when predicting variables of time series data. RNN was processed to process time series data, which has the characteristic that the hidden layers of the neural network are connected. The input of a hidden layer includes the input at this moment and the output of the previous layer. Another characteristic of RNN is that the parameters are shared at different neurons, which also greatly reduces the number of the neural network variables and may effectively reduce the training time [33]. Fig. 1 is a typical RNN structure.

The backpropagation through time (BPTT) algorithm used in the training stage of RNN has the inherent drawback that backpropagated gradients either grow or shrink at each time step and will cause gradient explosion or vanishment eventually [34]. LSTM was proposed to settle the conundrum [22]. Compared with the ordinary RNN structure, the biggest feature of LSTM network is introduction of the 'gate' structure to vary the information. Fig. 2 is a typical LSTM unit [6].

In Fig. 2, σ represents the sigmoid activation function, and tan*h* represents the tanh activation function. *C* is the long short-term memory through each structure. In each LSTM unit, there are a forget gate, an input gate, and an output gate to control the information. The computational formulas are showed as followed.

$$\boldsymbol{O}_f = \sigma\left(\boldsymbol{W}_{\mathrm{f}} \cdot \left[\boldsymbol{x}^t, \boldsymbol{h}^{(t-1)}\right] + \boldsymbol{b}_{\mathrm{f}}\right) \tag{2}$$

$$\boldsymbol{O}_i = \sigma\left(\boldsymbol{W}_{i1} \cdot \left[\boldsymbol{x}^t, \boldsymbol{h}^{(t-1)}\right] + \boldsymbol{b}_{i1}\right) \odot \tanh\left(\boldsymbol{W}_{i2} \cdot \left[\boldsymbol{x}^t, \boldsymbol{h}^{(t-1)}\right] + \boldsymbol{b}_{i2}\right) \tag{3}$$

$$\boldsymbol{C}^{(t)} = \boldsymbol{O}_{\mathrm{f}} \odot \boldsymbol{C}^{(t-1)} + \boldsymbol{O}_i \tag{4}$$

$$\boldsymbol{h}^{(t)} = \boldsymbol{O}_o = \sigma\left(\boldsymbol{W}_o \cdot \left[\boldsymbol{x}^t, \boldsymbol{h}^{(t-1)}\right] + \boldsymbol{b}_{\mathrm{o}}\right) \odot \tanh\left(\boldsymbol{C}^{(t)}\right) \tag{5}$$

2.2. DiPCA

Single data sample $x \in R^m$ represents a data vector with *m* variables. Assuming that each variable has *n* samples, the total data

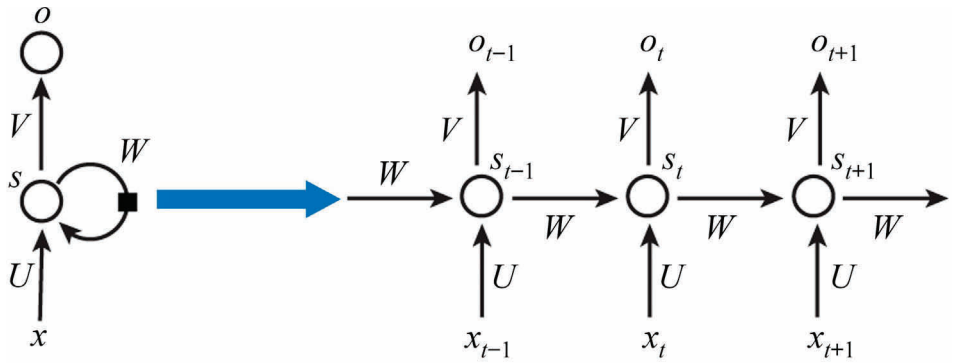

Fig. 1. Recurrent neural network structure.

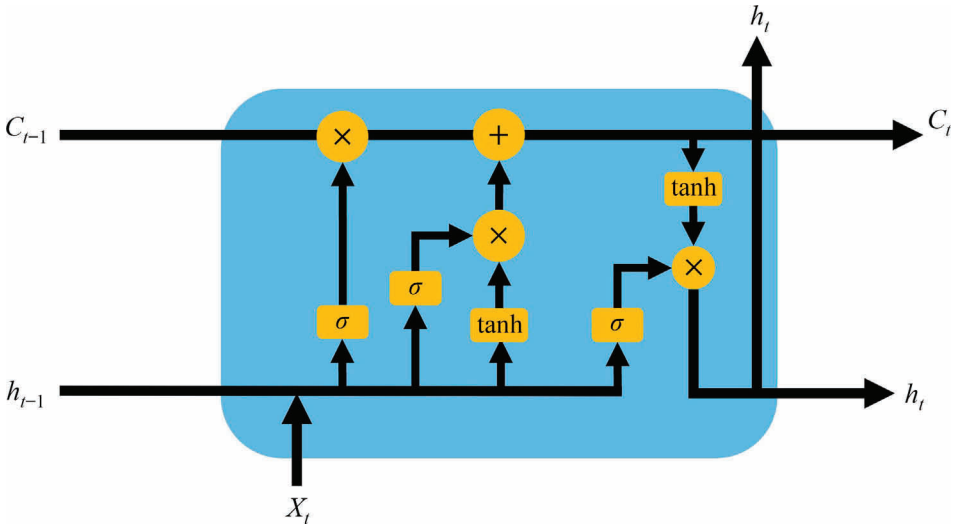

Fig. 2. LSTM unit structure.

matrix is $X \in R^{n\times m}$. The rows represent samples, and the columns represent the collection of a variable over time.

The goal of PCA is to extract the component of the largest variance in the m-dimensional measurement space, $\boldsymbol{p}$ is a loading vector, and the measurement space is projected into the expected subspace [31,35]. The relationship above can be expressed as:

$$\max_{p} \boldsymbol{p}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{p} \text{ s.t.} \|\boldsymbol{p}\| = 1 \tag{6}$$

The above optimization problem can be solved by Lagrange multiplier:

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{p} = \lambda\boldsymbol{p} \text{ where}, \lambda = \boldsymbol{p}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{p} \tag{7}$$

So, $\boldsymbol{p}$ is the eigenvector of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, and λ is the corresponding maximum eigenvalue. The original data matrix can be transformed by the score vector $\boldsymbol{t} = \boldsymbol{X}\boldsymbol{p}$ to extract the next largest eigenvalue and eigenvector:

$$\boldsymbol{X} := \boldsymbol{X} - \boldsymbol{t}\boldsymbol{p}^{\mathrm{T}} \tag{8}$$

The loading vectors $\boldsymbol{p}$ extracted l times are combined to form a loading matrix, and the m-dimensional data matrix coverts to an l-dimensional data matrix.

The goal of DiPCA is to extract the dynamic characteristics of hidden variables, which are also the most predictive variables [31]. The dynamic characteristics of latent variables can be expressed as:

$$\boldsymbol{t}_k = \beta_1\boldsymbol{t}_{k-1} + \beta_2\boldsymbol{t}_{k-2} + \cdots + \beta_s\boldsymbol{t}_{k-s} + \boldsymbol{r}_k \tag{9}$$

The algorithm extracts the latent variable expression as follows, which is a linear combination of the original observed variables:

$$\boldsymbol{t}_k = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{w} \tag{10}$$

$\boldsymbol{x}_k$ is the sample vector at time k, and $\boldsymbol{w}$ is the weight vector. $\boldsymbol{r}_k$ is the prediction residual which can be ignored. The prediction model of latent variables is:

$$\hat{\boldsymbol{t}}_k = \boldsymbol{x}_{k-1}^{\mathrm{T}}\boldsymbol{w}\beta_1 + \boldsymbol{x}_{k-2}^{\mathrm{T}}\boldsymbol{w}\beta_2 + \cdots + \boldsymbol{x}_{k-s}^{\mathrm{T}}\boldsymbol{w}\beta_s = [\boldsymbol{x}_{k-1}^{\mathrm{T}}\cdots\boldsymbol{x}_{k-s}^{\mathrm{T}}](\boldsymbol{\beta}\otimes\boldsymbol{w}) \tag{11}$$

The objective function of DiPCA is:

$$\max_{\beta,w} \frac{1}{N}\sum_{k=s+1}^{s+N} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_k^{\mathrm{T}}[\boldsymbol{x}_{k-1}^{\mathrm{T}}\cdots\boldsymbol{x}_{k-s}^{\mathrm{T}}](\boldsymbol{\beta}\otimes\boldsymbol{w}) \tag{12}$$

For a matrix $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2\cdots\boldsymbol{x}_{N+s}]^{\mathrm{T}}$, matrix can be transformed to:

$$\boldsymbol{X}_i = [\boldsymbol{x}_i\boldsymbol{x}_{i+1}\cdots\boldsymbol{x}_{N+i-1}]^{\mathrm{T}}, i = 1, 2\cdots, s+1 \tag{13}$$

$$\boldsymbol{Z}_s = [\boldsymbol{X}_1\boldsymbol{X}_2\cdots\boldsymbol{X}_s] \tag{14}$$

Through this transformation, the above objective function can be rewritten as:

$$\max_{\beta,w} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}_{s+1}^{\mathrm{T}}\boldsymbol{Z}_s(\boldsymbol{\beta}\otimes\boldsymbol{w}) \quad \text{s.t.} w = 1, \beta = 1 \tag{15}$$

$\boldsymbol{w}$ can be optimized by Lagrange multiplier method. In the process of extracting dynamic features, the following 4 steps are used

to extract the loading vector matrix to implement the DiPCA algorithm, which were shown in Ref. [31]:

(1) The data matrix $\boldsymbol{X}$ is normalized and adjusted to zero mean and unit variance. The loading vector $\boldsymbol{w}$ is initialized to a random unit vector.

(2) Extract latent variables and record the loading vector $\boldsymbol{w}$ at convergence.

(3) After extracting a loading vector, the original data matrix is compressed.

(4) Return the compressed data matrix to step (2) to extract the next latent variable until l latent variables are extracted.

These loading vectors are combined into a loading matrix $\boldsymbol{W}$, and the score matrix $\boldsymbol{X}'$ is a linear transformation of the original variables with lower dimensions, which is the combination of $\boldsymbol{t}_k$ at the time for each input. Then we train the models to learn the changes of the score matrix $\boldsymbol{X}'$. The matrix $\boldsymbol{X}'$ can be calculated as:

$$\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{W} \tag{16}$$

The data matrix obtained by DiPCA has a lower dimension than the original data, but in the data prediction, the experiment needs to restore the low-dimensional data to the original data space to verify the accuracy of the prediction.

3. The Proposed Key Alarm Variables Forecast Method

The proposed key alarm variables prediction method is shown in the Fig. 3. During the data preprocessing stage, the historical process data under normal and abnormal working conditions are collected and preprocessed for training. In the models designing and training stage, LSTM models are designed and trained by low-dimensional training datasets. During the data forecasting stage, testing data is preprocessed by normalization and feature extraction. Then we use the low-dimensional testing datasets to get the key alarm variables predictions. The specific experimental steps are shown in the following description.

Data preprocessing:

Step 1: The data under normal working conditions and abnormal conditions are selected from the historical dataset for model training.

Step 2: Normalization is used for data pre-processing. Then loading vectors are calculated by DiPCA.

Step 3: The feature extraction is realized by the loading vectors. Then a time window is selected to transform the original data into fixed-length multivariate time series matrices.

Models designing and training:

Step 4: LSTM models are designed by architecture optimization.

Step 5: The models are trained by the low-dimensional training datasets.

Data forecasting:

Step 6: Testing data are collected and normalized.

Step 7: The feature extraction is realized by DiPCA. The testing data are transformed into sample matrices in the same time window as the model training step.

Step 8: The trained models get the key alarm variables prediction results using low-dimensional testing datasets.

4. Case Study Experiment

In this section, we use two benchmarks to verify the effectiveness of the proposed key alarm variables prediction model, including a continuous stirred tank heater (CSTH) model and an industrial pre-hydrogenation model. The results of the real data will be compared with RNN, LSTM, PCA-RNN, DiPCA-RNN, PCA-LSTM. These methods will be evaluated for performance on the same dataset we generated. The experiments are implemented on a computing server with Intel Xeon Gold 5118 CPU and NVIDIA TITAN RTX GPU.

4.1. Application on the CSTH system

The continuous stirred tank heater (CSTH) is a typical mixing vessel in chemical processes. In 2008, Thornhill *et al.* [36] developed a MATLAB simulation for a CSTH pilot plant, which is displayed in the Fig. 4. The tank is supplied with one stream of hot water and one stream of cold water, and heated by steam then. The water is well mixed and therefore the temperature of the outlet water is regarded as it in the tank. A total of five variables are monitored in this work: cold water flow, cold water valve, tank level, outlet temperature, and steam valve. All the measurements are electrical signals in the range of 4–20 mA. The closed-loop simulation code can be accessed at https://personal-pages.ps.ic.ac.uk/~nina/CSTHSimulation/index.htm. The Simulink program only provides the structure with real disturbances under normal conditions, while the simulation data under abnormal conditions needs to be generated by users' own modification of the program. The CSTH model has been used as a benchmark in many studies to validate the effectiveness of the method [7,37,38].

In this simple CSTH simulation model, the temperature of the tank is chosen as the key variable for the study, because the temperature would affect states of reactions in the tank largely, which deserves a lot of constant attention as a key variable. In addition, the temperature might have a great impact on the safety of the tank, which would be chosen as an alarm variable in real chemical industry plants. Two kinds of abnormal cases are designed and investigated as shown in Table 1. Fault 1 stands for the step drop in the tank level, and the control loop would make the cold water valve and steam valve increase and recover, which could result in the abnormal changes in the key variable of the temperature in the tank. Fault 2 represents the instantaneous increase in outflow, which would make the level gradually low. And it in turn will affect the key variable of the temperature in the tank.

In order for the proposed method to learn the dynamic characteristics of the model, small fluctuations are introduced in the first 2000 samples when simulating the two abnormal states. The data under two abnormal conditions is divided into a training dataset and a validation dataset, with the training data set consisting of 2000 samples and the validation data set consisting of 400 samples. In both abnormal cases, the fault is introduced after 2100 samples of normal operation. Fig. 5 shows the variation of each variable under various conditions.

The hyper-parameters chosen for the proposed method are described as follows. According to the proposed method, the top 3 important dynamic principal components of the 5 variables are extracted using the method of DiPCA. And 90.5% data variation is explained by the 3 principal components. The following neural network structure is designed and selected, containing two hidden layers of LSTM with 64 nodes and one fully connected layer to obtain the predicted values of the key variable of the temperature in the tank. The target of the model is to predict the changes of the key variable under abnormal conditions (10) samples in advance using previous 30 samples, with the expectation that the key variable could be determined in advance if they exceed the threshold and thus preventive measures could be taken.

Fig. 6 illustrates the performance of the key variable forecast method proposed in this work under two failure conditions. Under Fault 1, the step-down in level rapidly increases the cold water flow because of the control loop, and gradually regulates temperature and level by the stream valve, resulting that the temperature in the tank will have an abnormal drop. Similarly, the temperature in the tank rises abnormally after the introduction of the fault under Fault 2. Under both faults, the method could predict changes

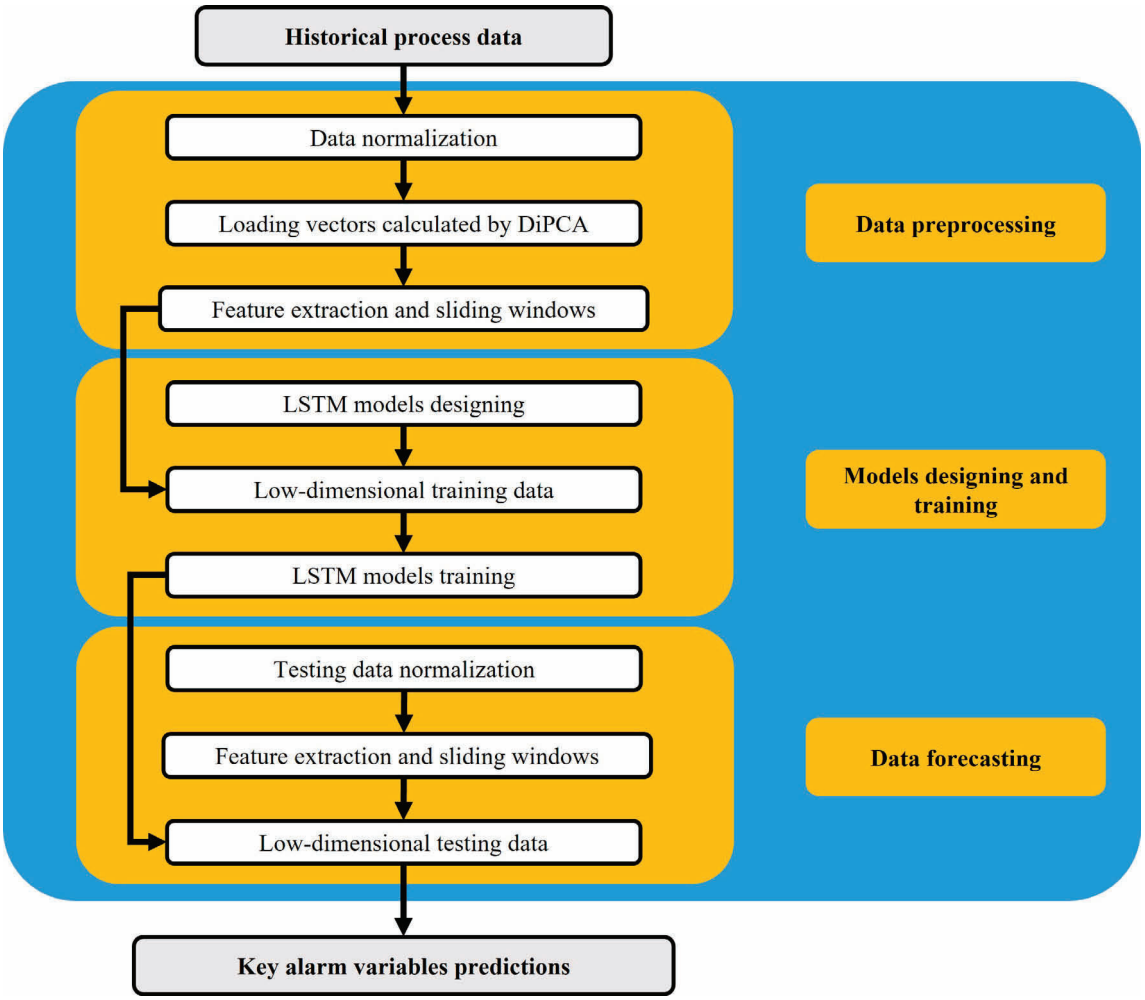

Fig. 3. The flowsheet of the proposed key alarm variables prediction method.

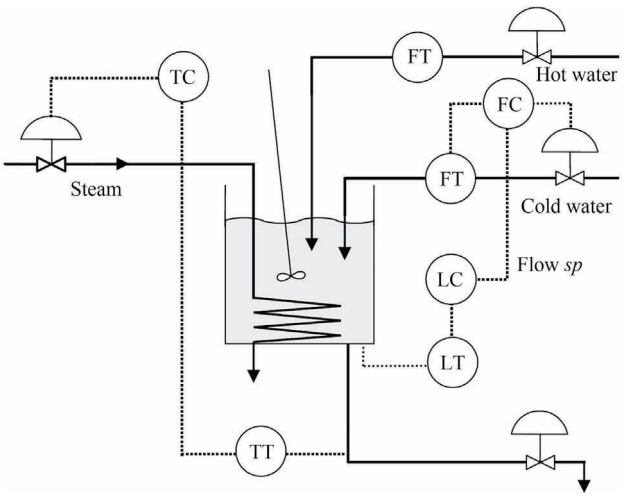

Fig. 4. The continuous stirred tank heater [36].

Table 1
Abnormal cases in the CSTH

Fault number	Description	Type
1	Tank level	Step drop
2	Outflow	Step increase

in key variables accurately 10 samples ahead after the introduction of the fault at the 2100th sample.

4.2. Application on industrial pre-hydrogenation instance

The data collected in a pre-hydrogenation zone of a petrochemical company, which is located in front of the refined oil reforming zone, is used for applications. The propose of the pre-hydrogenation zone is to reduce the content of sulfur in petroleum by the reaction of hydrogen and naphtha for environmental protection.

The system selected in the experiment is the reaction system of the pre-hydrogenation zone. Fig. 7 is a schematic diagram of the overall process framework of it. The main instruments include storage tank V101, combustion furnace F101, reactor R101 and high sub-tank V102. The naphtha and hydrogen stored in V101 are mixed into hydrogen combustion furnace F101 and heated to 279.8 °C into R101. After the layer-by-layer reaction, it leaves the reactor and enters V102 to separate the unreacted hydrogen from the reacted naphtha. The hydrogen returns to the heating furnace again. After the reaction, the naphtha is collected and enters the refining system of the pre-hydrogenation zone.

Variables are selected combined with chemical engineering principles and reaction processes in advance. The relevant variables are shown in Table 2. Among the 11 selected variables, the

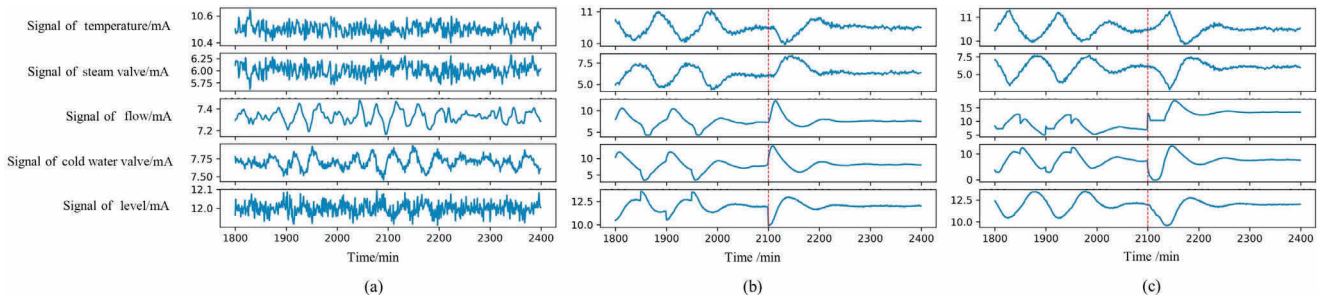

Fig. 5. The time series data of variables of CSTH: (a) under normal condition, (b) under Fault 1, (c) under Fault 2.

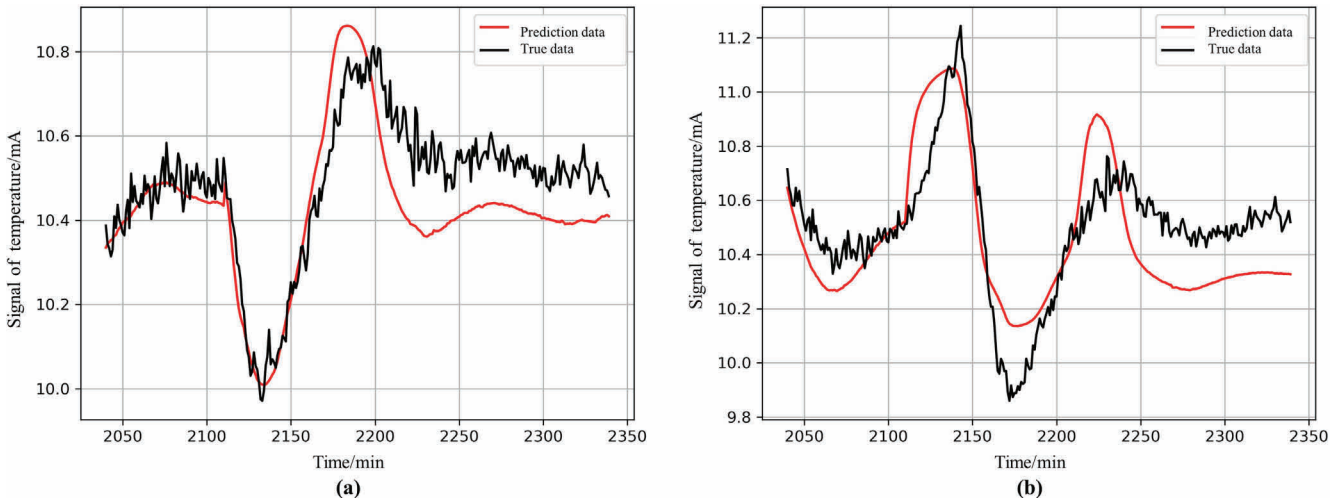

Fig. 6. Key variable forecast results 10 samples ahead and true changes: (a) under Fault 1, (b) under Fault 2.

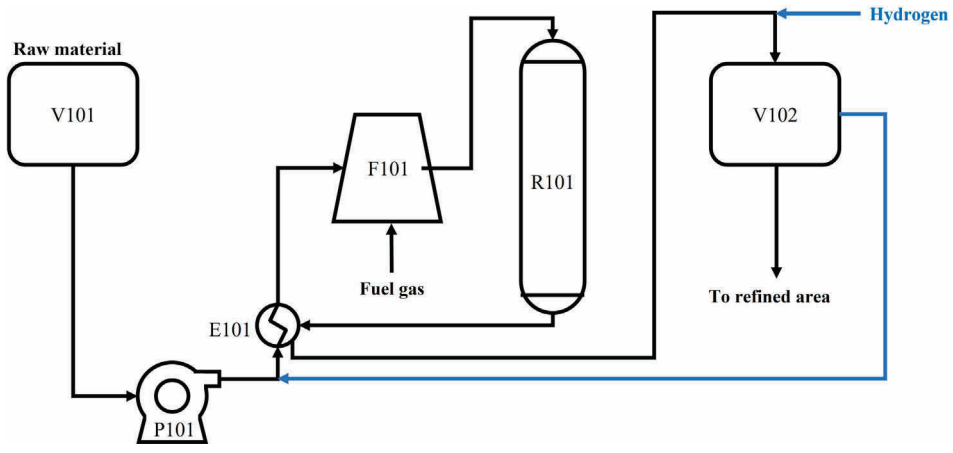

Fig. 7. Pre-hydrogenation reaction zone.

seventh variable F101 fuel gas volume is selected as the key alarm variable.

The values of the above variables are measured every-three minutes, and each variable has more than 10000 measured values. The measurement of the key variable, F101 fuel gas volume, is shown in Fig. 8.

Chemical engineering process could be a complex nonlinear process, which may refer a great deal of variables, physical processes and chemical processes. In order to reduce the impact of noisy data, it is necessary to make use of data smoothing methods before the experiment. Moving average, one of data smoothing

Table 2
Selected variables of pre-hydrogenation reaction zone

Number	Description	Number	Description
1	V101 level	7	F101 fuel gas volume
2	Pre-hydrogenation feed	8	Bed temperature of reactor
3	F101 output temperature	9	Bed temperature of reactor
4	F101 fuel gas pressure	10	High sub-tank pressure
5	F101 radiation temperature	11	High sub-tank level
6	F101 negative pressure		

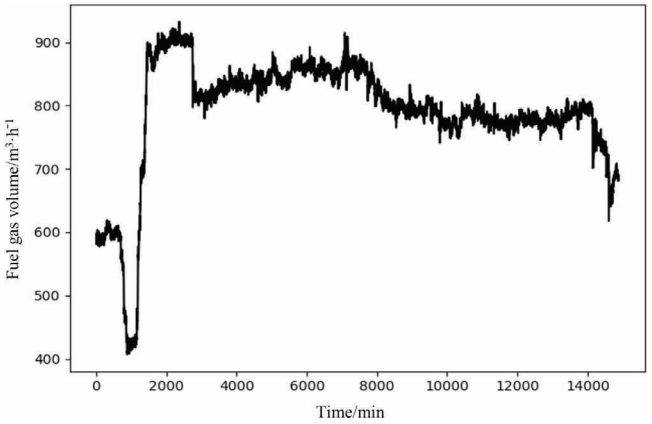

Fig. 8. Measurement data of fuel gas volume.

methods, is applied in the preprocessing of chemical process data, which makes the process data smooth with keeping the original data information to an extreme.

In order to compare the degree of data change and information loss, it is inevitable to use information theory to deal with it. Information entropy expresses the uncertainty of data in the form of value, which is used to measure the amount of information. For a random variable X, $p(x)$ represents the probability that variable X takes the value x. And the information entropy $H(X)$ of random variable X can be expressed as [39]:

$$H(X) = -\int p(x)\lg p(x)\mathrm{d}x \tag{17}$$

Random variable X with large information entropy may have more uncertainty and information to a certain degree. On this basis, in order to measure the relationship between two sets of variables, mutual information entropy is introduced. The value of mobile window is selected as 11 by mutual information.

$$H(X) = -\int p(x)\lg p(x)\mathrm{d}x \tag{18}$$

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{19}$$

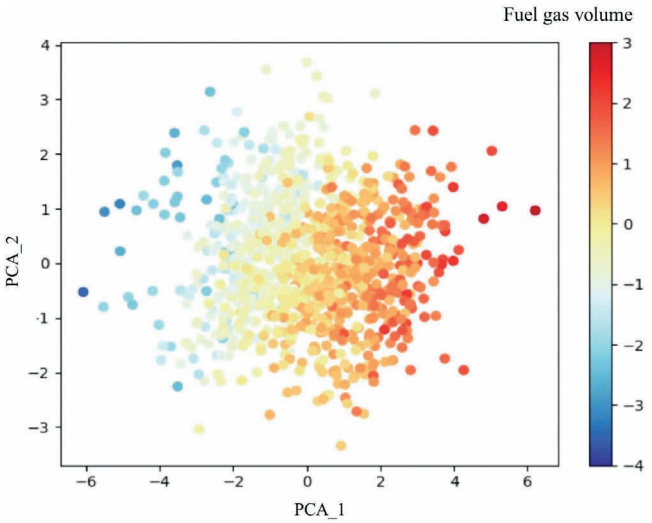

Fig. 9. Plotting PCA of features extracted by DiPCA.

Table 3
LSTM model candidates for variable prediction of the pre-hydrogenation process

Model	Architecture
Model 1	LSTM (128)
Model 2	LSTM (64) – LSTM (32)
Model 3	LSTM (64) – LSTM (16)
Model 4	LSTM (64) – LSTM (8)
Model 5	LSTM (32) – LSTM (16)
Model 6	LSTM (32) – LSTM (8)
Model 7	LSTM (16) – LSTM (8)
Model 8	LSTM (64) – LSTM (16) – LSTM (8)
Model 9	LSTM (32) – LSTM (16) – LSTM (8)

Table 4
Prediction errors of different architectures

Model	Absolute error/$m^3 \cdot h^{-1}$	Relative error/%
Model 1	0.89	0.11
Model 2	2.15	0.26
Model 3	0.84	0.10
Model 4	0.70	0.08
Model 5	0.80	0.10
Model 6	0.71	0.09
Model 7	1.71	0.21
Model 8	3.61	0.44
Model 9	1.03	0.12

4.2.1. Feature validity verification

Plotting PCA is a method which could visually display data in a two-dimensional figure by extracting two dimensions of multi-dimensional data with PCA. Plotting PCA method is used to verify the effectiveness of the features extracted by DiPCA method.

In Fig. 9, each point represents one training principal component, and the color of each point from blue to red represents the value of fuel gas volume. The abscissa of points in Fig. 9 represent the first principal components extracted by PCA method, which describe the principal components straightway. It could be obtained that as the abscissa of points varies from negative to positive, the color changes from blue to red, which means the value of target variable increases. From the above results, there should be a positive correlation between the principal components extracted by DiPCA and the training targets. Such results have initially veri-

Table 5
Structure settings of RNN, LSTM, PCA-RNN, PCA-LSTM, PCA-BiLSTM, DiPCA-PCA, DiPCA-LSTM for key variable forecast of industrial pre-hydrogenation instance

Model	Feature extraction	Structure
RNN	—	RNN: 11 → 64 RNN: 64 → 8; Linear: 8 → 1
LSTM	—	LSTM: 11 → 64; LSTM: 64 → 8; Linear: 8 → 1
PCA-RNN	11 → 5	RNN: 5 → 64; RNN: 64 → 8; Linear: 8 → 1
PCA-LSTM	11 → 5	LSTM: 5 → 64; LSTM: 64 → 8; Linear: 8 → 1
PCA-BiLSTM	11 → 5	BiLSTM: 5 → 64; BiLSTM: 64 → 8; Linear: 8 → 1
DiPCA-RNN	11 → 5	RNN: 5 → 64; RNN: 64 → 8; Linear: 8 → 1
DiPCA-LSTM	11 → 5	LSTM: 5 → 64; LSTM: 64 → 8; Linear: 8 → 1

Table 6
Prediction errors of different models

Model	Absolute error/$m^3 \cdot h^{-1}$	Relative error/%
RNN	1.54	0.19
LSTM	2.05	0.25
PCA-RNN	1.38	0.17
PCA-LSTM	1.75	0.21
PCA-BiLSTM	1.47	0.18
DiPCA-RNN	1.48	0.18
DiPCA-LSTM	0.5662	0.07

fied the effectiveness of DiPCA in chemical process variable forecast.

4.2.2. Network architecture optimization

It is a common issue that there is no scientific guidance for designing an optimal LSTM architecture for a certain problem. In order to find a proper model, several LSTM models are tried in our experiments, the architectures of which are shown in Table 3 [5]. Then an architecture with the most outstanding prediction performance was selected from Table 3. Parameters of the selected architecture were studied through experiments. The parameters mainly contain the number of hidden layers and the numbers of neurons in each layer.

In the following, Model 4 is explained as an example. Here we use two LSTM hidden layers with 64 neurons and 8 neurons. These models were studied through experiments, and each model was trained to predict the key chemical process variable 15 min ahead with principal components data extracted by DiPCA. The prediction errors of each model are summarized in the Tables 4 and 5.

In the light of the Table 4, prediction results could be related to the structure of model, and too simple or too complex model may not obtain satisfactory prediction results. For example, a LSTM model with three hidden layers may be too complicated for the pre-hydrogenation process in the experiments, or more training data and training time may be imperative. Forecasting errors and the trend of prediction results were evaluated comprehensively, and the Model 4 with two hidden layers is selected as the best architecture for the pre-hydrogenation process.

4.2.3. Performance of key variable forecast

In order to verify the effectiveness of the model proposed in the paper, 7 methods are used in the experiments. Two typical structures, simple RNN layers and LSTM layers, are used in models. Original data, principal components extracted by PCA and DiPCA are applied to the models. The hyper-parameters we choose for different methods are described as following Table 5.

The models are trained to predict the key chemical process variable 15 min ahead. The prediction errors of each model are summarized in the Table 6. It could be easily found from the results that the DiPCA-LSTM model has the smallest prediction errors. The key variable prediction results of various methods are displayed in Fig. 10. From the intuitive view of the prediction results, the proposed DiPCA-LSTM model show the best results than other methods.

4.2.4. Time intervals in advance

To verify the performance of the DiPCA-LSTM model at different time advances, the Model 4 with two LSTM layers is chosen in the experiments. The model is trained to predict the key chemical process variable 9, 15 and 21 min ahead respectively. The prediction results of different time intervals in advance are shown in Fig. 11.

From the prediction results, it could be found that the forecasting results of the different three intervals in advance can accurately describe changes in real data to some extent. The prediction errors of the three methods are summarized in the Table 7. It could be found that the model with 9 min in advance has the smallest prediction errors and the prediction errors would increase when the prediction interval ahead enlarging. It would be a challenging problem that how to predict process variable more accurately a long-time interval beforehand.

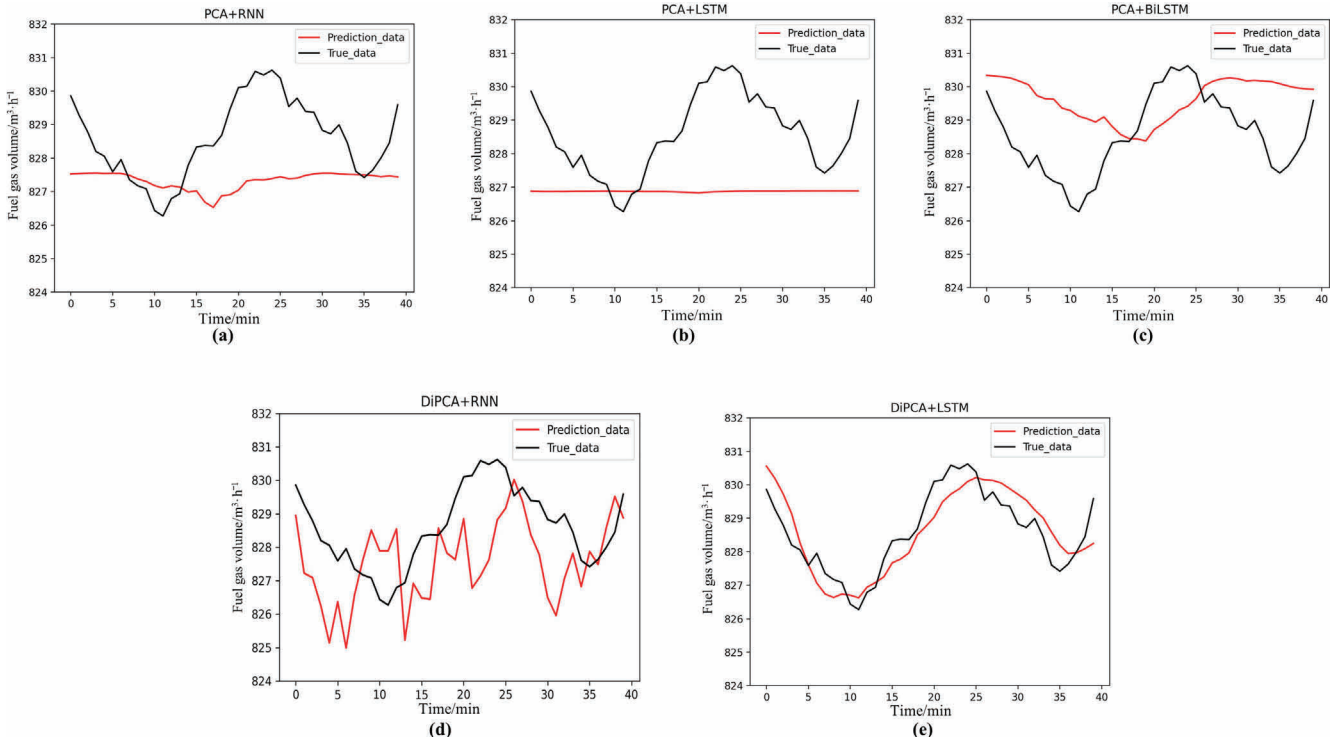

Fig. 10. Key variable prediction results of the model of (a) PCA-RNN, (b) PCA-LSTM, (c) PCA-BiLSTM, (d) DiPCA-RNN, (e) DiPCA-LSTM.

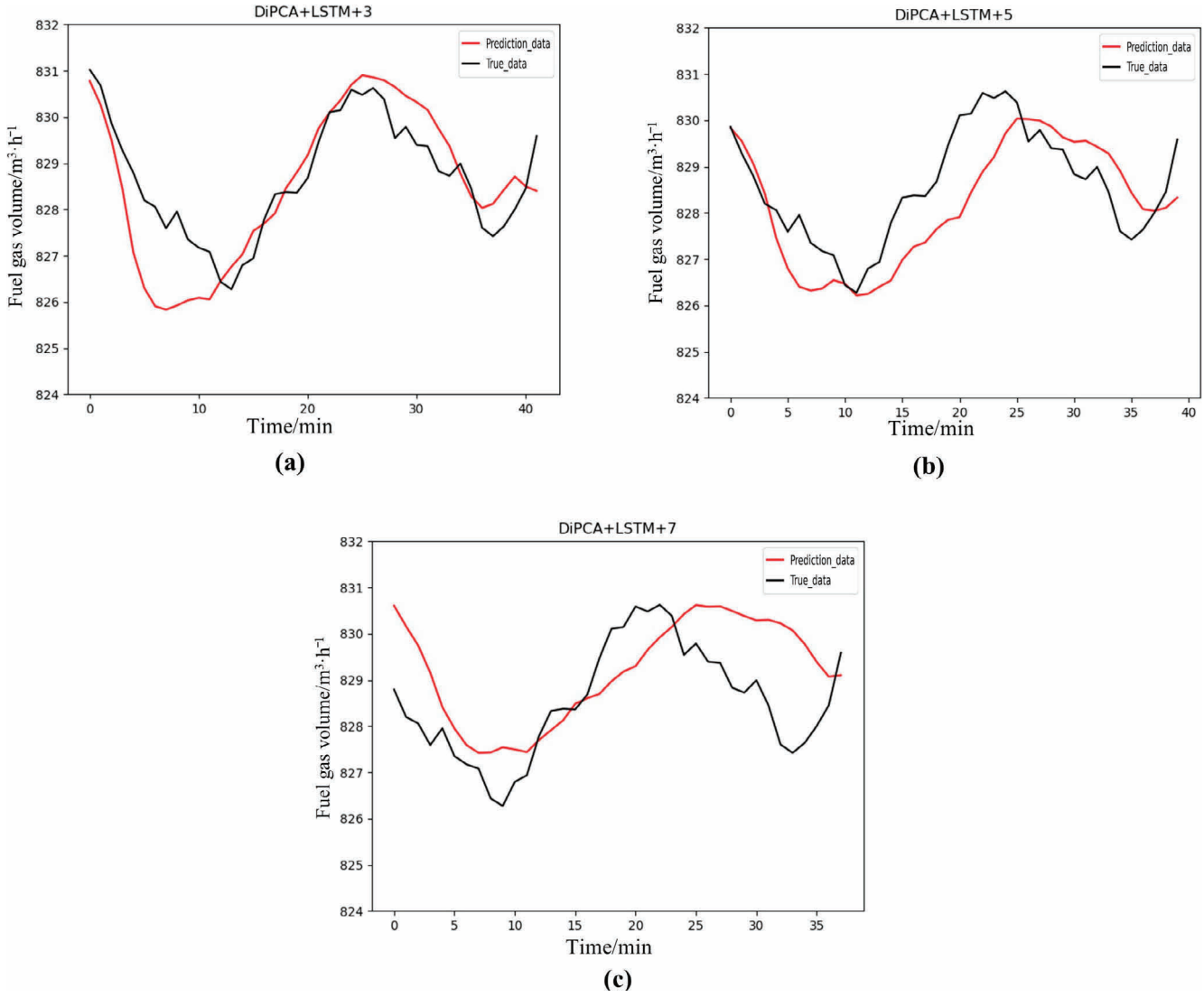

Fig. 11. Key variable prediction results: (a) 9 min ahead, (b) 15 min ahead, (c) 21 min ahead.

Table 7
Prediction errors of different time intervals ahead

Prediction interval/min	Absolute error/$m^3 \cdot h^{-1}$	Relative error/%
9	0.68	0.08
15	0.78	0.09
21	1.04	0.13

4.2.5. Model validity analysis

In order to understand the advantages of the model of DiPCA-LSTM in key process variable prediction, we analyze the effectiveness of the model from the perspective of principal components.

In this work, 5-dimensional principal components were extracted from the samples in the training set by PCA and DiPCA, and the changes are shown in the Figs. 12 and 13.

Fig. 12 show the principal components extracted by DiPCA, and Fig. 13 show them extracted by PCA. The trends and ranges are mainly concerned, which might play an important role in the forecasting process. For principal components extracted by DiPCA, each principal component has a change trend to a certain extent. For principal components extracted by PCA, only the first principal component has an obvious change trend.

DiPCA is designed to extract the principal components with the largest self-correlation in the original data, and PCA is proposed to extract the principal components with the largest variance. It could be found that each principal component extracted by DiPCA may contain certain dynamical information, while only the first principal component extracted by PCA contain most dynamical information. Then it could be explained that the prediction results of PCA fluctuate slightly mainly based on the change of the first principal component. And it also can be comprehended that the prediction of DiPCA may perform well because each principal component contains dynamical information. The effectiveness of the proposed method could be explained by the principal components extracted by DiPCA.

5. Summary and Outlook

In this work, a novel DiPCA-LSTM prediction method is proposed for key alarm chemical process variables prediction. The most predictive principal components of process variables are extracted from high-dimensional data by DiPCA and then computed by LSTM models to get the key alarm chemical process variables prediction results. The key alarm chemical process variables

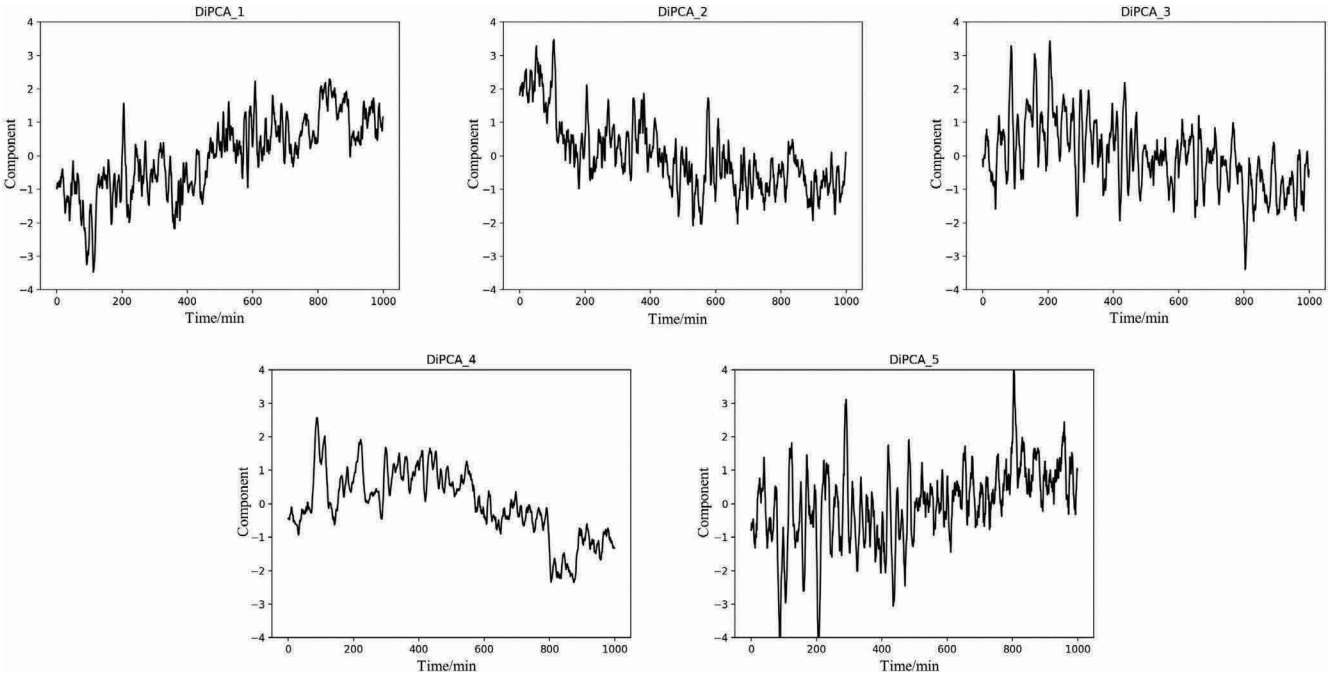

Fig. 12. Principal components extracted by DiPCA.

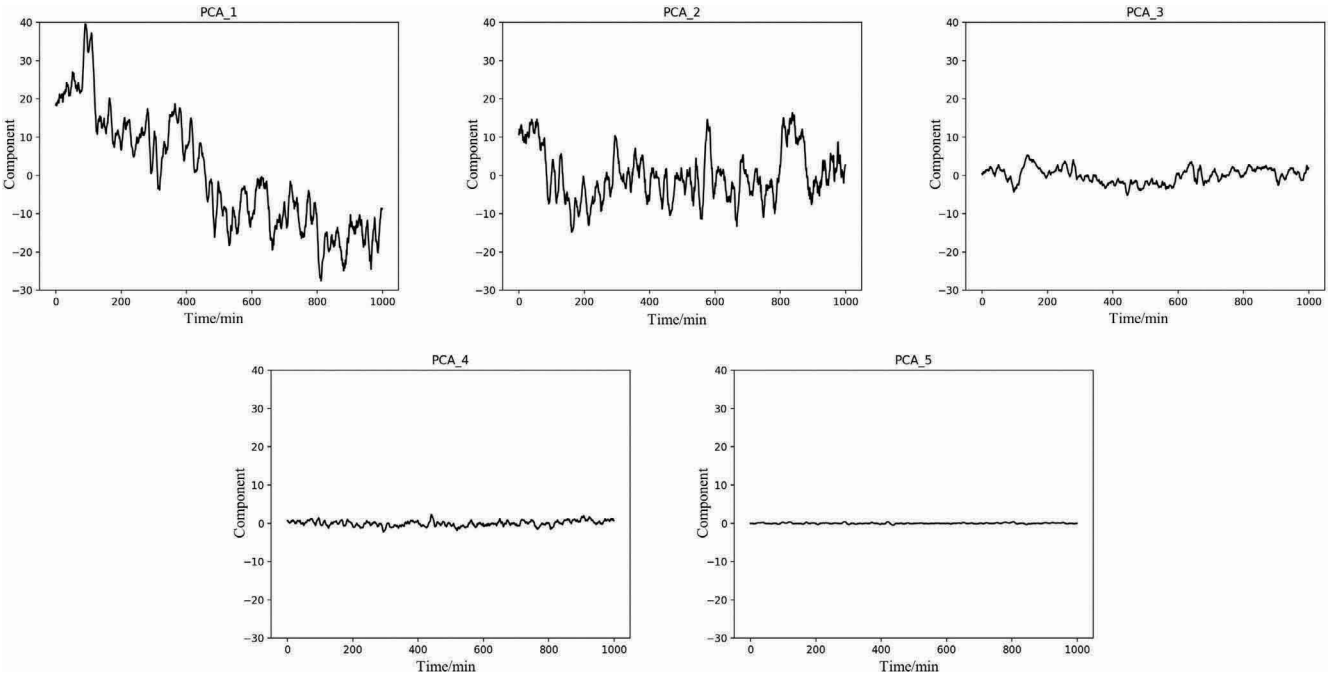

Fig. 13. Principal components extracted by PCA.

prediction model is trained by historical data offline and used for the key variable prediction online.

The proposed method is applied on a simulation CSTH and a pre-hydrogenation zone of a petrochemical company. The trend of fuel gas could be effectively predicted as an example in the pre-hydrogenation model. A LSTM model is selected by network architecture optimization. The MSE is defined as error function to evaluate the prediction results of different models. The relative error of the proposed model for the chemical process variable prediction 15 min in advance is 0.07%. In addition, the changes of principal components extracted by PCA and DiPCA are studied to understand the advantages of DiPCA-LSTM method in key process variables prediction.

This method is dependable as the basis of chemical process variables prediction when confronting chemical process historical data. However, the method does not perform well in a long-term process variables predictions. To overcome this obstacle, more efficient feature extraction techniques and neural network structures will be investigated in the near future. It is worth trying to integrate transformer structure to chemical process variables prediction methods or to use attention to extract efficient feature.

CRediT Authorship Contribution Statement

Yiming Bai: Methodology, Writing – original draft, Software, Validation, Formal analysis, Investigation, Data curation, Visualization. **Shuaiyu Xiang:** Methodology, Investigation. **Feifan Cheng:** Methodology, Investigation. **Jinsong Zhao:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge support from the National Natural Science Foundation of China (21878171).

References

[1] P.A. Carson, C.J. Mumford, An analysis of incidents involving major hazards in the chemical industry, *J. Hazard. Mater.* 3 (2) (1979) 149–165.
[2] D. Crowl, J. Louvar, Chemical Process Safety: Fundamentals with Applications, Pearson Education, 2001.
[3] M. Madakyaru, F. Harrou, Y. Sun, Improved data-based fault detection strategy and application to distillation columns, *Process. Saf. Environ. Prot.* 107 (2017) 22–34.
[4] X.T. Bi, R.S. Qin, D.Y. Wu, S.D. Zheng, J.S. Zhao, One step forward for smart chemical process fault detection and diagnosis, *Comput. Chem. Eng.* 164 (2022) 107884.
[5] H. Wu, J.S. Zhao, Deep convolutional neural network model based chemical process fault diagnosis, *Comput. Chem. Eng.* 115 (2018) 185–197.
[6] S.D. Zheng, J.S. Zhao, A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis, *Comput. Chem. Eng.* 135 (2020) 106755.
[7] X.T. Bi, J.S. Zhao, A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification, *Process. Saf. Environ. Prot.* 156 (2021) 581–597.
[8] H. Wu, J.S. Zhao, Fault detection and diagnosis based on transfer learning for multimode chemical processes, *Comput. Chem. Eng.* 135 (2020) 106731.
[9] D.Y. Wu, J.S. Zhao, Process topology convolutional network model for chemical process fault diagnosis, *Process. Saf. Environ. Prot.* 150 (2021) 93–109.
[10] B. Bhadriraju, J.S.I. Kwon, F. Khan, Risk-based fault prediction of chemical processes using operable adaptive sparse identification of systems (OASIS), *Comput. Chem. Eng.* 152 (2021) 107378.
[11] S.Y. Xiang, Y.M. Bai, J.S. Zhao, Medium-term prediction of key chemical process parameter trend with small data, *Chem. Eng. Sci.* 249 (2022) 117361.
[12] M.A.I. Khan, S. Imtiaz, F. Khan, Predictive warning system for nonlinear process plants, *J. Process. Control* 100 (2021) 1–10.
[13] W.D. Tian, M.G. Hu, C.K. Li, Fault prediction based on dynamic model and grey time series model in chemical processes, *Chin. J. Chem. Eng.* 22 (6) (2014) 643–650.
[14] K. Zhong, M. Han, B. Han, Data-driven based fault prognosis for industrial systems: A concise overview, *IEEE/CAA J. Autom. Sin.* 7 (2) (2020) 330–345.
[15] J.V. Kresta, J.F. Macgregor, T.E. Marlin, Multivariate statistical monitoring of process operating performance, *Can. J. Chem. Eng.* 69 (1) (1991) 35–47.
[16] B.C. Juricek, D.E. Seborg, W.E. Larimore, Predictive monitoring for abnormal situation management, *J. Process. Control* 11 (2) (2001) 111–128.
[17] C. Hamzaçebi, D. Akay, F. Kutay, Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting, *Expert Syst. Appl.* 36 (2) (2009) 3839–3844.
[18] A.E. Pankratz, Forecasting with Univariate Box-Jenkins Models: Concepts and Cases, John Wiley & Sons, 1983.
[19] Ü.Ç. Büyükşahin, Ş. Ertekin, Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition, *Neurocomputing.* 361 (2019) 151–163.
[20] X.F. Yuan, C. Ou, Y.L. Wang, C.H. Yang, W.H. Gui, A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes, *Chem. Eng. Sci.* 217 (2020) 115509.
[21] J.T. Connor, R.D. Martin, L.E. Atlas, Recurrent neural networks and robust time series prediction, *IEEE Trans. Neural Netw.* 5 (2) (1994) 240–254.
[22] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput* 9 (8) (1997) 1735–1780.
[23] J.C. Song, L.Y. Zhang, G.X. Xue, Y.P. Ma, S. Gao, Q.L. Jiang, Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model, *Energy Build.* 243 (2021) 110998.
[24] X. Zhang, Y.Y. Zou, S.Y. Li, S.H. Xu, A weighted auto regressive LSTM based approach for chemical processes modeling, *Neurocomputing* 367 (2019) 64–74.
[25] Z.X. Guo, J.Z. Zhao, Z.J. You, Y.M. Li, S. Zhang, Y.Y. Chen, Prediction of coalbed methane production based on deep learning, *Energy* 230 (2021) 120847.
[26] F.F. Cheng, Q.P. He, J.S. Zhao, A novel process monitoring approach based on variational recurrent autoencoder, *Comput. Chem. Eng.* 129 (2019) 106515.
[27] S.X. Ji, X.H. Han, Y.C. Hou, Y. Song, Q.F. Du, Remaining useful life prediction of airplane engine based on PCA–BLSTM, *Sensors* 20 (16) (2020) 4537.
[28] J. Huang, X. Yan, Dynamic process fault detection and diagnosis based on dynamic principal component analysis, dynamic independent component analysis and Bayesian inference, *Chemom. Intell. Lab. Syst.* 148 (2015) 115–127.
[29] W.F. Ku, R.H. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, *Chemom. Intell. Lab. Syst.* 30 (1) (1995) 179–196.
[30] G. Li, B.S. Liu, S.J. Qin, D.H. Zhou, Dynamic latent variable modeling for statistical process monitoring, *IFAC Proc.* 44 (1) (2011) 12886–12891.
[31] Y.N. Dong, S.J. Qin, A novel dynamic PCA algorithm for dynamic data modeling and process monitoring, *J. Process. Control* 67 (2018) 1–11.
[32] A.V. Ooyen, B. Nienhuis, Improving the convergence of the back-propagation algorithm, *Neural Netw.* 5 (3) (1992) 465–471.
[33] J.A. Le, H.M. EI-Askary, M. Allali, D.C. Struppa, Application of recurrent neural networks for drought projections in California, *Atmospheric Research.* 188 (2017) 100–106.
[34] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, *IEEE Trans. Neural Netw.* 5 (2) (1994) 157–166.
[35] J. Forkman, J. Josse, H.P. Piepho, Hypothesis tests for principal component analysis when variables are standardized, *J. Agric. Biol. Environ. Stat.* 24 (2) (2019) 289–308.
[36] N.F. Thornhill, S.C. Patwardhan, S.L. Shah, A continuous stirred tank heater simulation model with applications, *J. Process. Control* 18 (3–4) (2008) 347–360.
[37] M.T. Amin, F. Khan, S. Ahmed, S. Imtiaz, A data-driven Bayesian network learning method for process fault diagnosis, *Process. Saf. Environ. Prot.* 150 (2021) 110–122.
[38] J. Yu, S.J. Qin, Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models, *AIChE J.* 54 (7) (2008) 1811–1829.
[39] C.E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* 27 (4) (1948) 623–656.