

Available online at www.sciencedirect.com

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherdIChemE
ADVANCING
CHEMICAL
ENGINEERING
WORLDWIDE

Prediction of gasoline yield in fluid catalytic cracking based on multiple level LSTM

Fan Yang^a, Yongsheng Sang^{a,*}, Jiancheng Lv^a, Jun Cao^{b,*}

^a College of Computer Science, Sichuan University, 610041, China

^b School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai, 200237, China

ARTICLE INFO

Article history:

Received 7 April 2022

Received in revised form 27 May 2022

Accepted 23 June 2022

Available online 28 June 2022

Keywords:

Fluid Catalytic Cracking

LSTM

Neural Network

AI

Gasoline yield

Reaction

ABSTRACT

Data-driven method has been widely used in Fluid Catalytic Cracking (FCC) process modeling. However, due to the complexity of chemical process both in time and spatial domain, how to reflect the time and spatial characteristics of FCC units and build corresponding model is important to construct a better model for the gasoline yield prediction. In this paper, a special neural network structure was developed to deal with the input variables with different time scales considering the collection characteristics of various variables, as well as the time continuity of large-scale process manufacturing units, LSTMs with different time scales are stacked to extract temporal and spatial features to help capture the relationship between influencing factors and product yield. The characteristics of FCC process are also fully reflected in data processing and building model. It is demonstrated from the conclusions that the new model developed in this paper performs better than the traditional LSTM networks, which will be of great help to the intelligent upgrading of the FCC process.

© 2022 Institution of Chemical Engineers. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Fluid Catalytic Cracking (FCC) is one of the most important chemical processes in the oil and gas industry. At present, 30 % of diesel and 70 % of gasoline in China are from the FCC process (Souza et al., 2018; Salvado et al., 2017; Lu et al., 2018; Yang et al., 2017), thus it is significant to build an accurate predicting model for improving the utilization rate of feed oil and the yield of high-value products. However, FCC is a highly complex process affected by many non-linear and underlying interactions. Numerous factors, such as feed quality, catalyst characteristics, as well as operating conditions, will affect the reaction process and product yield. In this case, it is a complex task to build a chemical process model to analyze the relationship between numerous factors and product yield. Different kinds of methods have been

proposed to handle the product yield predicting. These methods can be classified into two categories: mechanism-based methods and data-driven methods. Mechanism-based methods ignore many factors, which might be hard to completely simulate the chemical process, while data-driven methods could draw the relationship between numerous factors and product yield, which have been paid more and more attention though the mechanisms of the reactions are unclear or too complicated.

As one of the data-driven methods, artificial neural network (ANN) is skilled in dealing with non-linear problems and has been widely used in the chemical process modeling (Ma et al., 2001; Wang, 2015; Deng et al., 2011; Jiang et al., 2015). Various methods have been proposed to improve the performance of ANN and can be summarized as the following two categories. The first one is to adjust the hyper parameters by optimization algorithms, which could accelerate the convergence speed, and avoid the model to fall into local optimum. For example, the genetic algorithm (GA) and particle swarm optimization (PSO) methods are used to search the initial value and threshold of BP neural network

* Corresponding authors.

E-mail addresses: sangys@scu.edu.cn (Y. Sang),

caojun@ecust.edu.cn (J. Cao).

<https://doi.org/10.1016/j.cherd.2022.06.040>

0263-8762/© 2022 Institution of Chemical Engineers. Published by Elsevier Ltd. All rights reserved.

respectively in [Su et al., \(2016a\)](#) and [Gao et al. \(2017\)](#)., while the GA and PSO methods are applied to optimize BP neural network in [Su et al. \(2016b\)](#), which indicates that the performance of GA-BP is better than PSO-BP. The second category is the hybrid of ANN and mechanism model. A neural network model is developed to fit the differences between the lumped model results and the real values of product yield, then the results of lumped are modified ([Ouyang and Liu, 2017](#)). Furthermore, the results of the lumped model are jointed in the input variables of the neural network model to predict product yield in [Yang et al. \(2020\)](#). However, these methods still have some shortcomings. Adjusting the hyper parameters of ANN cannot improve the fitting ability for the set of hypothesis functions, although it can help the researcher find better fitting functions from a given hypothesis space. The hybrid of neural network and mechanism model has strict requirements on data quality, although it can improve the prediction accuracy. In the actual production process, it is difficult to satisfy its requirements. In addition, the chemical process is a continuous process in the time domain, and temporal features of time series data should be considered as well, while ANN ignores the temporal features in the time domain.

To improve the performance of the above methods, Long Short-Term Memory (LSTM) model which could extract temporal features, is adopted for chemical process modeling ([Ke et al., 2017](#); [Wang et al., 2018](#)). A deep bidirectional LSTM is developed for predicting product yield in ref ([Zhang et al., 2018](#)). Moreover, a WAR-LSTM model is proposed to predict product yield in ref ([Zhang et al., 2019](#)), where correction information and prior knowledge are extracted from historical yield for the predictive model. However, methods in ref ([Zhang et al., 2018, 2019](#)). cannot be applied to adjust controllable operating variables, which is significant for the guidance to improve the production results in the actual production process. It is hard to obtain the future variables which are implemented in the deep bidirectional LSTM in [Zhang et al. \(2018\)](#). In [Zhang et al. \(2019\)](#), the historical values are taken as the input variables, which will cause two bad effects. On the one hand, the weight of controllable operating variables in the model will be reduced, since the difference between the product yields at the adjacent time is very small. On the other hand, it is difficult to adjust the parameters in a short time, since the model can only predict the yield at next moment. Besides, the recurrent neural network for chemical process modeling usually uses limited length time series, which cannot reflect the time delay characteristics of various factors in large-scale equipment, such as Fluid Catalytic Cracking Units (FCCU).

While collecting FCCU data, the values of various operating conditions in the distributed control system (DCS) are recorded in real-time, and the values of feed quality in the laboratory information management system (LIMS) are manually recorded, which leads to a much larger recording interval of LIMS than that of DCS. DCS and LIMS data are used as input variables and these various variables need to be aligned in the time domain. According to the characteristics of FCCU, we propose a Multi-Level LSTM (ML-LSTM). Input variables with different time scales will be aligned through a special network structure, and significant temporal as well as spatial features could be extracted by stacking multiple LSTM structures. The main contributions are as follows:

- (1) A special network structure is built to deal with the difference of data recording frequency, and takes the more comprehensive factors into consideration, which influence on product yield.
- (2) A new neural network model is proposed to predict the product yield of FCCU, which reflects the time continuity of the production process of large-scale process manufacturing units.
- (3) The effectiveness of ML-LSTM is proved by analyzing the relationship between the product yield and the features, which are extracted by multiple LSTM structures.

In this paper, the principle of FCCU, as well as the collection and storage methods of production data, and analysis on gasoline yield used in the paper are introduced in [Section 2](#). In [Section 3](#), we present the structure of LSTM, and the ML-LSTM is proposed to predict gasoline yield according to the characteristics of FCCU. In [Section 4](#), the actual process data from FCCU are utilized to verify the effectiveness of our model, and the prediction results are analyzed and discussed. Finally, the concluding remarks are pointed out in [Section 5](#) of this paper.

2. Problem formulation

2.1. Overview of FCC process

The main purpose of FCCU is to convert heavy oil to high-value light transportation fuels in the condition of suitable temperature, pressure, and catalyst. There are three sub-systems in FCCU: reactor-regenerator system, fractionation system, and absorption-stabilization system. The main task of the reactor-regenerator system is to convert feed oil, wax oil, and residual oil to high-value fuels. In the fractionation system, according to the different boiling points, the high-temperature oil and gas from the reactor-regenerator system are cut into rich oil, naphtha, diesel oil and slurry oil. In the absorption-stabilization system, naphtha and rich oil are separated to dry gas, gasoline, and liquid hydrocarbons according to different solubility in liquids.

Among the above products from FCCU, gasoline is a high-value product and an important engine fuel. To build the prediction model of gasoline yield is the key research goal in the paper. In the process of FCC, feed quality, catalyst characteristics, and operating conditions have an important impact on the gasoline yield ([Ma et al., 2001](#)). Feed quality and catalyst characteristics mainly include aromatics, saturates, resins and coke. The operating conditions mainly consist of temperature and pressure variables.

2.2. Collection methods and analysis on FCCU data

LIMS and DCS systems are widely used in collecting FCCU data. The LIMS system mainly manages the data of feed oil and regenerated catalysts, while the DCS system mainly records operating variables and manage the mass balance data.

2.3. The goals for modeling

The ultimate goal of chemical process modeling is that the model can be used to adjust operating variables, and then improve product yield. The premise is that a prediction model for product yield needs to be built, of which the

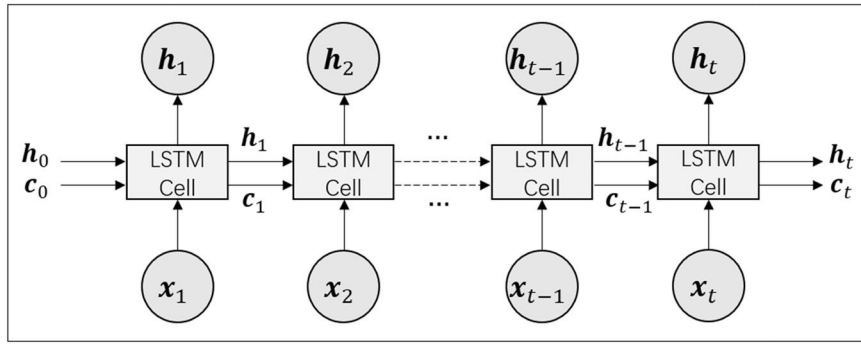


Fig. 1 – The structure of LSTM. t denotes all the time steps, x_t denotes input variables, h_t and c_t denote the hidden state and cell state of LSTM, respectively, h_0 is the initial value of hidden state, and c_0 is the initial value of cell state.

predicted result is closest to the real value. It can be defined as follows:

$$\operatorname{argmin}_f L(f(\mathbf{X}), \mathbf{Y}) \quad (1)$$

where \mathbf{X} denotes the time series data of various factors, \mathbf{Y} denotes the real values of product yield, $f(\cdot)$ denotes the relationship between various factors and product yield, and $L(\cdot)$ measures the difference between the predicted values and the real values.

2.4. Challenges for modeling

The challenges of building a gasoline yield predicting model are as follows: Firstly, in the production process of FCCU, it is hard to clarify which parameters affect the gasoline yield at each time, since the feed oil stays in the FCCU for a while and the operating conditions are dynamic. Secondly, due to the different collection frequency of data, when taking the DCS and LIMS data as input variables, it is essential to align variables in the time domain. In addition, while applying the historical yields as the input variables, the model weights of the controllable operating variables might be too small to be analyzed.

3. Multi-level LSTM models

Long Short-Term Memory (LSTM) is a typical recurrent neural network and can extract the temporal features in time series data. However, there are large differences between DCS data and LIMS data on the collection frequency. In this case, various variables need to be aligned in the time domain while being implemented as input variables. Multi-level LSTM (ML-LSTM) is proposed in the paper, where a special network is built to deal with the input variables with different collection frequencies, and a novel training method is used to help the model to capture the long-term dependencies in FCCU data.

3.1. The structure of long short-term memory

As a typical recurrent neural network, LSTM (Hochreiter and Schmidhuber, 1997) is successfully applied in many tasks, such as speech recognition and natural language processing (Zhang and Yang, 2018; Tian et al., 2019). The structure of LSTM is shown in Fig. 1.

In Fig. 1, h_t is the output of LSTM at t th time step, and is the accumulation of all input variables, such as x_1, x_2, x_{t-1}, x_t . The computations of h_t and c_t are affected by each input

variable. We use $G_h(\cdot)$ and $G_c(\cdot)$ to denote the relationship between $(x_t, x_{t-1}, \dots, x_2, x_1, h_0, c_0)$ and h_t as well as c_t , and they can be described as follows:

$$h_t = G_h(x_t, x_{t-1}, \dots, x_2, x_1, h_0, c_0; \Theta_h) \quad (2)$$

$$c_t = G_c(x_t, x_{t-1}, \dots, x_2, x_1, h_0, c_0; \Theta_c) \quad (3)$$

where Θ_h and Θ_c denote the parameters of $G_h(\cdot)$ and $G_c(\cdot)$, respectively.

A LSTM cell is the calculation at each time step, and its structure is shown in Fig. 2. The cell has input to input gate, forget gate, output gate and cell state connections, and they are parameterized by weight matrices U_i, U_f, U_o and U_c , respectively. It also has hidden state to input gate, forget gate, output gate and cell state connections, and they are parameterized by weight matrices W_i, W_f, W_o and W_c , respectively.

LSTM cell can be described by the following equations:

$$i_t = \sigma(U_i x_t + W_i h_{t-1}) \quad (4)$$

$$f_t = \sigma(U_f x_t + W_f h_{t-1}) \quad (5)$$

$$o_t = \sigma(U_o x_t + W_o h_{t-1}) \quad (6)$$

$$\tilde{c}_t = \tanh(U_c x_t + W_c h_{t-1}) \quad (7)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (8)$$

$$h_t = o_t \circ \tanh(c_t) \quad (9)$$

where \circ denotes the element-wise product, \tilde{c}_t and c_t denote cell state and final cell state, and h_t denotes hidden state. $\sigma(\cdot)$ and $\tanh(\cdot)$ denote sigmoid and tanh activation functions. For a given scalar (denoted as a), the computations of $\sigma(\cdot)$ and $\tanh(\cdot)$ are described as follows:

$$\sigma(a) = \frac{1}{1 + e^{-a}} \quad (10)$$

$$\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \quad (11)$$

It can be seen that h_t and c_t are affected by h_{t-1} and c_{t-1} , and the computations of h_t and c_t are recursive. We use $L_h(\cdot)$ and $L_c(\cdot)$ to represent the computations, which are described as (12) and (13).

$$h_t = L_h(x_t, h_{t-1}, c_{t-1}; \theta_h) \quad (12)$$

$$c_t = L_c(x_t, h_{t-1}, c_{t-1}; \theta_c) \quad (13)$$

where θ_h and θ_c denote the parameters of $L_h(\cdot)$ and $L_c(\cdot)$.

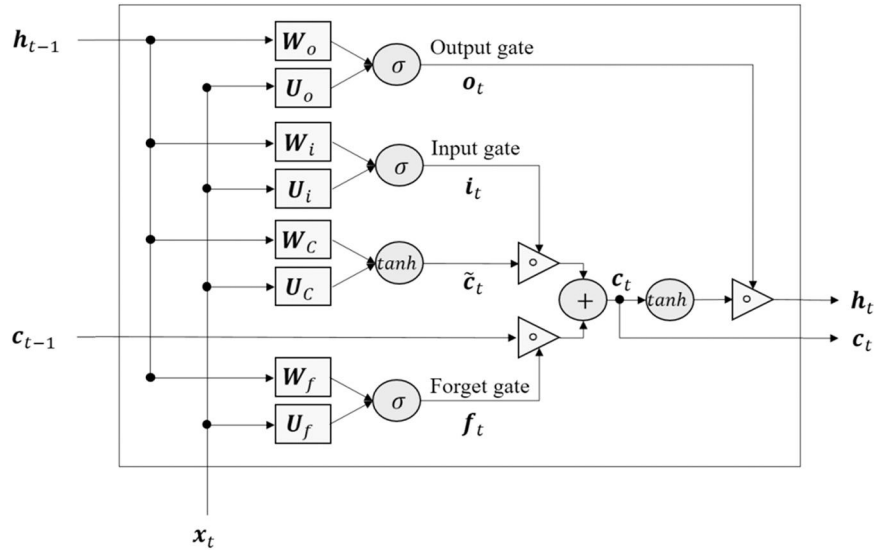


Fig. 2 – The structure of LSTM cell.

3.2. The development of multi-level LSTM

While data-driven methods are applied in developing a predicting model for product yield, researchers mainly utilize feed quality, catalyst characteristics, and operating conditions (Ma et al., 2001; Wang, 2015; Deng et al., 2011; Jiang et al., 2015; Su et al., 2016a). The models which use historical product yield as input variables is hard to be used in adjusting the controllable operating variables. Based on data analysis and expert experience, a few variables of feed quality are filtered from the LIMS system, such as feed density, weight of carbon residue as well as the mass weight percentage of aromatics, and lots of operating variables are selected from the DCS system, such as temperatures and pressures in all units. According to the characteristics of FCC, the ML-LSTM is proposed, and its structure is shown in Fig. 3. The ML-LSTM is composed of Input Layer, LSTM Layer, FC Layer and Output Layer. LSTM Layer consists of Layer LSTM-I, Concat-Layer, and Layer LSTM-II. \mathbf{X}^l denotes the input variables of LIMS data, \mathbf{X}^d denotes the input variables of DCS data, and $\hat{\mathbf{Y}}$ denotes the outputs of our model. N denotes the number of time steps.

ML-LSTM receives data of different time scales through the input layer and extracts features through the LSTM layer and the fully connected layer. The output layer gives the predicted value. The prediction of our model can be described as follows:

$$\hat{\mathbf{Y}} = f(\mathbf{X}^l, \mathbf{X}^d; \alpha) \quad (14)$$

where α denotes all the parameters of our model.

The Input Layer, LSTM Layer, FC Layer and Output Layer of our model are detailed as follows.

(1) Input Layer

The Input Layer of the ML-LSTM receives LIMS and DCS data, of which sampling time interval are 60 min and 5 min, respectively, in the paper. The input variables \mathbf{X}^l and \mathbf{X}^d on training set can be shown in Fig. 4. The row and column denote the time and feature dimensions, respectively. The shape of \mathbf{X}^l is $(N * m)$, and the shape of \mathbf{X}^d is $(12N * n)$.

In Fig. 4, the numbers of records of LIMS and DCS data are N and $12N$, respectively. LIMS data in green box and DCS

data in yellow box are at the same time period. The relationship between the LIMS and DCS data in time domain is shown in Fig. 5.

T_i denotes the i th time period. At T_i , we collect 1 record of LIMS data and 12 records of DCS data, which are denoted as \mathbf{x}_i^l and $(\mathbf{x}_{12i-11}^d, \mathbf{x}_{12i-10}^d, \dots, \mathbf{x}_{12i}^d)$. \mathbf{x}_i^l has the same beginning time with \mathbf{x}_{12i-11}^d , and the same ending time with \mathbf{x}_{12i}^d .

When predicting the gasoline yield at T_i , the input variables of our model consist of \mathbf{x}_i^l and $(\mathbf{x}_{12i-11}^d, \mathbf{x}_{12i-10}^d, \dots, \mathbf{x}_{12i}^d)$.

(2) LSTM Layer

LSTM Layer consists of Layer LSTM-I, Concat-Layer and Layer LSTM-II. In the Layer LSTM-I, two LSTM structures are used to process LIMS and DCS data, respectively, and the computations are described as follows.

$$\mathbf{h}_i^l = \begin{cases} \mathbf{h}_0^l, & \text{if } i = 0 \\ L_h(\mathbf{x}_i^l, \mathbf{h}_{i-1}^l, \mathbf{c}_{i-1}^l; \theta_h^l), & \text{if } 1 \leq i \leq N \end{cases} \quad (15)$$

$$\mathbf{c}_i^l = \begin{cases} \mathbf{c}_0^l, & \text{if } i = 0 \\ L_c(\mathbf{x}_i^l, \mathbf{h}_{i-1}^l, \mathbf{c}_{i-1}^l; \theta_c^l), & \text{if } 1 \leq i \leq N \end{cases} \quad (16)$$

$$\mathbf{h}_j^d = \begin{cases} \mathbf{h}_0^d, & \text{if } j = 0 \\ L_h(\mathbf{x}_j^d, \mathbf{h}_{j-1}^d, \mathbf{c}_{j-1}^d; \theta_h^d), & \text{if } 12i - 11 \leq j \leq 12i \end{cases} \quad (17)$$

$$\mathbf{c}_j^d = \begin{cases} \mathbf{c}_0^d, & \text{if } j = 0 \\ L_c(\mathbf{x}_j^d, \mathbf{h}_{j-1}^d, \mathbf{c}_{j-1}^d; \theta_c^d), & \text{if } 12i - 11 \leq j \leq 12i \end{cases} \quad (18)$$

where $\theta_h^l, \theta_c^l, \theta_h^d$ and θ_c^d denote parameters, and $\mathbf{h}_0^l, \mathbf{c}_0^l, \mathbf{h}_0^d$ and \mathbf{c}_0^d are initial vectors, of which the elements are set as 0.

Then, the alignment of various variables with different time scales is carried out in the Concat-Layer, which is shown in Fig. 6.

At T_i , \mathbf{x}_i^l and $(\mathbf{x}_{12i-11}^d, \mathbf{x}_{12i-10}^d, \dots, \mathbf{x}_{12i}^d)$ are processed by Layer LSTM-I, and \mathbf{h}_i^l and \mathbf{h}_{12i}^d are obtained, which can denote the accumulated information of LIMS and DCS data at the end of i th time period, respectively. The concatenation of \mathbf{h}_i^l and \mathbf{h}_{12i}^d can be used to denote the accumulated information of all input variables. The concatenation can be described as follows.

$$\mathbf{h}_i^c = \text{Concat}(\mathbf{h}_i^l, \mathbf{h}_{12i}^d) \quad (19)$$

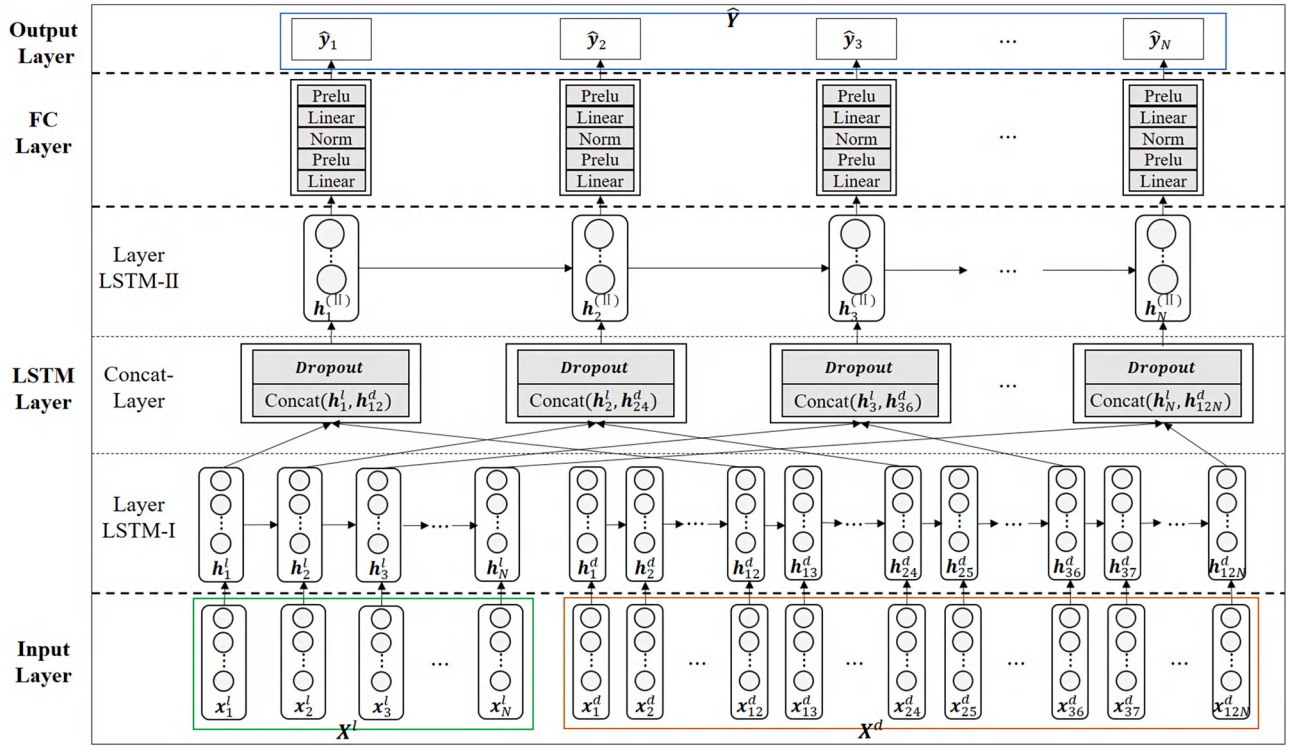


Fig. 3 – The structure of ML-LSTM.

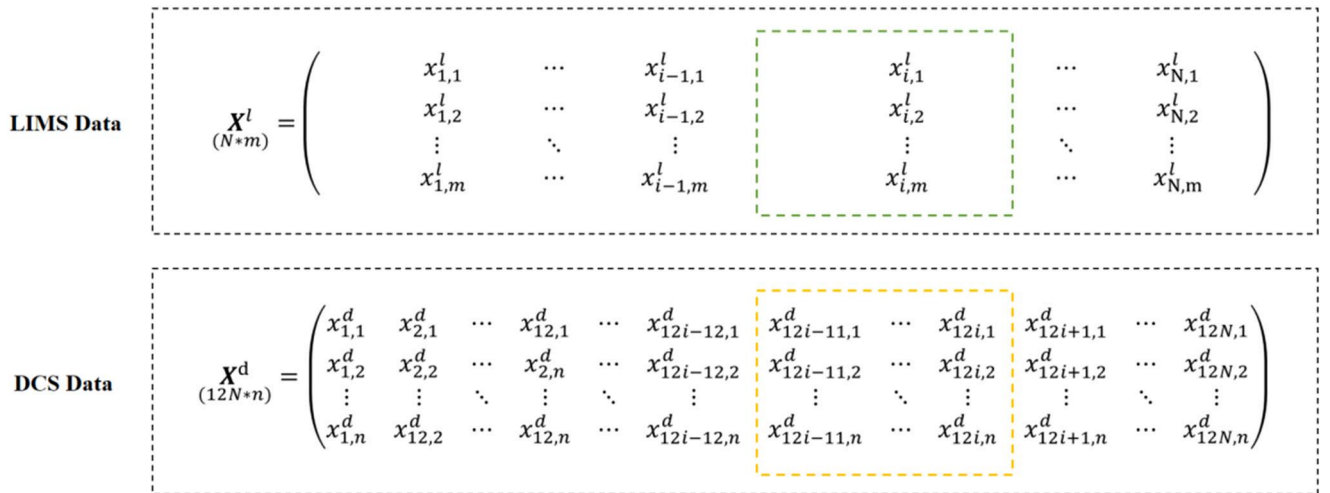


Fig. 4 – Presentation of LIMS and DCS time series data on training set.

Then, we use *Dropout* method to process h_i^c , and get \tilde{h}_i^c . The calculation is shown in Eq. (20).

$$\tilde{h}_i^c = r_i \odot h_i^c \quad (20)$$

where r_i is a vector of independent Bernoulli random variables, which has the same dimension with h_i^c , and each variable has a probability p of being 1. Furthermore, the computation of Layer LSTM-II can be described as Eqs. (21) and (22).

$$h_i^{(II)} = \begin{cases} h_0^{(II)}, & \text{if } i = 0 \\ L_h(\tilde{h}_i^c, h_{i-1}^{(II)}, c_{i-1}^{(II)}; \theta_h^{(II)}), & \text{if } 1 \leq i \leq N \end{cases} \quad (21)$$

$$c_i^{(II)} = \begin{cases} c_0^{(II)}, & \text{if } i = 0 \\ L_c(\tilde{h}_i^c, h_{i-1}^{(II)}, c_{i-1}^{(II)}; \theta_c^{(II)}), & \text{if } 1 \leq i \leq N \end{cases} \quad (22)$$

where $\theta_h^{(II)}$ and $\theta_c^{(II)}$ denote parameters, and $h_0^{(II)}$ and $c_0^{(II)}$ are initial vectors, of which the elements are set as 0.

(3) FC Layer

This section describes the FC Layer. Let $R(\cdot)$ and $N(\cdot)$ denote activation function and Layer Normalization, respectively. $w^{(1)}$ and $b^{(1)}$ denote the weights and biases at the first fully connected layer, and $w^{(2)}$ and $b^{(2)}$ denote the weights and biases at the second fully connected layer. FC Layer can be described as (23):

$$FC(h_i^{(II)}) = R(w^{(2)}(N(R(w^{(1)}h_i^{(II)} + b^{(1)}))) + b^{(2)}) \quad (23)$$

In Eq. (23), $N(\cdot)$ and $R(\cdot)$ are described as follows:

$$N(z) = \frac{z - u}{v} \odot \gamma + \beta \quad (24)$$

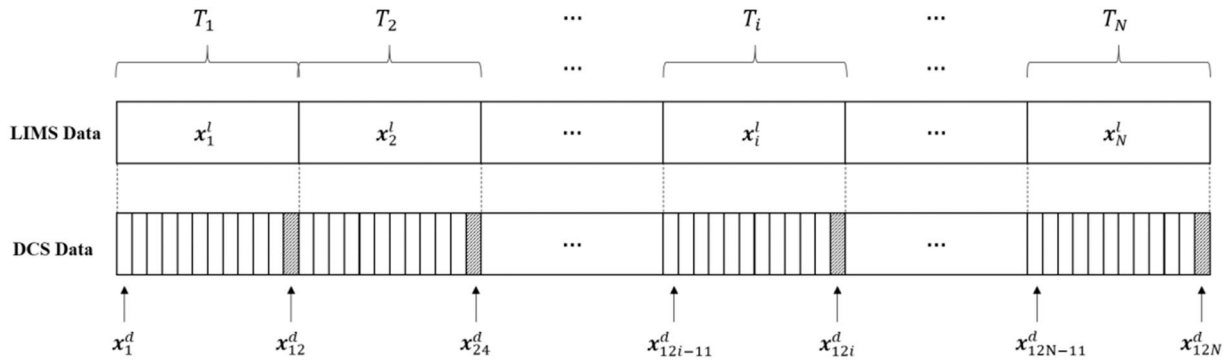


Fig. 5 – The relationship of LIMS and DCS data in time domain. x_i^l denotes i th LIMS data, and x_{12i}^d denotes 12th DCS data.

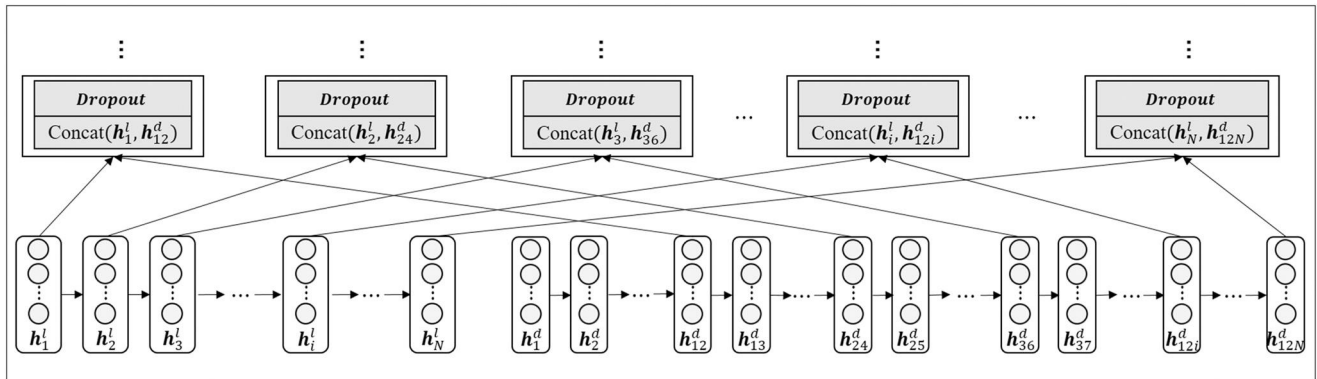


Fig. 6 – Alignment of input variables with different time scales. $\text{Concat}(h_i^l, h_{12i}^d)$ is the alignment result of x_i^l and $(x_{12i-11}^d, x_{12i-10}^d, \dots, x_{12i}^d)$.

$$R(a) = \begin{cases} a, & \text{if } a \geq 0 \\ \rho a, & \text{if } a < 0 \end{cases} \quad (25)$$

where \mathbf{z} denotes the vector of outputs from the first fully connected layer, u denotes mean value of \mathbf{z} , v denotes variance of \mathbf{z} , β and γ are defined as the bias and gain parameters of the same dimension as \mathbf{z} , respectively, ρ is learnable parameter, and a is a scalar.

(4) Output Layer

\hat{y}_i denotes the final output of model, which is the predicted value of gasoline yield at i th time period, and is one value of $\hat{\mathbf{Y}}$.

3.3. The computation of training and predicting of ML-LSTM

(1) In the training phase

The whole calculation of the neural network on the training set is described as follows:

$$h_i^l = G_h(\mathbf{x}_i^l, \mathbf{x}_{i-1}^l, \dots, \mathbf{x}_1^l, \mathbf{h}_0^l, \mathbf{c}_0^l; \Theta_h^l) \quad (26)$$

$$h_{12i}^d = G_h(\mathbf{x}_{12i}^d, \mathbf{x}_{12i-1}^d, \dots, \mathbf{x}_1^d, \mathbf{h}_0^d, \mathbf{c}_0^d; \Theta_h^d) \quad (27)$$

$$h_i^{(II)} = L_h(\mathbf{r}_i^l \text{Concat}(h_i^l, h_{12i}^d), h_{i-1}^{(II)}, \mathbf{c}_{i-1}^{(II)}; \Theta_h^{(II)}) \quad (28)$$

$$\hat{y}_i = FC(h_i^{(II)}) \quad (29)$$

where Θ_h^l and Θ_h^d denote parameters.

It can be seen that the calculation of $h_i^{(II)}$ is recursive. The prediction of our model is consistent with the time con-

tinuity of large-scale process manufacturing device. To find the non-linear map ($f(\cdot)$) shown in Eq. (1) as accurate as possible, in the training phase, the difference between predicted value and real value is measured by (30).

$$L(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (30)$$

where \mathbf{Y} denotes all the real values of gasoline yield, and y_i is the i th real value.

(2) In the predicting phase

The first record of the test set and the last record of the training set are adjacent in time domain. The k th predicted value in the test set is denoted as \hat{y}_{N+k} , and the calculation can be described as follows:

$$\hat{y}_{N+k} = FC(L_h(\mathbf{r}_{N+k}^l \text{Concat}(h_{N+k}^l, h_{12(N+k)}^d), h_{N+k-1}^{(II)}, \mathbf{c}_{N+k-1}^{(II)}; \Theta_h^{(II)})) \quad (31)$$

where, h_{N+k}^l , $h_{12(N+k)}^d$, $h_{N+k-1}^{(II)}$ and $\mathbf{c}_{N+k-1}^{(II)}$ are:

$$h_{N+k}^l = G_h(\mathbf{x}_{N+k}^l, \mathbf{x}_{N+k-1}^l, \dots, \mathbf{x}_1^l, \mathbf{h}_N^l, \mathbf{c}_N^l; \Theta_h^l) \quad (32)$$

$$h_{12(N+k)}^d = G_h(\mathbf{x}_{12(N+k)}^d, \mathbf{x}_{12(N+k)-1}^d, \dots, \mathbf{x}_1^d, \mathbf{h}_N^d, \mathbf{c}_N^d; \Theta_h^d) \quad (33)$$

$$h_{N+k-1}^{(II)} = G_h(\tilde{h}_{N+k}^c, \tilde{h}_{N+k-1}^c, \dots, \tilde{h}_{N+1}^c, h_N^{(II)}, \mathbf{c}_N^{(II)}; \Theta_h^{(II)}) \quad (34)$$

$$\mathbf{c}_{N+k-1}^{(II)} = G_c(\tilde{h}_{N+k}^c, \tilde{h}_{N+k-1}^c, \dots, \tilde{h}_{N+1}^c, h_N^{(II)}, \mathbf{c}_N^{(II)}; \Theta_h^{(II)}) \quad (35)$$

The computations of h_{N+k}^l , $h_{12(N+k)}^d$, $h_{N+k-1}^{(II)}$ and $\mathbf{c}_{N+k-1}^{(II)}$ are recursive, and the prediction process of our model on the test set can be shown in Fig. 7.

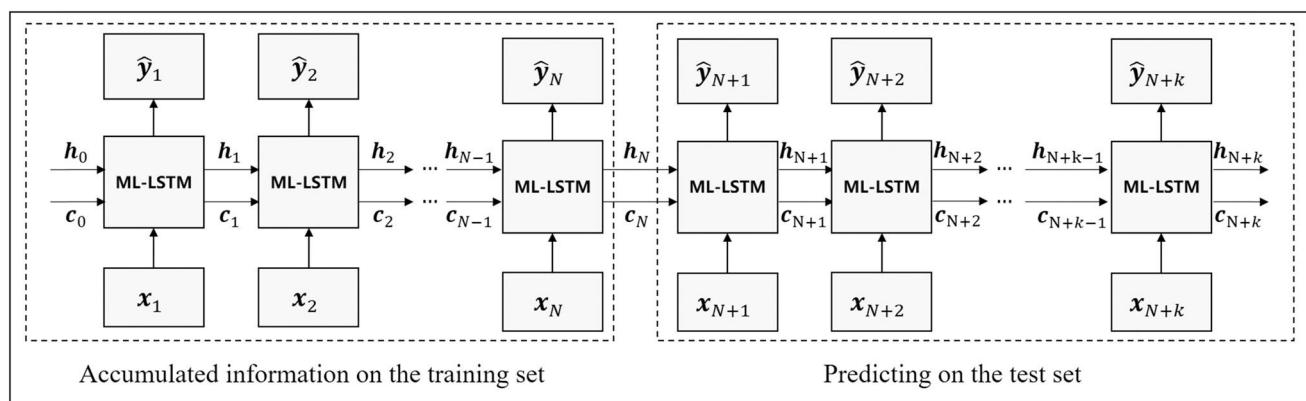


Fig. 7 – Prediction of ML-LSTM model on the test set.

In Fig. 7, h_N and c_N are the simple representation of ML-LSTM hidden states and final cell states on training set, and they are essential for the prediction of our model on test set.

3.4. Apply ML-LSTM to predict gasoline yield

Fig. 8 shows the steps of ML-LSTM offline modeling and on-line prediction.

(1) Offline modeling

Step 1 Collect LIMS data and DCS data, interpolate LIMS data, smooth DCS data and then obtain training dataset.
Step 2 Use Z-score to standardize each feature of training set, and it is described as Eq. (36).

$$x_i^{(j)} = \frac{x_i^{(j)} - \bar{x}^{(j)}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^{(j)} - \bar{x}^{(j)})^2}} \quad (36)$$

where N denotes the length of training set, $x_i^{(j)}$ denotes the value of the j th variable at the i th moment, and $\bar{x}^{(j)}$ denotes the mean value of the j th variable.

Step 3 Convert the train dataset into the form of $S \times B \times D$, where S is the sequence length of one sample, B is the batch size, and D is the number of features.

Step 4 Choose the suitable optimizer and hyper parameters, and train the model.

Step 5 Calculate the loss in Eq. (30). If the loss is less than

the threshold or the iteration reaches the maximum number, stop training.

Step 6: Save the parameters of model.

(2) Online prediction

Step 1: Process the collected data, and obtain the test dataset.

Step 2: Standardize the test dataset according to the means and variances of features in the training set.

Step 3: Predict gasoline yield using the method shown in Fig. 7.

4. Experiments

4.1. Experiment setup

Theoretically, the data applied here should have the same starting time as FCCU, but this requirement is difficult to be satisfied in data collection. In this paper, the starting time of data collected from a refinery is set as 1, and the data is preprocessed as follows: Firstly, the DCS data sets with sampling interval of 5 min and 60 min are obtained by sliding time window and calculating the mean value of the window. Secondly, the LIMS data set with sampling interval of 60 min is obtained by B-spline interpolation. Then, the LIMS and DCS sets, of which durations are 2256 h, are split into training set and test set. The first 2088 h of data sets are used as training set, and the next 168 h are used as test set. In addition, the batch-size of the training and test sets are set as 8 and 1, respectively.

Five models are built in the paper, and are listed as follows. 1) MLP is a three-layer neural network, which uses LIMS and DCS data with sampling interval of 60 min as input; 2) LSTM with DCS data is a deep neural network based on LSTM, which uses DCS data with sampling interval of 5 min as input; 3) LSTM with LIMS data is a deep neural network based on LSTM, which uses LIMS data with sampling interval of 60 min as input; 4) LSTM with AVG data, which has the same structure as ML-LSTM, uses LIMS and DCS data with sampling interval of 60 min as input; 5) ML-LSTM is based on LIMS data with sampling interval of 60 min and DCS data with sampling interval of 5 min

The results of the five models are analyzed and compared. Firstly, the prediction result of LSTM with LIMS data is analyzed. Secondly, the effectiveness of LSTM structure is

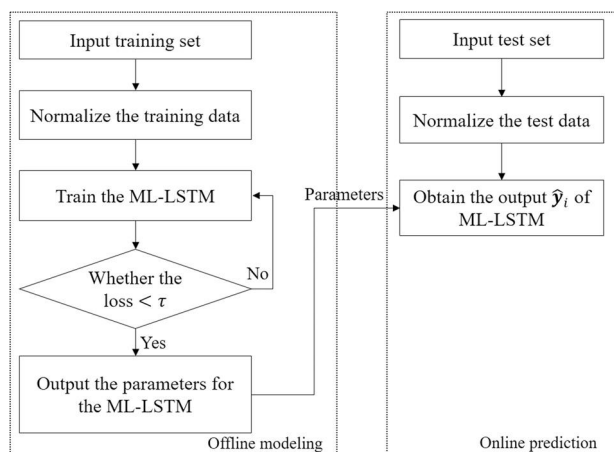


Fig. 8 – Steps of gasoline yield prediction based on ML-LSTM.

Table 1 – Performance of all models on the training set and test set.

Model	RMSE		PearsonR		R ²	
	Train	Test	Train	Test	Train	Test
MLP	0.6432	0.7133	0.9012	0.5271	0.7814	0.1653
LSTM with LIMS data	0.9560	0.8194	0.7432	0.2865	0.5172	-0.1014
LSTM with DCS data	0.5632	0.6066	0.9161	0.6533	0.8324	0.3964
LSTM with AVG data	0.6214	0.6236	0.8990	0.6622	0.7960	0.3620
ML-LSTM	0.5732	0.5851	0.9100	0.7671	0.8264	0.4384

verified by comparing the prediction results of MLP, LSTM with AVG data and LSTM with DCS data. Then, the differences of prediction results between LSTM with DCS data and LSTM with AVG data are analyzed. Finally, the effectiveness of the proposed method is verified by comparing the results of ML-LSTM and other models.

What's more, we use Adam as optimizer, and set the alpha1, alpha2 and learning rate of Adam as 0.9, 0.99 and 0.001, respectively. The weight decay of ML-LSTM, LSTM with AVG data, MLP, LSTM with LIMS data and LSTM with DCS data are 0.00003, 0.000015, 0.00002, 0.00002, 0.00002, respectively.

4.2. Evaluation criterion

The criteria to compare performance are Root Mean Square Error (RMSE), Pearson correlation coefficient (PearsonR) and R square (R²).

RMSE represents the absolute error between the prediction and the real values, and reflects the accuracy of the model on the test set. RMSE is calculated as follows.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (37)$$

where n is the number of samples, y_i is the i th real value, and \hat{y}_i is the i th predicted value.

PearsonR measures the relationship between the real values and the predicted values, it is computed as (38).

$$PearsonR = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (38)$$

where \bar{y} denotes the mean value of the real values, and $\bar{\hat{y}}$ is the mean value of the predicted values. The value of PearsonR is between -1 and 1. The closer its value is to 1, the more relevant the predicted value is to the real data.

R² reflects the goodness of the model fitness, and it is calculated as (39).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (39)$$

The maximum value of R² is 1. The closer its value is to 1, the better model fits.

4.3. Experiment results

Table 1 shows the RMSE, PearsonR and R² of MLP, LSTM with LIMS data, LSTM with DCS data, LSTM with AVG data, and ML-LSTM on the training set and test set.

It can be seen that the RMSE of the ML-LSTM model is 0.5732 on the training set, which shows that the ML-LSTM model can fit the gasoline yield well, while the RMSE of the ML-LSTM model is 0.5851 on the testing set, and the error is smaller than other models. It can be seen that ML-LSTM has better prediction ability. On the training set and the test set, PearsonR of ML-LSTM is 0.91 and 0.7667, respectively. The ML-LSTM model can accurately fit and predict the trend of gasoline yield over time. R² of ML-LSTM on the test set is 0.4384, which is larger than other approaches. ML-LSTM gets better performance.

The prediction results of MLP, LSTM with LIMS data, LSTM with DCS data, LSTM with AVG data and ML-LSTM on the test set are shown in Fig. 9. It can be seen that LSTM with LIMS data can only roughly predict the trend of gasoline yield over time. Although MLP, LSTM with DCS data and LSTM with AVG data can predict the trend of gasoline yield over time well, the accuracy is still insufficient. ML-LSTM accurately predicts the values and trend of gasoline yield over time. Therefore, our approach shows a competitive advantage.

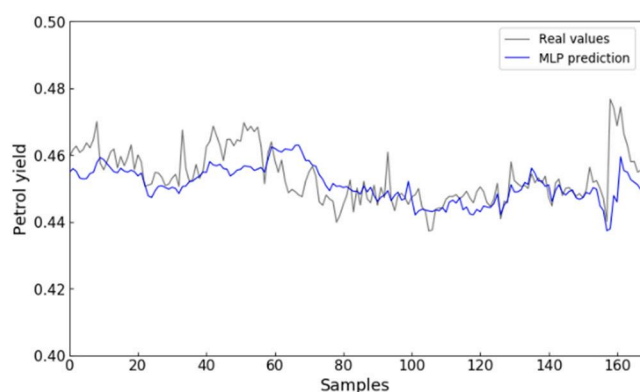
4.4. Analysis and discussion

LSTM with LIMS data gets poor performance compared with MLP on the test set. The reasons may be that LSTM with LIMS data ignores operating variables and is based on poor-quality data. However, LSTM with LIMS data can roughly predict the trend of gasoline yield over time, showing that LIMS data is helpful. Furtherly, LSTM with DCS data and LSTM with AVG data show great advantages in the prediction of gasoline yield. Obviously, it is easier for LSTM structure to find the relationship between factors and gasoline yield.

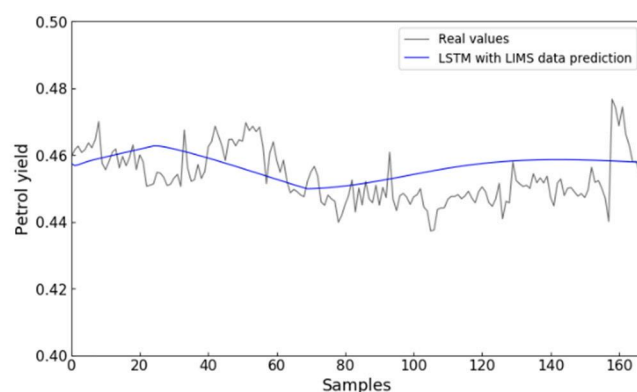
LSTM with AVG data significantly outperforms LSTM with DCS data on PearsonR. It can be seen that it is helpful to capture the trend of gasoline yield over time by combining LIMS and DCS data. However, RMSE of LSTM with AVG data is higher than RMSE of LSTM with DCS data. It is inferred that the DCS data with high sampling frequency can provide more information for the model.

ML-LSTM, which uses LIMS data, DCS data with high sampling frequency, and LSTM structure, gets the best performance. It is obvious that ML-LSTM with different sampling frequencies data performs an excellent ability to capture the relationship between various factors and gasoline yield.

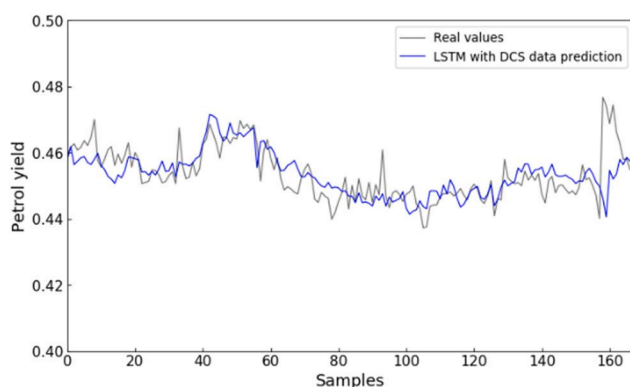
To analyze the effectiveness of ML-LSTM, the hidden states of ML-LSTM and LSTM with AVG data on the testing set are extracted to calculate PearsonR between each feature



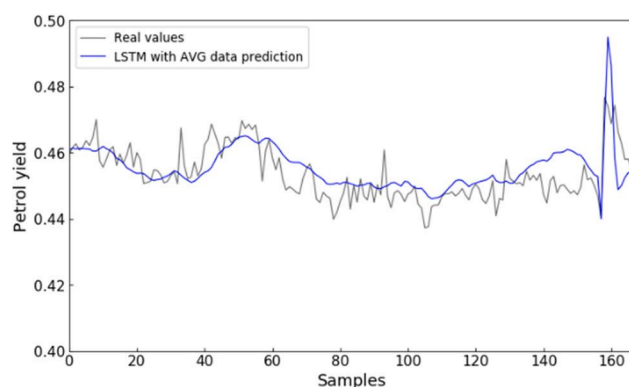
(a) Prediction result of MLP



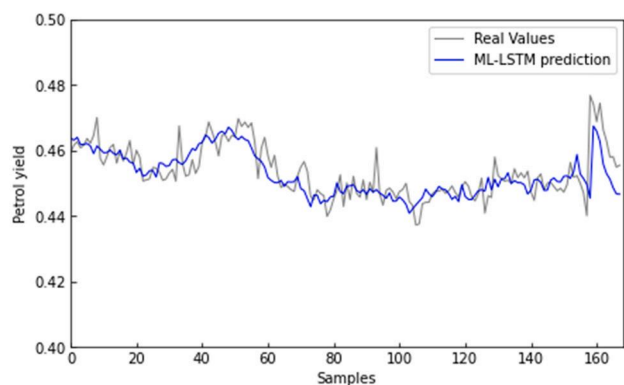
(b) Prediction result of LSTM with LIMS data



(c) Prediction result of LSTM with DCS data



(d) Prediction result of LSTM with AVG data



(e) Prediction result of ML-LSTM

Fig. 9 – Prediction results of all models on the test set. (a) Prediction result of MLP. (b) Prediction result of LSTM with LIMS data. (c) Prediction result of LSTM with DCS data. (d) Prediction result of LSTM with AVG data. (e) Prediction result of ML-LSTM.

in the hidden state and the gasoline yield. The results are shown in Fig. 10, where the larger values are represented by darker lines, and the smaller values are represented by lighter lines. It can be seen that the correlation between ML-LSTM hidden state and gasoline yield has more dark lines than the correlation between LSTM hidden state and gasoline yield.

The histogram to compare the difference is shown in Fig. 11, where the horizontal axis and the vertical axis represent the Pearson correlation coefficient and the number of features, respectively. In Fig. 11(a), the correlation is centered on 0. Fig. 11(b) shows a bimodal distribution of correlation, with peaks of 0.5 and - 0.5. On the whole, ML-LSTM shows a great competitive advantage in feature extraction.

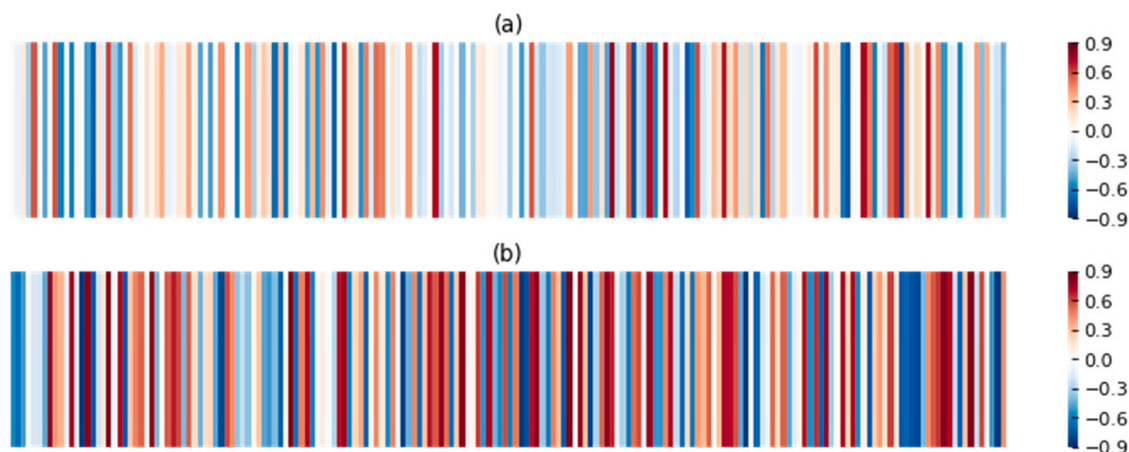


Fig. 10 – PearsonR between features in the hidden state and gasoline yield on the test set: (a) PearsonR between hidden state of LSTM with AVG data and gasoline yield, (b) PearsonR between hidden state of ML-LSTM and gasoline yield.

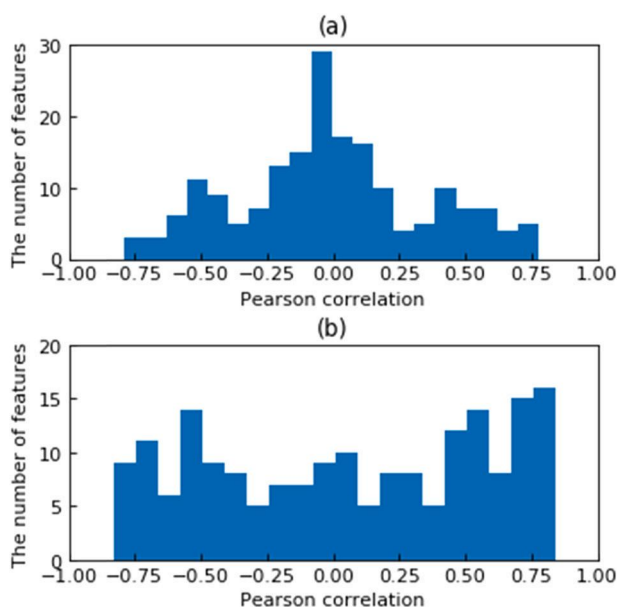


Fig. 11 – Histogram of PearsonR: (a) histogram of PearsonR between hidden state of LSTM with AVG data and gasoline yield, (b) histogram of PearsonR between hidden state of ML-LSTM and gasoline yield.

Figs. 10 and 11 demonstrate that ML-LSTM is superior to LSTM model in feature extraction.

5. Conclusions

According to the characteristics of FCCU, ML-LSTM is proposed to predict the gasoline yield in this paper. Firstly, a special network structure is considered to align the input variables with different collection frequencies in the time domain. Secondly, LSTMs with different time scales are stacked to extract high-level representations from multi variables, and a novel training method is applied to help the model to capture the long-term dependencies in the data of FCCU. These experiments illustrate that compared with the LSTM model, RMSE of ML-LSTM is lower, and the Pearson correlation coefficient and R-square of ML-LSTM are higher. In addition, the correlation between features extracted by our approach and gasoline yield is calculated and analyzed to verify the model performance. The further study will focus

on explaining the impact process of various factors on product yields in the time domain.

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

This work is supported by the Science and Technology Major Project of Sichuan Province, China (Grant No. 2019ZDZX0006), Shanghai Natural Science Foundation, China (No. 20ZR1413200) and the Science and Technology Development Project of SINOPEC, China (No. 320131-3).

References

- Deng, Y., Jiang, Q., Cao, Z., 2011. ANN modeling of industrial FCC unit. *Comput. Appl. Chem.* 09, 64–67.
- Gao, Y., Xing, Y., Fu, J., et al., 2017. The application of BP neural network based on a particle swarm optimization to the catalytic cracking reaction-regeneration process. *Comput. Appl. Chem.* 34, 68–72.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Jiang, H., Tang, J., Ouyang, F., 2015. A new method for the prediction of the gasoline yield of the MIP process. *Pet. Sci. Technol.* 33, 1713–1720.
- Ke W., Huang D., Yang F., et al (2017) Soft sensor development and applications based on LSTM in deep neural networks. 2017 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2017.
- Lu, C., Fan, Y., Liu, M., 2018. Advances in key equipment technologies of reaction system in RFCC unit. *Acta Pet. Sin. (Pet. Process. Sect.)* 34, 441–454.
- Ma, F., Li, Z., Weng, H., 2001. A model of deep catalytic cracking product yields using artificial neural networks. *Pet. Refin. Eng.* 02, 35–38.
- Ouyang, F., Liu, Y., 2017. Prediction of the product yield from catalytic cracking process by lumped kinetic model combined with neural network. *Petrochem. Technol.* 46, 9–16.
- Salvado, F.C., Teixeira-Dias, F., Walley, S.M., 2017. A review on the strain rate dependency of the dynamic viscoplastic response of FCC metals. *Prog. Mater. Sci.* 88, 186–231.
- Souza, N.L., Tkach, E., Morgado Jr., E., Krambrock, K., 2018. Vanadium poisoning of fcc catalysts: a quantitative analysis

- of impregnated and real equilibrium catalysts. *Appl. Catal. A Gen.* 560, 206–214.
- Su, X., Pei, H., Wu, Y., 2016a. Predicting coke yield of FCC unit using genetic algorithm optimized BP neural network. *Chem. Ind. Eng. Prog.* 35, 389–396.
- Su, X., Yu, Y., Pei, H., et al., 2016b. Prediction of coke yield of FCC unit using different artificial neural network models. *China Pet. Process. Petrochem. Technol.* 18, 102–109.
- Tian G., Yuan Y., Liu Y. (2019) Audio2Face: Generating Speech/Face Animation from Single Audio with Attention-Based Bidirectional LSTM Networks. (<https://doi.org/10.48550/arXiv.1905.11142>).
- Wang, G., 2015. Application of neural network technology in prediction models for the yield of hydrocracking products. *Petrochem. Technol. Appl.* 04, 73–77.
- Wang, K., Gopaluni, B., Chen, J., et al., 2018. Deep learning of complex batch process data and its application on quality prediction. *IEEE Trans. Ind. Inform.* 16, 7233–7242.
- Yang, C., Chen, X., Li, C., 2017. Challenges and opportunities of fluid catalytic cracking technology. *J. China Univ. Pet. (Ed. Nat. Sci.)* 41, 171–177.
- Yang, F., Dai, C., Xuan, J., 2020. A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance. *Chem. Eng. Res. Des.* 155, 202–210.
- Zhang X., Zou Y., Li S., et al (2018) Product yields forecasting for FCCU via deep Bi-directional LSTM network. *Proceedings of the 37th Chinese Control Conference, IEEE*, 2018.
- Zhang, X., Zou, Y., Li, S., et al., 2019. A weighted auto regressive LSTM based approach for chemical processes modeling. *Neuro Comput.* 367, 64–74 64–74.
- Zhang Y., Yang J. (2018) Chinese NER Using Lattice LSTM. (<https://arxiv.org/abs/1805.02023v4>).