

Received May 12, 2020, accepted May 23, 2020, date of publication May 29, 2020, date of current version June 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2998515

Smart Online Fuel Sulfur Prediction in Diesel Hydrodesulfurization Process

JINGKAI MA^{ID}, NAN LI^{ID}, YATING HUANG^{ID}, AND YUE MA^{ID}

College of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

Corresponding author: Nan Li (linan@th.btbu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61877002 and Grant 51405005, and in part by the National Key Research and Development Project of China under Grant 2019YFC1605306.

ABSTRACT In the process industry, online operators usually adjust the entire production process in time according to changes in certain key indicators. However, it is difficult to directly obtain the value of key indicators due to the complex structure and the numerous processes of the process industry. This paper proposes a method to achieve high precision on-line prediction of key indicators in the process industry. The method first performs data preprocessing of multi-source heterogeneous time series data involved in industrial processes based on the professional knowledge, which not only keeps the prediction error within 10%, but also reduces the prediction time. Then a model framework is constructed based on LSTM neural network and the error correction algorithm is proposed to improve prediction accuracy based on real-time error, which directly causes the error to drop by 3%-5%. At the end, a set of multi-mode online training strategies and related trigger conditions are designed to perform predicting online. Diesel hydrodesulfurization is a typical case in the process industry. The effectiveness of the proposed method is empirically studied by applying its actual data sets. Through the comparison with other traditional well-known forecasting models, and models optimized by adjusting parameters, the experimental results demonstrate that the method can achieve the great prediction performance in terms of both accuracy and stability.

INDEX TERMS Fuel sulfur, LSTM, online predicting, sensor.

I. INTRODUCTION

Process industry is an industry in which raw materials are separated, mixed, formed or changed by physical or chemical changes to increase their value. Process industry is an important part of manufacturing industry and plays an important role in the improvement of national economy and the development of social, its production process is usually continuous or batch. However, at present, process industry is generally faced with the process time-consuming, complex structure, various procedures and many other prominent problems. Therefore, it is urgent to improve the automation level of process industry, which is of great significance to reduce resource consumption and improve practical benefits [1].

The process industry usually involves many process indicators or variables with important reference values. Operators often judge the changes of these key process variables or

indicators based on experience to adjust the entire production process timely and ultimately improve production efficiency, such as the temperature of molten iron and the content of silicon. The production process indicators are roughly divided into two categories. One is often obtained directly based on sensors and other equipment, such as temperature and pressure. These indicators are collectively referred to as production process operating parameters. However, a large number of process variables included in this type of indicator are usually multi-source time series data with time offsets between each other and different time series, which makes our application of these data very difficult. The other is the defined production process indicators, such as the real-time sulfur content in the device during the hydrodesulfurization of diesel. These indicators are often difficult to measure directly or cannot be monitored at all because they are limited by the actual conditions in the industrial process. However, this kind of production process indicators usually has a complex non-linear relationship with other process variables, so it is

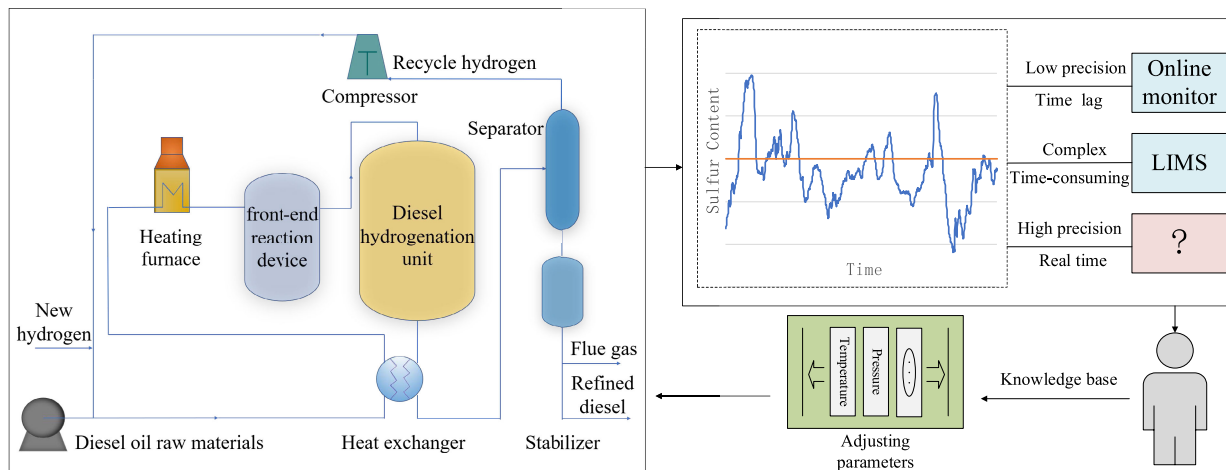


FIGURE 1. Flow chart of diesel hydrodesulfurization.

particularly important to establish an effective mathematical model for real-time prediction. On the other hand, from the perspective of real-time regulation of industrial process trends, in terms of operators, they hope to grasp the future change trends of certain key process variables in advance, which also makes the index trend prediction an important task of the current intelligent production process in the process industry.

The traditional production process index prediction uses mechanism-based modeling method, which is based on the internal mechanism of the process. It used to apply some known laws to establish an accurate mathematical model, such as dynamic principles, material balance equations, energy balance equations, etc. However, the mechanism modeling depends heavily on the cognition of process mechanism. Due to the characteristics of non-equilibrium, non-stability and non-linearity in industrial production process, this kind of mechanism model has high cost and difficulty and its accuracy and reliability is difficult to guarantee. Therefore, there are always problems such as low model accuracy and easy model mismatch [2]. With the popularization of computer applications and the arrival of the era of big data, enterprises have obtained a large amount of industrial big data based on process monitoring and data acquisition systems. Under the situation of continuous integration of information technology and industrialization, these rich data and information materials are combined with machine learning technology to promote the development of industrial enterprises towards intelligence [3]. At present, the most popular method for predicting production process indicators is based on the data-based modeling method, which usually collects the key variable values obtained in the production process, and the relationship between input and output variables is established by machine learning methods to complete the prediction task. This type of method does not need to study the mechanism information of the production process.

Diesel hydrodesulfurization is a typical case of process industry, its device is shown in Fig.1. Because the diesel hydrogenation reaction device is affected by the production environments with high temperature and high pressure, its internal sulfur content cannot be directly measured. A commonly used method now is through the monitor at the end of the outlet to estimate the sulfur content in the diesel hydrogenation reaction device, due to the effect of actual distance factors, the data collected by the monitor is not the real-time sulfur content in the diesel hydrogenation device, but the sulfur content some time ago in the diesel hydrogenation device. Another method is taking a sample at the outlet and then performing expert analysis in a laboratory environment to obtain the sulfur content in the diesel hydrogenation reactor. Although the results of this method are very accurate, its operation is complex and time-consuming. The method is not suitable for real-time prediction of sulfur content in a diesel hydrogenation reactor in a production line. Through the analysis of these existing measuring methods of sulfur content, we cannot predict sulfur content with high precision in real time.

In order to realize timely and accurate closed-loop control of diesel hydrodesulfurization process to stabilize the production quality of fuel, this paper designed an online sulfur content prediction method based on LSTM, which greatly improved the timeliness and accuracy of process regulation. However, due to the existence of real-time errors and unknown changes in the production environment, the prediction process of sulfur content is always unstable. In order to improve the accuracy and robustness of the prediction method, we designed an online correction strategy that can correct the predicted results of the model online based on real-time errors. In addition, equipment damage and process upgrading are inevitable production links in the process industry, and a single prediction model is not enough to cope with complex production conditions. Therefore, we designed a set of online training strategies combined with offline

prediction to improve the automation level of the prediction method.

The whole structure of this paper is given below. Firstly, Section1 and Section2 are the introduction containing the description of background on the prediction of key physical quantities in the process industry and related researches on this field. Then, Section3 introduces the data preprocessing algorithm based on multi-source heterogeneous time series data. Section4 introduces the components of the prediction model, including the model framework and error correction algorithm. Next, the online training strategies and related trigger conditions are included in Section5. Section6 is the experimental simulation and discussion. Finally, Section7 includes conclusion that summarize the ideas in the whole paper.

II. RELATED WORK

Generally, the key technologies for real-time prediction of production process indicators based on data include three aspects, namely feature extraction; the establishment of prediction models; and the construction of online learning systems. The feature extraction process refers to a large number of candidate input features are effectively calculated to calculate some equivalent features or to select the features most relevant to the prediction index as input variables of the prediction model. Data-based predictive modeling can be completed using machine learning methods. Unlike mechanism models, such methods only focus on the input and output of the model. The input is the selected relevant feature variable and the output is the key indicator to be predicted [4]. The construction of the online learning system will realize the online training of the prediction model and the real-time prediction of key indicators by combining with data collected on site.

A. FEATURE EXTRACTION

In the process industry, the predicted index value often has a complex non-linear relationship with multiple process variables. To predict the production process index, we need to select some of the most effective features from the many original variables to reduce the data set dimension. This is an important method of improving learning algorithm performance and is also a key step in data preprocessing. Whether the feature selection based on the data is independent of subsequent modeling algorithms can be divided into two types: filter and wrapper [5]. The basics idea of filter feature selection is to define an evaluation criterion in advance to determine the degree of correlation between the process variable and the predicted index, and then select a series of highly relevant process variables to participate in the modeling. The method of wrapper combines the feature selection steps with the modeling algorithms. Although the method will improve the prediction accuracy, the time cost will be relatively high. At present, the most common feature selection method applied in industrial production processes is the filter methods, such as correlation-based

analysis methods [6]–[8]; wrapper methods, such as variable pruning methods [9], [10], genetic algorithm-based methods [11].

B. THE ESTABLISHMENT OF PREDICTION MODELS

With the further development of machine learning technology, its application in the field of engineering is more common. In various types of artificial neural networks, radial basis functions (RBF) [12], [13], least squares support vector machine (LSSVM) [14]–[16], convolutional neural network (CNN) [17], [18], long short-term memory networks (LSTM) and other models have attracted the attention of many researchers. Long short-term memory networks (LSTM) are better than recurrent neural networks (RNN). It inherits most of the characteristics of the RNN models and solves the problem of gradient explosion and gradient disappeared during gradient back propagation. Specific to engineering tasks, the LSTM model is very suitable for dealing with problems that are highly related to time series, and it has become a very hot research model in the deep learning framework. In the process industry, due to the existence of many aspects of chemical transformation and different catalyst deactivation mechanisms in the reaction system, coupled with the lack of comprehensive understanding of the process, it is difficult to fully grasp the process. Machine learning technology is used to help establish prediction and optimization models and achieve efficient quality control and more accurate monitoring processes. This technology has triggered wide concern in the industry and academia, which not only can be used to understand the importance of process factors, but also can apply historical data to predict the future. At the same time, it saves a lot of time and energy and reduces the empirical error.

For some data-driven systems, a variety of sensor data is used as input to select and extract features that characterize the state of the system. Sensor data is essentially data with a certain time sequence, which is sampled by the sensor and represented in sequential form. Previous research mainly focused on multi-domain feature extraction, including statistics (variance, skewness, kurtosis), frequency (spectral skewness), and time-frequency (wavelet coefficient) features. However, these methods cannot model the intrinsic sequence characteristics behind the sensor data [19]. These models require a lot of expert knowledge or feature engineering. In addition to these methods based on artificial engineering features, some sequences models including Markov models, Kalman filters, and conditional random domains, which have powerful capabilities only for accessing the sequence data of the original time series [20], [21]. However, they have been unable to capture long-term dependencies. In recent years, recurrent neural network (RNN) and its derivative long short-term memory network (LSTM) [22], [23] and gated recurrent neural network [24] show great advantages in terms of sequence prediction tasks. Among them, the network of long short-term memory has achieved good results in

petrochemical field [25], transportation field [26], [27] and medical field [28], [29]. They are widely used in speech recognition, caption generation, machine translation, image and audio classification. Recurrent neural networks have proven to be superior to convolutional neural networks in processing data that is tightly connected [30]. However, the problem of the disappearance or explosion of the error gradient during the model back-propagation directly affects the performance of the neural network. This means that RNN cannot capture the long-term dependence of the data. The LSTM model makes up for the shortcomings in this area, it is capable of capturing long-term dependencies and nonlinear dynamics in time series data.

C. THE CONSTRUCTION OF ONLINE LEARNING SYSTEM

The multivariate statistical process control (MSPC) method has proven to be an effective tool for process monitoring, modeling, and fault detection. This method has achieved a lot in real-time monitoring and on-line modeling of continuous multi-scale operation process in factory operation [31]–[33]. However, the traditional method is based on assumption that process variables are independently sampled, Gaussian distribution and linear correlation. In industrial production, the process variables are always non-linear due to changes in actual operating conditions or in parameters setting. In addition, some variables may not obey the Gaussian distribution. A method based on kernel-PCA (KPCA) is proposed by [34]. Compared with other non-linear principal component analysis methods, the main advantage of the KPCA is that it does not require non-linear optimization. A new principal component non-linear measurement method is proposed by [35] and it also discussed the criteria for selecting linear or non-linear principal component analysis for a specific process. Because MSPC is not efficient in multi-mode processes, a monitoring strategy is proposed by [36], which updates the monitoring model by establishing a recursive PCA model. In order to avoid fault adaptation, Kruger *et al.* [37] proposed a fault monitoring strategy and also introduced a method based on model library.

On-line modeling based on current data is a key link for real-time monitoring in the process industry. Traditional modeling methods such as neural networks, fuzzy set methods and methods based on parametric model all rely on pre-collected data sets. It is difficult to solve the time-dependent model structure determination and parameter optimization during the actual industrial problem [38]. However, the idea of local modeling is to set a non-linear model within a limited range of data to predict the key physical quantities of the local area. The famous models based on the idea of local modeling include neural fuzzy networks and T-S fuzzy models. However, the difficulty of local modeling lies in the need to use prior knowledge to determine a certain operating area. Once the prior knowledge or experience is insufficient, a complex training strategy is needed to determine the optimal structure and parameters of the local model.

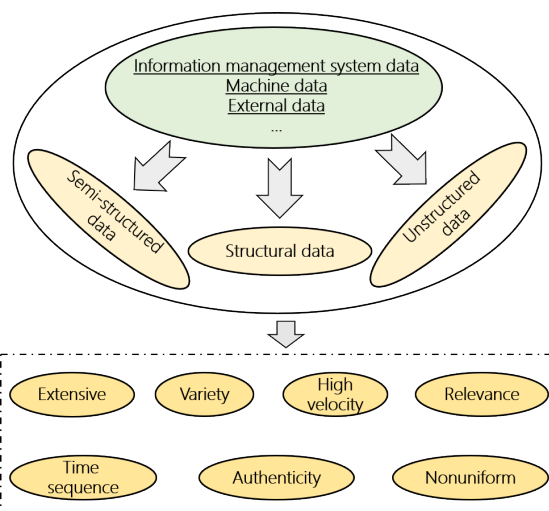


FIGURE 2. Multi-source time series data fusion.

III. TIME SERIES DATA

A. CHARACTERISTICS AND COMPLEXITY

The process industry has a complex structure and a variety of processes. The entire process usually runs at different levels, in which a large number of key process indicators are involved. These important variables have important reference values for the control of the entire production process. Multi-time series data fusion is a significant feature of the process industry.

The data of the process industry has many characteristics, as shown in the Fig.2. It is mainly reflected in (1) extensive, the process industry usually has the characteristics of long sampling time, high rate, many machines and equipment, high information density, and large data storage. (2) variety, there are many types of data in the process industry, and data acquisition is diverse, usually including information management system data, machine equipment data, external data, etc., and its storage methods are complex and diverse, including structured data, semi-structured data, unstructured data, etc. In addition, irregular sampling of data is also one of the reasons for the diversity of data. (3) high velocity, the process industry usually needs online modeling and real-time updating, which requires that the data acquisition rate should meet certain requirements. (4) nonuniform, in the data modeling process of process industry, the value density of data is often considered, that is the proportion of valid data in a batch of data. However, the value density of data in the process industry is usually uneven. (5) authenticity, in the process industry, due to the interference of abnormal factors such as the failure of monitoring instruments or equipment, the sampling data may contain unreal data or data missing, which are the problems to be solved in the process of data call. (6) time sequence, the data collected in the process industry has a certain time series and has the characteristics of high dimensional and dynamic sampling. (7) relevance, the data collected at the same stage in the process industry has strong

relevance, such as the status of machinery and equipment at a certain stage, etc. In addition, the data between different links in the product life cycle also have a certain correlation.

In the case of diesel hydrodesulfurization, multiple sources of time series data are involved in the process of predicting the sulfur content in the diesel hydrodesulfurization device. The data sources are divided into two categories, one type is the sensor data with a certain time series in the diesel hydrodesulfurization device and the front-end reaction device, and the other type is the sulfur content value in the diesel hydrodesulfurization device with time series obtained through multiple schemes. Due to the limitation of actual distance factor between the front-end reaction device and hydrogenation reaction device, it takes some time for the change of the front-end reaction device to have some effect on the hydrogenation reaction device. Therefore, when the data values of two types of sensors in the front-end reaction device and the hydrogenation reaction device are applied to represent the sulfur content in the hydrogenation reaction device, the time series of the two types of data should have a certain time lag when they correspond to each other. The sulfur content value obtained by different scheme is divided into three categories. One is the value obtained by the online monitor called YSYL. This scheme uses the monitor at the outlet to estimate the sulfur content in the hydrogenation reaction device. Also due to the limitation of distance, the data collected by the monitor represents the sulfur content in the hydrogenation reaction device some time ago. Another type is laboratory testing value called LIMS, this solution obtains a very accurate sulfur content in the hydrogenation reaction device by sampling at the outlet and then performing expert analysis in a laboratory environment, but this solution usually requires a lot of analysis work and it can only obtain a small number of accurate values. It is not suitable for the prediction of the real-time sulfur content in the diesel hydrogenation reaction device in the production line. The last category is the real-time prediction value called JZCYS. In this scheme, the sulfur content in the hydrogenation reactor can be predicted in real time by the model designed by us.

B. DATA ALIGNMENT AND DATA FILTERING

In the process industry, the prediction of certain key physical quantities based on multi-source time series data often involves a series of related variables. Because the process industry has the characteristics of continuous and batch processing, the effects of the values of related variables collected at the same time on the key physical quantities usually have a certain time offset. In addition, due to the characteristics of the machine or the equipment failure, the corresponding time series of the monitoring values obtained by each variable may be different. We use data alignment algorithm to process this kind of multi-source time-series data with time offset and different time series. After the original data is data aligned, the values of each variable are evenly distributed on the same time series. Then multivariate values are filtered to achieve dimensionality reduction. At the same time, data filtering

can effectively extract features based on expert knowledge, thereby improving the prediction accuracy of the model and speeding up the calculation of the model. The results of data alignment algorithm and data filtering algorithm are shown in the Fig.3.

Some concepts in data alignment and data filtering algorithms are defined as follows:

Definition 1: The base timestamp is defined as BT that is the end time of the sampling. If base timestamp is not specified, the current timestamp of the system is used as the base timestamp.

Definition 2: The collection of time offsets on the base timestamp is defined as T , $T = \{T_1, T_2, \dots, T_j\}$, j is the j -th time offset on the base timestamp in the T . When $j = 0$, it represents the time offset is 0.

Definition 3: A collection of related variables with a certain time offset j is defined as $S_j = \{S_{j1}, S_{j2}, \dots, S_{ji}\}$, S_{ji} represents the i -th related variable with time offset j .

Definition 4: The number of variables of a certain collection S_j with time offset j is defined as SN_j .

Definition 5: The sampling time is defined as S_r . The sampling time step is defined as S_p .

Definition 6: The sampling range of the time series is defined as T'_n , it is expressed as an ordered collection, $T'_n = \{0, 1, 2, 3, \dots, m\}$. T'_{nm} represents the m -th element in the set T'_n , its meaning is to generate a time series with m timestamps at equal time intervals in the sampling time.

Definition 7: Some kind of related variable S_j with time offset j , its end timestamp and start timestamp are defined as ET_j and ST_j , they are obtained according to the following formula:

$$ET_j = BT + j \quad (1)$$

$$ST_j = ET_j - S_r \quad (2)$$

Then according to T'_n , the sampling minimum timestamp and the sampling maximum timestamp of the certain type of variable are defined as $MinT_j$ and $MaxT_j$, they are calculated according to the following formula:

$$MinT_j = ST_j + T'_{nmin} \times S_p \quad (3)$$

$$MaxT_j = ST_j + T'_{nmax} \times S_p \quad (4)$$

where: T'_{nmin} , T'_{nmax} are the minimum and maximum values in T'_n .

According to the calculated $MinT_j$, $MaxT_j$, ET_j , and ST_j , the minimum and maximum timestamps are selected as the start and end timestamps of the relevant variable samples.

Definition 8: The time series to be calculated is defined as TS_j , it is expressed as an ordered collection whose elements have equal time interval, the time interval is equivalent to S_p , $TS_j = \{t1, t2, t3, \dots, tm\}$. Among them, the m -th timestamp in the set TS_j called TS_{jm} is obtained according to the following formula:

$$TS_{jm} = ST_j + T'_{nm} \times S_p \quad (5)$$

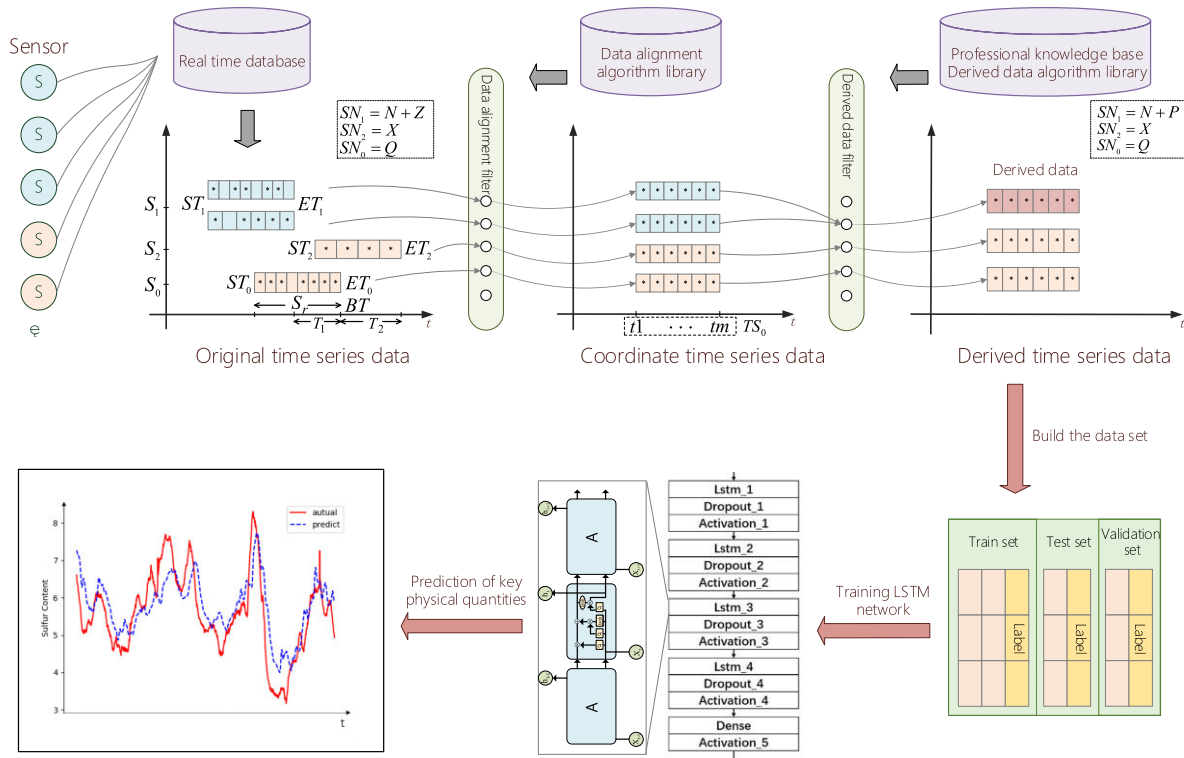


FIGURE 3. Schematic diagram of key physical quantity prediction in process industry.

In the data alignment algorithm, the related variable classes with different time offsets are all uniformly aligned according to the time series TS_0 .

Definition 9: t_n is the required calculation timestamp, V_n is the corresponding value of the timestamp. If t_n is within the time range of the sampled data and there is a sampling value corresponding to the current timestamp, the sampling value is V_m . If there is no corresponding value at that moment, define t_s as the timestamp after t_n , t_p as the timestamp after t_s , t_q as the timestamp before t_s , V_p as the value of t_p , and V_q as the value of t_q ; if t_n is not in the time range of the sampled data, t_r is defined as the timestamp closest to t_n in the time series of the sampled data, and V_r is the value corresponding to t_r .

The detailed steps of data alignment algorithm are described as follows.

Step 1. According to BT , T , S_j , S_r , S_p , and T'_n , the ST_j , ET_j , and TS_j are calculated.

Step 2. Obtain the sample data of each variable. The time interval is $ST_j < t < ET_j$. The original data is used to calculate data that corresponds to the time series calculated, and finally all the data have been calculated are unified under the TS_0 .

Step 3. To calculate the value V_n correspond to the required timestamp t_n , if t_n is not within the time range of the sampled data, $V_n = V_r$; if t_n is within the time range of the sampled data and there is corresponding value at that moment, $V_n = V_m$; if t_n is within the time range of the sampled data

but there is no corresponding value at that moment, V_n could be obtained according to the following formula:

$$V = \frac{\left[\left(\frac{1}{t_p - t_n}\right) \times V_p\right] + \left[\left(\frac{1}{t_q - t_n}\right) \times V_q\right]}{\left(\frac{1}{t_p - t_n}\right) + \left(\frac{1}{t_q - t_n}\right)} \quad (6)$$

After the original data is aligned, the values of each measurement point are evenly distributed on the same time series. According to expert knowledge, a part of the measurement point data collected in the device on the production line can be better characterized by effective calculation. It can also improve the prediction accuracy of the model and speed up the calculation speed of the model. This process is called data filtering, the result is in the Fig.3.

The detailed steps of data filtering algorithm are described as follows.

Step 1. Ratio discrimination of abnormal values. That is the sensor data column of a measurement point contains too many abnormal values, which means it is less likely to contain valuable information. Therefore, the feature points that the ratio of abnormal values in the data column is greater than a certain threshold are removed.

Step 2. Low variance filtering. That is the little change in the value of a sensor data column. It also doesn't contain much valuable information. Therefore, the positions with low variance of their data column are removed.

Step 3. High correlation filtering. In other words, if the variation trend of sensor data of two measurement points is

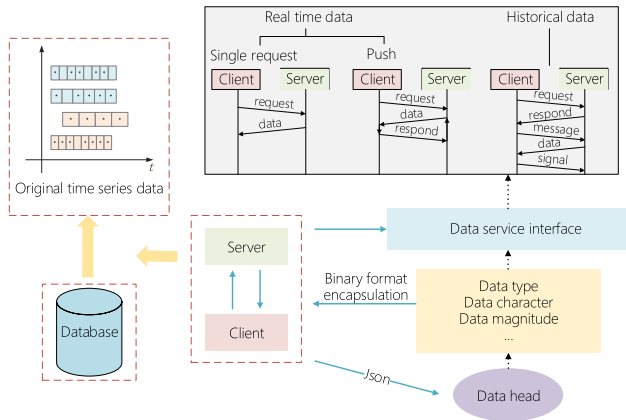


FIGURE 4. Schematic diagram of data publishing service.

similar, it is believed that the information contained in them is similar. The similarity is represented by the correlation coefficient of the two columns of data, while the correlation coefficient is greater than a certain threshold, only one column is reserved for the input of the machine learning model.

Step 4. Data point simplification. Its algorithms usually include accumulation, squaring, weighting, averaging, or a combination of several of them. The values of Z measurement points on the same time series are converted into P derivative values after data point simplification to participate in the training and testing of the model.

C. DATA PUBLISHING SERVICE

In order to unify the way of data call and improve the level of data application and management, the data service interface is designed and developed based on Socket communication framework, the schematic is drawn in Fig.4. When the client side interacts with the server side for information, the server side recognizes the data request pattern through the data request type, nature, magnitude and other aspects, and invokes the matching data service interface for data invocation. Currently, there are three types of request modes: historical data requests, real-time data single requests, and real-time data push requests. For the client, historical data requests and real-time data single requests are short connections. The client will actively interrupt the connection after obtaining the required data. The real-time data push request is a long connection for the client. The client will continuously receive real-time data until the client feedback indicates that it stops receiving data, the client will end the link by itself. Regardless of the form of data request, the data header of the first request sent by the client is uniformly encoded in JSON format. For different types of data requests, their binary packaging format design is also different. For example, in the complete historical data request, the server side and the client side need to conduct 5 data interactions. However, in the single query request of real-time data, the server side and the client side only need to conduct two data interactions.

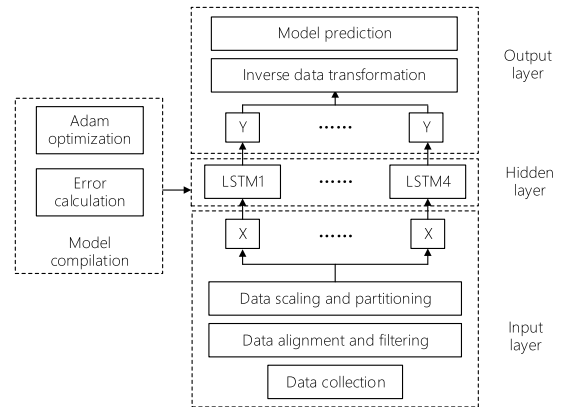


FIGURE 5. The mode framework of Sulfur-LSTM.

IV. PREDICTION MODELS BASED ON LSTM

A. MODEL FRAMEWORK

The model we built can be divided into three parts: input layer, hidden layer, and output layer. The input layer is mainly used for data preprocessing and dataset division of the multi-source heterogeneous timing data. The data can be normalized and scaled between 0 and 1 after passing through the input layer. The hidden layer is trained based on the training set. The setting of batch size and the number of network layers affects the learning ability and test time of the model, mean square error is used in training the LSTM model. Through the Adam optimizer, the weight of the LSTM structural unit is optimized, and the optimal parameter combination is obtained by continuously optimizing the batch size parameter of the network layer. The Linear activation function is used to improve the computing power and Dropout is added to prevent overfitting. The output layer predicts the data according to the model learned in the hidden layer and performs inverse data transformation. The model framework is shown in Fig.5.

B. CONSTRUCTION OF THE DATA SET

1) DIFFERENT FEATURES

Different data filtering algorithms will directly lead to changes in the input features of the model, which will have a significant impact on the final output of the model. In the case of diesel hydrodesulfurization, we use five different schemes to construct the dataset to find the best feature combination. The data sources of the diesel hydrodesulfurization production line come from the real-time values of the temperature sensors in the front-end reaction device and the hydrogenation reaction device as well as the real-time sulfur content value monitored by the online monitor at the outlet. The Table 1 defines different data sources.

Since the problem we need to solve is to use the real-time temperature values of the various positions in the front-end reaction device and the hydrogenation reaction device to predict the sulfur content in the hydrogenation reaction device at that time, then which of the measured temperature values can

TABLE 1. Definition of various data sources in diesel hydrodesulfurization production line.

Symbol	Definition
A	Values of each sensor in the front-end reaction device at time t
B	Values of each sensor in the hydrogenation reaction device at time t
C	Values of online monitor at time t
D	The value of B through data filtering

TABLE 2. Data sets of different features.

Data set	Combination	Train shape	Test shape
D1	B+C	(2000,7,25)	(1000,7,25)
D2	A+B+C	(2000,7,37)	(1000,7,37)
D3	A+B	(2000,7,36)	(1000,7,36)
D4	A+C+D	(2000,7,22)	(1000,7,22)
D5	A+D	(2000,7,21)	(1000,7,21)

TABLE 3. Data sets of different time lags.

Data set	Combination	Train shape	Test shape
D6	A+C+D	(2000,5,22)	(1000,5,22)
D7	A+C+D	(2000,6,22)	(1000,6,22)
D8	A+C+D	(2000,7,22)	(1000,7,22)
D9	A+C+D	(2000,8,22)	(1000,8,22)
D10	A+C+D	(2000,9,22)	(1000,9,22)

be better characterized the sulfur content in the hydrogenation reaction device at that time? In order to solve this problem, we proposed 5 different data processing schemes, the data collected were respectively made into data sets D1, D2, D3, D4, D5. The method of feature combinations of the five data sets, sample number and input shape are shown in the Table 2.

2) DIFFERENT TIME LAGS

When the data set is made into a supervised learning data set, in order to determine the optimal time lags, we have made 5 kinds of data sets with steps of 4, 5, 6, 7, 8 of time lags based on expert knowledge, named in sequence D6, D7, D8, D9, D10. The feature combination methods of five data sets, the number of samples and the input shape are shown in the Table 3.

C. THE MODIFIED ALGORITHM

Due to the influence of various uncertain factors, there is always an error between the true value and the predicted value in the prediction process. In order to improve the prediction accuracy, we propose a correction strategy based on real-time errors, which can be used to continuously modify the output of the model. The algorithm flow is shown in the Fig.6. Here are some concepts in the modified algorithm:

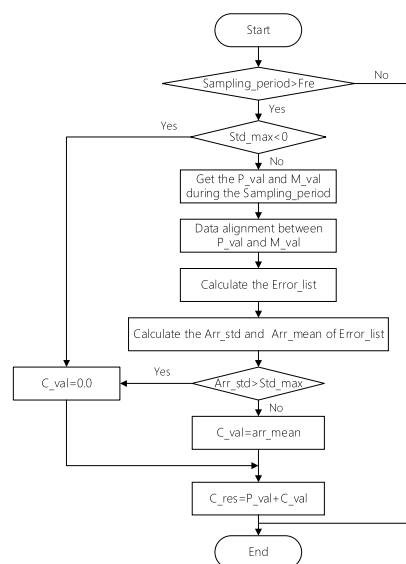


FIGURE 6. Flow chart of modified algorithm.

Definition 10: The sampling period is defined as *Sampling_period*. The minimum value of the sampling period is defined as *Fre*. If *Fres* is the larger sampling frequency of predicted and real values, $Fre \geq Fres$.

In this period, the real value is defined as *M_val* and the predicted value is defined as *P_val*, the correction value is defined as *C_val* and the correction result is defined as *C_res*.

Definition 11: The maximum standard deviation is defined as *Std_max*. It is used to determine the reasonableness of the error.

Definition 12: The error list is defined as *Error_list*, this list is formed by calibrating the measured value and the predicted value in the current sampling period to obtain the error.

Definition 13: The Standard deviation of the current error list is defined as *Arr_std*. And the mean value of the current error list is defined as *Arr_mean*.

Definition 14: The cycling time of correction service is defined as *Cycle_time*.

When the correction service is started, *C_val* is calculated for each batch of data. In the prediction of the next batch of data, the *C_res* is equal to the superposition of the *P_val* and the *C_val*. The detailed steps of correction algorithm are described as follows.

Step 1. Gets all predicted and true values during the sampling period and align the data to obtain one-to-one corresponding predicted values and true values on the same time series.

Step 2. Calculate the *Error_list* and then the standard deviation is calculated as the criterion to judge the degree of error.

Step 3. If the *Arr_std* is greater than *Std_max*, the error is unstable, $C_val = 0$. Otherwise, $C_val = Arr_mean$. If $Std_max < 0$, $C_val = Arr_mean$.

After experimental statistics, the shorter the correction service cycle, the better the result, but it is necessary to

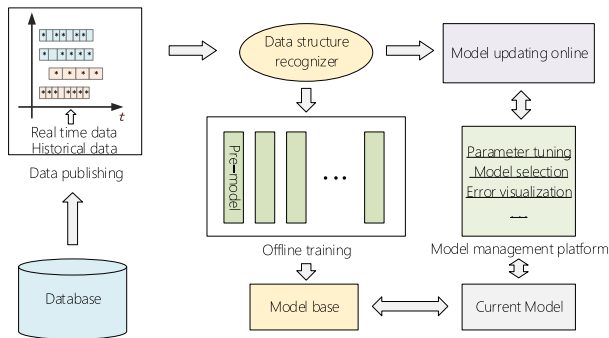


FIGURE 7. Tuning of pre-trained models.

consider the limitations of practical factors, such as sampling frequency for key indicator in the process industry.

V. ONLINE LEARNING STRATEGIES
A. TUNING OF PRE-TRAINED MODELS

The data obtained on the production line is preprocessed and stored in the statistical database of the training server. The client can perform offline model training and testing by calling the data in the statistical database and can also implement online training. There are two modes of online training strategies based on whether the data structure has changed.

When the data source is unstable and the input data structure of the model changes, a semi-automatic strategy is adopted, it is shown in the Fig.7. The client trains the model offline by calling historical data in the database to obtain a pre-model. Then the trained pre-model file is transmitted to the model base for called by the client. During the using of the pre-model, the model is continuously optimized to increase the prediction accuracy by combining online data and using a full-automatic tuning strategy.

In the training process of the network, the data must be normalized firstly. Due to the different data properties and the big difference in order of magnitude, in order to avoid the large prediction error caused by them, it is necessary to normalize the data and convert the data into values between 0 and 1. The mean absolute percentage error and the root mean square error are used as the criteria for evaluating the model.

B. MODEL UPDATING ONLINE

When the data source is stable and the input data structure of the model is unchanged, a full-automatic strategy shown in the Fig.8 is adopted. As long as the time reaches the *Cycle_time*, all data in the sampling period prior to the current moment is called. The trigger conditions of model updating are divided into two modes. The first mode is to determine the similarity of current batch input data, and the second mode is to compare the trend of the predicted value and the measured value. The first mode is in the priority sequence.

The decision rule of mode one is as follows: determine the similarity between the input data and the training data of current model. If the similarity factor is greater than the

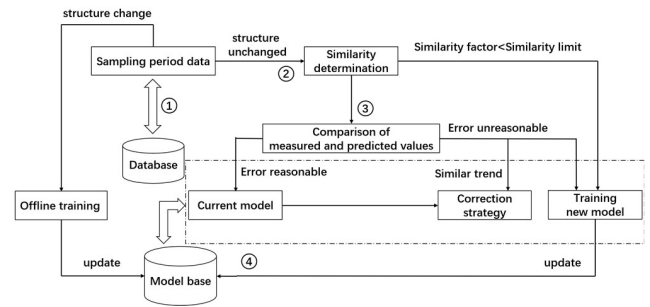


FIGURE 8. Model online updating strategy.

limit of similarity, enter mode 2 for judgment. If the similarity factor is less than the similarity limit, then the batch data is used as training data to retrain the model and the trained model is transferred to the model base for prediction of the next batch of data. The decision rule of mode two is as follows: the trend of the predicted value and measured value of the current batch is performed. In comparison, if the trend of the measured value is the same as the predicted value, but there is a certain deviation, the model’s prediction accuracy is improved by calling the correction service. If the error index calculated from the measured values and predicted values is within the expected range, the current model is still invoked after parameter tuning to predict the next batch of data. If the error index exceeds the expected range, the online training of the new model will be performed by using the current batch data and the trained model files are transferred to the model base. And the new model is called when the next batch of data is predicted. When the new model is built online, the training process status is uploaded to model management platform and the model parameters are continuously optimized. After the training is completed, the model is verified. One defines the desired prediction accuracy as *EA*, $EA = 90\%$. If the verification accuracy is higher than *EA*, the model is saved as a binary file and stored in the model base. If the verification accuracy is not up to standard, the previous model is still called for prediction.

VI. EXPERIMENTS AND DISCUSSION
A. SYSTEM IMPLEMENTATION AND EXPERIMENTAL CONDITIONS

Based on the problem of sulfur content prediction in diesel hydrodesulfurization production line, a data online learning prediction system as shown in the Fig.9 is designed. Server 1 is used as the InfoPlus21 real-time database, which stores real-time monitoring data and historical data. Server 2 is used as the client of the real-time database to communicate with server 1 through IP21Api to obtain real-time data and historical data. Server 2 also serves as a server to provide external data query services, and server 3 acts as a client to establish a connection request with it through a socket to obtain historical data and real-time data, which are encapsulated in a custom binary format. After acquiring the data,

TABLE 4. The experiment condition.

	Data resource		Experimental configuration	Model parameters	Experiments design
	First batch	Second batch			
Feature numbers	37	19	Computer: 3, GPU, i7, 4GRAM High-performance server: 3 SQLite Expert Professional Pycharm, Tensorflow=1.12, Cuda9.0	Batch size=15 4 nerve layers 130 neurons Linear activation function	Parameter tuning comparison experiment: different features different time lags different models different data structure
Sample step	360s				
Data volume	Train set:2000 Test set:1000				

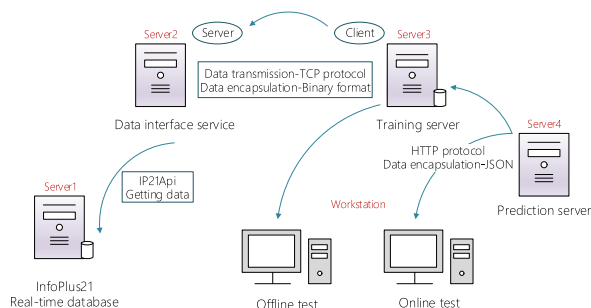


FIGURE 9. Online learning and prediction system.

server 3 performs necessary data cleaning and preprocessing and then carries out machine learning. Server 4 acts as a client to call the trained predictive model from the model library. Server 4 also provides a prediction service server and a web-service based on the HTTP protocol. Users call data through the client, which is encapsulated in binary format. The workstation is the user terminal, which connects to the server 4 through the HTTP protocol, which can realize visual analysis of the data and prediction of important indicators.

In this paper, the whole system and its related algorithms are verified by collecting the actual data of diesel hydrodesulfurization production line, the operating condition of this experiment is showed in Table 4. The data source is divided into two batches of data based on whether the process pattern changes. And the original feature numbers of the two batches of data are 37 and 19. Both batches of data are managed in a unified structured data form. After the data is preprocessed, the data is unified into a data set with a data volume of 3000, the first 2000 are used for training and the last 1000 are used for prediction. In addition, the system model is trained based on a single GPU and Tensorflow, the online tuning of the model is implemented based on the model management platform and multiple models are stored in the model library for invocation.

B. DIFFERENT FEATURES

To measure the effectiveness of different data filtering algorithms, the Mean Absolute Percentage Errors (MAPE) and Root Means Squared Errors (RMSE) are computed. Table 5 demonstrate the prediction performance of different algorithms, the algorithms with the best performance is marked in bold. When the real-time sulfur content monitored by the online monitor is used as input feature, a better prediction result can be observed compared with the one with no

TABLE 5. Comparison of MAPE and RMSE of sulfur content in the model trained by data sets with different features.

Data set	Test-MAPE	Test-RMSE	Time Consume(s)
D1	10.368	0.662	0.792
D2	9.779	0.648	0.772
D3	15.684	1.175	0.741
D4	9.652	0.640	0.750
D5	15.435	1.132	0.748

sulfur content monitored input. In addition, the MAPE and RMSE of D4 are the lowest excluding D3 and D5, which shows that the real-time temperature value of each sensor in the front-end reaction device, the value of the real-time temperature value of each sensor in the hydrogenation reaction device through data filtering and the real-time monitoring value of the online monitor are combined to be input features is effective.

C. DIFFERENT TIME LAGS

To examine the prediction performance in a more intuitive way, the comparison between the actual sulfur content in the test set of the five data sets and the predicted sulfur content by the Sulfur-LSTM model was drawn in Fig.10. We can see from the curve trend in the figure that all the models trained from the five data sets have good fitting effect on the test set. Fig. 11 further presents the prediction performances for different time lags, we can see that D8 data set performs best among all other data sets, but the difference is marginal, which means that the prediction error caused by the change of the time lag of the data set is not great. In addition, we can clearly observe from the histogram that the time required to predict the sulfur content of the test set increases as the time lags become longer. It shows that while changing the time lags of the data set to improve the prediction accuracy, it should also consider the time consumed when a large amount of data is used for prediction.

D. DIFFERENT NETWORK PARAMETERS

1) BATCH SIZES

After setting up the model framework, we performed a series of experimental adjustments to its parameters to find the best results of the model. We studied the impact of the batch size on the model firstly, and the results are shown in the Fig.12.

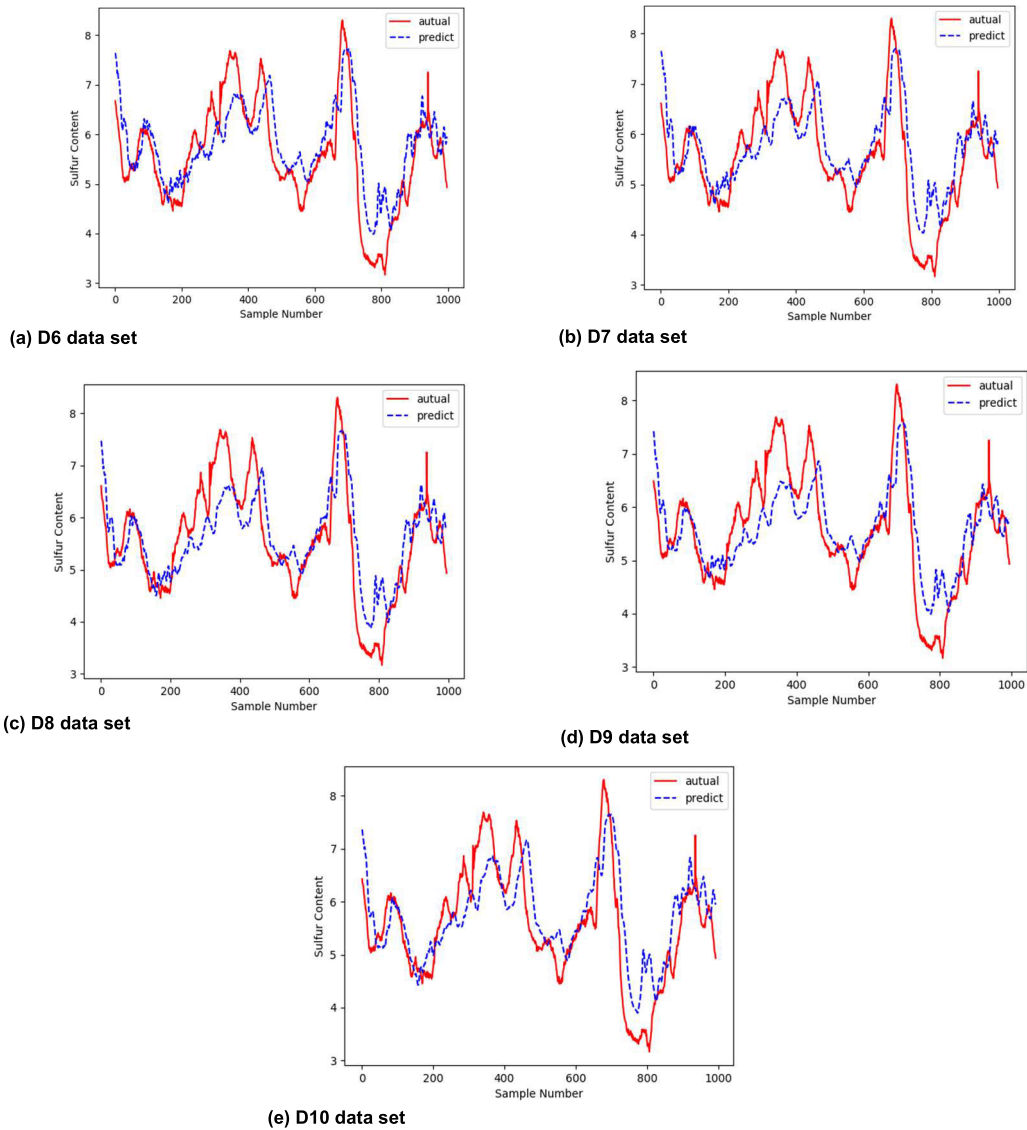


FIGURE 10. Comparison of the predicted values by the model trained in data sets of different time lags and true values about sulfur content.

The experimental results show that when the batch size is too small, the randomness is large and the oscillation is too obvious, and it is difficult for the model to reach convergence. When the batch size is too large, the model takes a long time to train and the direction of gradient descent also basically does not change. Taking into account the prediction accuracy, fitting effect and other factors of the model, when the batch size is 15, the prediction effect of the model is the best.

2) ACTIVATIONS FUNCTIONS

The choice of activation function is critical to the performance of the model. We have studied the effects of four common activation functions on the model, and the results are shown in the figure. From the Fig. 13, we can see that the linear activation function has the best effect in the four sets of experiments we did, and the MAPE and RMSE are the lowest

in the test set. Therefore, all four LSTM layers in our model are activated by the linear activation function. The problem we studied is a regression problem, so the last fully connected layer still uses a linear activation function.

3) THE NUMBER OF HIDDEN LAYER NEURONS

The number of hidden layer units will affect the learning ability of the model, so we designed a series of experiments to find the optimal number, and the results are shown in the figure. We can observe from the Fig.14 that when the number of hidden layer units is less than 130, the MAPE and RMSE show a downward trend with the increase of the number of hidden layer units. When the number of hidden units is equal to 130 or greater than 130, the MAPE and RMSE increase sharply. The results show that the prediction performance is closely related to the complexity

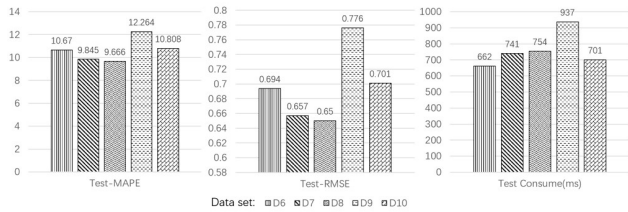


FIGURE 11. Comparison of MAPE and RMSE of sulfur content in the model trained by data sets with different time lags.

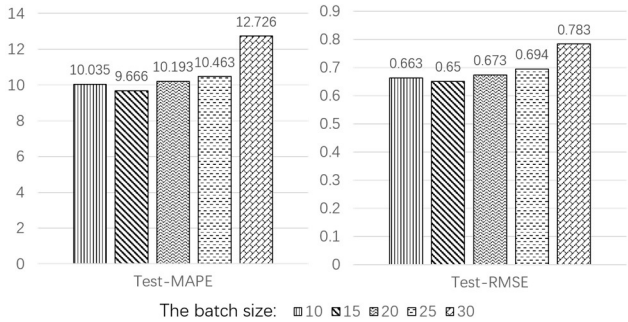


FIGURE 12. Prediction performance of the model with different batch sizes.

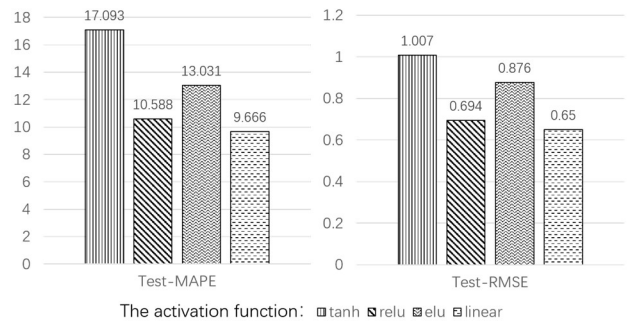


FIGURE 13. Prediction performance of the model with different activation functions.

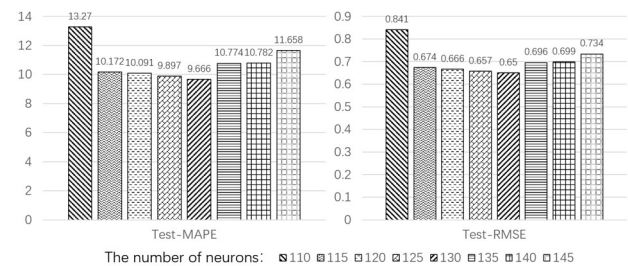


FIGURE 14. The influence of the number of hidden layer neurons on the prediction accuracy of the model.

of the model. When the complexity of the model structure is lower than a certain degree, the prediction results are very poor. When the complexity of the model structure exceeds a reasonable interval, the prediction error increases rapidly.

TABLE 6. Comparison of prediction performance of different models.

	Sulfur-LSTM	SVR	Decision Tree	MLP	KNN
RMSE	0.650	0.678	0.783	0.683	1.386
MAPE	9.666	11.020	11.764	9.621	20.684

E. COMPARISON OF DIFFERENT NETWORK MODELS

We compare the Sulfur-LSTM model with several classic regression models on the same data set and find that the Sulfur-LSTM model has higher prediction accuracy and better generalization ability. In the methods based on artificial intelligence, support vector regression (SVR) is considered to be an effective algorithm. The essence of the algorithm is to map the data into a high-dimensional feature space through a non-linear relationship, and then perform linear regression in the high-dimensional feature space. The support vector regression model of radial basis function (RBF) kernel is used in the experiment. Decision tree models and multilayer perceptron models are classic neural network models that both can solve the problem of non-linear features and also played a comparative role in our experiments. The model based on k-nearest neighbor regression algorithm is also a classical regression model. The algorithm finds the K nearest samples of a sample and assigns the average value of some attributes of these samples to the sample to obtain the corresponding value of the sample. The experimental results are shown in the Table 6. The Sulfur-LSTM model shows good results on the data set. Therefore, this paper mainly studies the application of long short-term memory neural networks in the process industry. And it did not do too many studies on the SVR and other traditional methods.

F. COMPARISON OF TIME SERIES METHODS AND LSTM NETWORK

In the time dimension, a common method for predicting large amounts of data is time series analysis. In the field of transportation, there are precedents for using time series methods to predict traffic accidents, for example, somebody study the trend of traffic accidents by studying the trend changes, periodic changes, seasonal changes and random changes of the time series of traffic accidents. The method of time series analysis is to analyze a series of data with constant time intervals to find long-term trends. Its purpose is to predict and guide real-world problems. Three analysis methods commonly used in time series are moving average models (MA), auto-regressive models (AR) and auto-regressive moving average models (ARMA).

We divide the data of the data set D4 into a training set and a test set, and the test set takes 10%, 15%, 20%, 25%, 30%, 35%, 40%, and 45% of the total data respectively. From the perspective of time series analysis, the ARMA model is established with the same data. This is a classic statistical time series model. Then several data sets of different proportions

TABLE 7. Comparison of ARMA and LSTM.

Ration of test set	MAPE		RMSE	
	LSTM	ARMA	LSTM	ARMA
10%	14.641	13.911	0.789	0.767
15%	12.526	11.960	0.765	0.751
20%	11.249	11.792	0.679	0.709
25%	10.381	10.573	0.697	0.729
30%	9.786	10.956	0.634	0.689
35%	9.714	11.717	0.724	0.713
40%	9.623	9.135	0.725	0.684
45%	9.212	9.322	0.716	0.712

are applied to the time series model and the Sulfur-LSTM model. In order to better demonstrate the prediction effect of the two methods, we still choose mean absolute percentage error (MAPE) and root mean square error (RMSE) as the evaluation criterion. We can see from the Table 7 that the prediction accuracy of the two models is relatively high and stable under the same partition of training set. When the number of test sets is small, time series models show greater advantages. But it is not difficult to see that the MAPE and RMSE values of the LSTM model continue to fall and most of them are lower than that of the time series model with the number of test sets increases, which shows that although the time series is good at capturing linear relationships, it only works well in the short-term prediction range. It does not have good performance when data is mixed in dealing with real industrial problems. In most cases, the LSTM model is more powerful than the time series model. Although the prediction results of the ARMA model and the LSTM model are comparable, the ARMA model requires a lot of preliminary parameter calibration work, which not only consumes a lot of energy, but also is not suitable for the online updating of the model in the process industry online learning system.

G. THE MODIFIED ALGORITHM

Due to the existence of real-time errors, the application of modified algorithm is of great significance to improve the

prediction accuracy. According to the previous algorithm, it is known that the calculation of the C_{val} involves two important variables: *Sampling_period* and *Cycle_time*. Since the time series of the supervised learning data set designed by us has equal time interval, the *Sampling_period* and *Cycle_time* can be expressed equivalently as the number of time series data, that is the length of time can be expressed equivalently as the amount of data. In order to study the relationship between *Sampling_period* and *Cycle_time* in the modified algorithm, we have made the following experiment, the result is shown in Table 8.

In the experiment, when $Sampling_period < Cycle_time$, the collected data is only a part of the data in the *Sampling_period*, which results in the discontinuity of the time series and the incompleteness of the data, this type of experiment is meaningless. When the *Cycle_time* is constant, the average absolute percentage error and the root mean square error both increase as the *Sampling_period* increases. When $Sampling_period = Cycle_time$, the average absolute percentage error and the root mean square error also increase as the *Sampling_period* increases. It can be obtained that when the cycle time is equal to the sampling period, the smaller the time, the better the effect. However, the practical factors must be considered, such as the cycle time and sampling period in this experiment are subject to the sampling frequency of the monitoring value of the online monitor at the outlet on the diesel hydrodesulfurization production line.

H. COMPARISON OF DIFFERENT DATA STRUCTURES

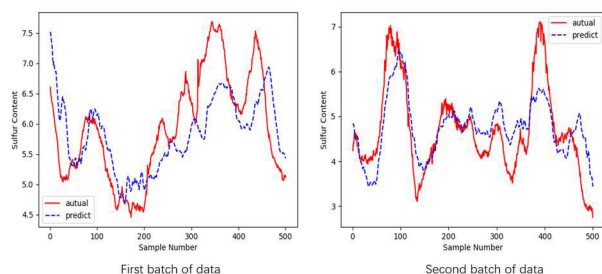
In the process industry, equipment maintenance and replacement are common. In addition, when the processing design is modified or some machines and equipment fail, the data source will change, which will lead to the change of data structure. In order to verify whether the model designed by us can have higher prediction accuracy when the input data structure of the model changes, two batches of data with different process modes on the diesel hydrodesulfurization production line are tested, and the experimental results are shown in the Fig.15. We can observe that when the structure of input data changes, the model we designed can still better

TABLE 8. The relationship between *Sampling_period* and *Cycle_time* in the modified algorithm.

Sampling period	2		5		8		10		15		20	
Cycle time	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
2	4.395	0.312	5.288	0.377	6.150	0.438	6.681	0.475	7.815	0.555	8.684	0.617
5	--	--	8.229	0.586	8.821	0.629	9.236	0.657	9.961	0.712	10.458	0.751
8	--	--	--	--	10.898	0.775	11.116	0.793	11.513	0.829	11.968	0.854
10	--	--	--	--	--	--	12.077	0.863	12.425	0.886	12.592	0.902
15	--	--	--	--	--	--	--	--	14.172	0.995	14.168	0.993
20	--	--	--	--	--	--	--	--	--	--	14.886	1.042

TABLE 9. The prediction performance of the model in two batches of data.

Sulfur-LSTM	Test-MAPE	Test-RMSE
First batch of data	8.640	0.647
Second batch of data	9.854	0.606

**FIGURE 15.** Comparison of the predicted values by the model trained in two data sets with different data structures and true values about sulfur content.

capture potential trends and the model has good reliability and generalization. It can be seen from the Table 9 that MAPE and RMSE of the two batches of data tested are within a reasonable range, which further verifies the effectiveness of the model.

I. RECOMMENDATION ON APPLICATION

In the process of data processing, due to the large difference in the magnitude of industrial data, the data should be normalized before the data alignment and data filtering methods are applied. The normalized data set must be redefined as three-dimensional data before it can be entered into the network. Its format is number of samples, time lag, feature number. Model parameters should be tuned according to experimental conclusion and actual predicted results. For the application of the modified algorithm, relevant parameters should be set reasonably according to the actual results and field experience, such as Std_max . In addition, in model online updating strategy, the model update mechanism is determined by a number of key parameters, such as similarity factors and similarity limits. Such parameter values can be set flexibly, and users can design by themselves according to the actual research field. Design metrics are usually mean value, absolute error, variance, etc.

VII. CONCLUSION

This paper presents a method for real-time prediction of sulfur content based on multi-source heterogeneous time series data in diesel hydrodesulfurization. The method is verified on the actual data set and has a promising prediction effect. Several useful findings can be generated in this paper:

(1) Data preprocessing method based on multi-source heterogeneous time series data of process industry is proposed, which can integrate and simplify multiple related features.

(2) Compared with several classic regression models as well as ARMA, the predicting effect of LSTM is better, the MAPE of sulfur content can be controlled within 10. An algorithm for online correction based on real-time errors is designed, which can further reduce the MAPE of sulfur content to 4.395.

(3) During the training process, the selection of batch size, activation function and number of hidden layer neurons has a great influence on the fitting effect and running time. The Sulfur-LSTM in this experiment used a batch size of 15, Linear activation function and 130 hidden layer neurons as the optimal model parameters.

(4) A set of multi-mode online training strategies and related trigger conditions are designed based on LSTM, which not only can quickly predict online, but also can perform offline tests. The user can adjust the decision according to the real-time trend of key indicators and change the working state in the reaction device to improve efficiency.

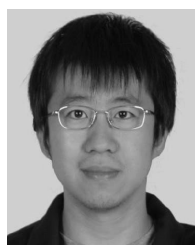
REFERENCES

- [1] L. Chen, Q. L. Liu, L. Q. Wang, J. Zhao, and W. Wang, "A review of process index forecasting methods based on data," *Acta. Auto. Sinica.*, vol. 43, no. 6, pp. 944–954, 2017.
- [2] X. J. Zhou, C. H. Yang, and W. H. Gui, "Modeling and control of non-ferrous metallurgical processes on the perspective of global optimization," *IET. Contr. Theo. A.*, vol. 32, no. 9, pp. 1158–1169, 2015.
- [3] J. C. Chen, "Development trend and typical application analysis of industrial big data," *Wireless. Interface Tech.*, vol. 16, no. 15, pp. 141–142, 2019.
- [4] Q. Liu and S. Z. Qin, "Research prospect of big data modeling in process industry," *Acta. Auto. Sinica.*, vol. 42, no. 2, pp. 3–13, 2016.
- [5] L. Wang, Y. Wang, and Q. Chang, "Feature selection methods for big data bioinformatics," *Methods.*, vol. 111, no. 21, pp. 21–31, 2016.
- [6] Y.-L. Yeh and T.-C. Chen, "Application of grey correlation analysis for evaluating the artificial lake site in pingtung plain, taiwan," *Can. J. Civil Eng.*, vol. 31, no. 1, pp. 56–64, Jan. 2004.
- [7] V. Singh, V. Tathavadkar, S. M. Rao, and K. S. Raju, "Predicting the performance of submerged arc furnace with varied raw material combinations using artificial neural network," *J. Mater. Process. Technol.*, vol. 183, no. 1, pp. 111–116, Mar. 2007.
- [8] Q. Li, Y. Liang, and Q. Zou, "Modeling the relationship between the influencing factors and the multiple responses of coal-like materials using taguchi-gray correlation analysis for their utilization in gas seepage studies," *Energy Sci. Eng.*, vol. 7, no. 3, pp. 835–850, Jun. 2019.
- [9] H. Chen and J. Chiang, "Design of an adjustable-way cache for energy reduction," *J. Chin. Inst. Eng.*, vol. 28, no. 4, pp. 691–700, Jun. 2005.
- [10] H. Saxén and F. Pettersson, "Method for the selection of inputs and structure of feedforward neural networks," *Comput. Chem. Eng.*, vol. 30, nos. 6–7, pp. 1038–1045, May 2006.
- [11] H. Saxén, F. Pettersson, and K. Gunturu, "Evolving nonlinear time-series models of the hot metal silicon content in the blast furnace," *Mater. Manuf. Processes.*, vol. 22, no. 5, pp. 577–584, Jun. 2007.
- [12] H. Yu, T. Xie, S. Paszczynski, and B. M. Wilamowski, "Advantages of radial basis function networks for dynamic system design," *IEEE Trans. Ind. Electron.*, vol. 58, no. 12, pp. 5438–5450, Dec. 2011.
- [13] A. Mohebbi, M. Ahmadi-Pour, and M. Mohebbi, "Accurate prediction of liquid phase equilibrium adsorption of sulfur compound," *Chem. Eng. Res. Design.*, vol. 126, pp. 199–208, Oct. 2017.
- [14] C. Corinna and V. Vladimir, "Multiclass least squares support vector machines," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [15] M. Curilem, G. Acuña, F. Cubillos, and E. Vyhmeister, "Neural networks and support vector machine models applied to energy consumption optimization in semiautogeneous grinding," in *Proc. Int. Conf. Energy Sav., Pollut Reduc. (PRES)*, Florence, Italy, 2011, pp. 761–766.
- [16] J. A. K. Suykens and J. Vandewalle, "Multiclass least squares support vector machines," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 1999, pp. 900–903.

- [17] I. Pulido-Calvo, P. Montesinos, J. Roldán, and F. Ruiz-Navarro, "Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems," *Biosyst. Eng.*, vol. 97, no. 2, pp. 283–293, Jun. 2007.
- [18] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial time-series data analysis using deep convolutional neural networks," in *Proc. 7th Int. Conf. Cloud Comput. Big Data (CCBD)*, Nov. 2016, pp. 87–92.
- [19] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, p. 273, Jan. 2017.
- [20] J. Taborri, E. Scalona, E. Palermo, S. Rossi, and P. Cappa, "Validation of inter-subject training for hidden Markov models applied to gait phase detection in children with cerebral palsy," *Sensors*, vol. 15, no. 9, pp. 24514–24529, Sep. 2015.
- [21] H.-D. Yang, "Sign language recognition with the kinect sensor based on conditional random fields," *Sensors*, vol. 15, no. 1, pp. 135–147, Dec. 2014.
- [22] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.
- [23] Y. Ma, Z. Zhang, and A. Ihler, "Multi-lane short-term traffic forecasting with convolutional LSTM network," *IEEE Access*, vol. 8, pp. 34629–34643, 2020.
- [24] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: <http://arxiv.org/abs/1412.3555>
- [25] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019.
- [26] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.
- [27] Y. Tian and L. Pan, "Short-term traffic flow by long short-term memory recurrent neural network," in *Proc. IEEE Int. Conf. Smart City*, Chengdu, China, Dec. 2016, pp. 153–158.
- [28] P. Kaushik, A. Gupta, P. P. Roy, and D. P. Dogra, "EEG-based age and gender prediction using deep BLSTM-LSTM network model," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2634–2641, Apr. 2019.
- [29] M. Bowie, E. Begoli, and B. Park, "Improving quality of observational streaming medical data by using long short-term memory networks (LSTMs)," in *Proc. IEEE 34th Int. Conf. Data Eng. Workshops (ICDEW)*, Apr. 2018, pp. 48–53.
- [30] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network traffic classifier with convolutional and recurrent neural networks for Internet of Things," *IEEE Access*, vol. 5, pp. 18042–18050, 2017.
- [31] A. AlGhazzawi and B. Lennox, "Monitoring a complex refining process using multivariate statistics," *Control Eng. Pract.*, vol. 16, no. 3, pp. 294–307, Mar. 2008.
- [32] S. Narasimhan and S. L. Shah, "Model identification and error covariance matrix estimation from noisy data using PCA," *Control Eng. Pract.*, vol. 16, no. 1, pp. 146–155, Jan. 2008.
- [33] N. F. Thornhill and A. Horch, "Advances and new directions in plant-wide disturbance detection and diagnosis," *Control Eng. Pract.*, vol. 15, no. 10, pp. 1196–1206, Oct. 2007.
- [34] J.-H. Cho, J.-M. Lee, S. Wook Choi, D. Lee, and I.-B. Lee, "Fault identification for process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 60, no. 1, pp. 279–288, Jan. 2005.
- [35] U. Kruger, D. Antory, J. Hahn, G. W. Irwin, and G. McCullough, "Introduction of a nonlinearity measure for principal component models," *Comput. Chem. Eng.*, vol. 29, nos. 11–12, pp. 2355–2362, Oct. 2005.
- [36] W. Li, H. H. Yue, S. Valle-Cervantes, and S. J. Qin, "Recursive PCA for adaptive process monitoring," *J. Process Control*, vol. 10, no. 5, pp. 471–486, Oct. 2000.
- [37] X. Wang, U. Kruger, and G. W. Irwin, "Process monitoring approach using fast moving window PCA," *Ind. Eng. Chem. Res.*, vol. 44, no. 15, pp. 5691–5702, Jul. 2005.
- [38] Y. M. Chai and C. F. Lei, "Review of online learning behavior research based on data mining technology," *Comput. Appl. Res.*, vol. 35, no. 5, pp. 1287–1293, 2018.



JINGKAI MA was born in 1996. He received the bachelor's degree from the College of Artificial Intelligence, Beijing Technology and Business University. He is currently pursuing the master's degree with the Machine Learning and Intelligent Engineering Group, Department of Artificial Intelligence, Beijing Technology and Business University. He is also involved in the applications of machine learning technology.



NAN LI was born in 1979. He received the Ph.D. degree in mechanical design and theory from Beijing Jiaotong University, China, in 2010.

He is currently an Associate Professor with the College of Artificial Intelligence, Beijing Technology and Business University. He has presided over and participated in projects, including the National Natural Science Foundation of China, the National Key Research and Development Project of China, and various types of enterprise customization systems. He is the author of one book and more than 50 articles. His research interests include intelligent engineering, intelligent manufacturing, CAD/CAE, distributed collaborative design, and environmental computational acoustics.



YATING HUANG was born in 1982. She received the B.S. degree in engineering and the Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, in 2005 and 2010, respectively.

From 2010 to 2017, she was a Lecturer with the School of Artificial Intelligence, Beijing Technology and Business University. Since 2017, she has been an Assistant Professor with the College of Artificial Intelligence, Beijing Technology and Business University. From 2018 to 2019, she visited with Texas A&M University, College Station, as a Visiting Scholar. In recent years, she has presided over and participated in projects, including the National Social Sciences Fund Project and various types of enterprise customization systems. She has published over ten articles. Her research interests focus in the broad area of surface and interface science and engineering. The topics are detection technology of nano-particle and mechanism of nano-particle adhesion and motion, experimental test and numerical computation of megasonic filed and failure of the surface nano structure, and biotribology and food tribology.



YUE MA was born in Beijing, China, in 1996. He received the bachelor's degree from the College of Artificial Intelligence, Beijing Technology and Business University, in 2018. He is currently pursuing the master's degree with the School of Artificial Intelligence, Beijing Technology and Business University. He is also involved in digital design and manufacturing.

...