Advanced Engineering Informatics 46 (2020) 101136

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Ensemble deep learning based semi-supervised soft sensor modeling method and its application on quality prediction for coal preparation process

Xianhui Yin[a], Zhanwen Niu[a], Zhen He[a,*], Zhaojun(Steven) Li[b], Dong-hee Lee[c]

[a] College of Management and Economics, Tianjin University, Tianjin 300072, China
[b] Department of Industrial Engineering and Engineering Management, Western New England University, Springfield, MA 01119, USA
[c] Division of Interdisciplinary Industrial Studies, Hanyang University, Wangsimniro 222, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Quality prediction
Soft sensor
Coal preparation process
Semi-supervised deep learning
Unlabeled data
Temporal dependency

ABSTRACT

Coal preparation is the most effective and economical technique to reduce impurities and improve the product quality for run-of-mine coal. The timely and accurate prediction for key quality characteristics of separated coal plays a significant role in condition monitoring and production control. However, these quality characteristics are usually difficult to directly measure online in industrial practices Although some computation intelligence based soft sensor modeling methods have been developed and reported in existing research for these quality variables estimation, some problems still exist, i.e., manual feature extraction, considerable unlabeled data, temporal dynamic behavior in data, which will influence the accuracy and efficiency for established soft sensor model. To address above-mentioned problem and develop an more excellent quality prediction model for coal preparation process, a novel deep learning based semi-supervised soft sensor modeling approach is proposed which combining the advantage of unsupervised deep learning technique (i.e., Stacked Auto-Encoder (SAE)) with the advantage of supervised deep bidirectional recurrent learner (i.e., Bidirectional Long Short-Term Memory (BLSTM)). More specifically, the unsupervised SAE networks are implemented to learn the representative features hidden in all available input data (labeled and unlabeled samples) and store them as context vector. Then, partial context vector with corresponding labels and the quality variable measure value at previous time are concatenated to form a new merged input feature vector. After that, the temporal and dynamic features are further extracted from the new merged input feature vector via BLSTM networks. Subsequently, the fully connected layers (FCs) are exploited to learn the higher-level features from the last hidden layer of the BLSTM. Lastly, the learned output features by FCs are fed into a supervised liner regression layer to predict the coal quality metrics. Meanwhile, to avoid over-fitting, some regularization techniques are utilized and discussed in proposed network. The application in ash content estimation for a real dense medium coal preparation process and some comparison experiment result demonstrate that the effectiveness and priority of proposed soft sensor modeling approach.

1. Introduction

Coal is one of the most primary and available energy resources, which provides a significant contribution to the economic development of many countries all around the world[1]. With the development of society, high-quality coal plays an important role in improving production efficiency and alleviating environmental pollution resulting from coal combustion[2]. Coal preparation is the most effective and economical solution to reduce impurities and improve coal quality, during which run-of-mine coal is transformed into required products (e.g., clean coal) by removing the ash-forming (inorganics) impurities and upgrading the carbon concentration [1–3]. Thus, it is critical to timely and accurately monitor the pivotal product quality variables (e.g., ash content and calorie value) so as to achieve real-time control and optimization for production process. However, most of these key quality variables are hard to directly measure online due to reasons such as high analyzer cost with online hard device, long time consummation or delay with offline lab analysis (industrial technique analysis) and extreme working environments, etc. [4,5].

To alleviate the above-mentioned problem, soft sensor, a kind of virtual sensing technique, has been widely adopted to estimate the difficult-to-measure quality variables with the help of acquired easy-to-measure process variables by some predictive mathematical models to offer accurate, reliable and economical alternatives to these expensive

* Corresponding author at: Collegement of Management and Econmics, Tianjin University, China.
E-mail address: zhhe@tju.edu.cn (Z. He).

https://doi.org/10.1016/j.aei.2020.101136
Received 31 August 2019; Received in revised form 27 June 2020; Accepted 28 June 2020

physical analysis [4–7]. These soft sensors can provide important real-time information that is necessary for industrial process monitoring, control and optimization. Typically, soft sensors can be broadly classified into two categories, which are first-principle models (FPMs) and data-driven models [4]. Especially, the convenience of industrial big data acquisition and progress in data analysis make the data-driven soft sensors modeling possible and applicable in practice, which has attracted increasing interests from academic researchers and industrial practitioners. And considerable research results in existing literatures indicate that the data-driven modeling especially for machine learning (ML) based methods have various prominent superiority than FPMs and conventional statistical models[8]. For example, the artificial neural networks (ANN), extreme learning machine (ELM), least square support vector regression (LSSVR), and extreme gradient boosting (XGBOOST) have been widely studied and successfully utilized in some industrial processes for predictive modeling [9–11]. Although these ML-based soft sensor modeling approaches have achieved promising prediction performance for nonlinear and highly dimensional process data, there still exist some limitations as follow, which significantly affects the prediction accuracy and efficiency:

1) Feature engineering problem. Usually, two additional manual procedures (i.e., feature extraction and dimensionality reduction) [12] are required for above-mentioned shallow ML approaches to ensure the prediction effectiveness, which is time-consuming and susceptible to utilized approaches.
2) Severe over-fitting problem. The developed model may have severe over-fitting when the correlations among variables become complicated but the extracted features have poor representation ability for some complex problems.

Recently, deep learning (DL) technique has demonstrated its outstanding performance in automatic representative feature extraction and complex correlation fitting for high dimensional data because it has hierarchical and deep network architecture[6,7]. Thus, many scholars attempt to establish a deep learning based end-to-end soft sensor directly on the raw data where no manual feature extraction and dimensionality reduction procedure are required[13,14], which provides the solution for the problems in ML based models. According to literature review, the 1D- convolutional neural networks (CNN), 2D-CNN, deep recurrent neural network (RNN), stacked auto-encoder (SAE), and deep belief networks (DBN) have been researched and exploited to construct the soft sensor models [4,15–23]. Although these developed DL-based soft sensors achieved more excellent performance in previous research work, some instinct limitation for method need to be further considered and addressed. Additionally, most of these modes are established with idea condition assumption, which maybe not reflect the actual characteristics for industrial process. The detailed description about these problems is presented as follow:

1) Considerable unlabeled data. The current soft sensor model always developed under the assumption that all the collected data has label. However, in practical coal preparation process, the online input variables used for quality prediction are those fast-sampling process variables, while output quality variables are difficult to collect due to lowing sampling frequency or expensive labeling cost. Thus, the collected datasets are mostly partially labeled. Usually, the amount of unlabeled data is much larger than that of the labeled data, which is also known as data rich but information poor(DRIP)[4,24].
2) Temporal dynamic behavior in production process. Usually, the variations of pivotal quality characteristics may be found extremely later than the changes in operating parameters [17,25–29]. Hence, dynamic features (temporal features) extraction should be conducted for collected sequential data. For this problem, the recurrent neural networks (RNN) and its updated version (i.e., long-short term memory (LSTM) and gate recurrent unit (GRU)) have been proposed and widely exploited in soft sensor modeling to capture the temporal features for complex industrial process [25,26,30–32]. However, these RNN based soft sensor only consider the temporal dependency in input variable. Actually, the measured value for quality metric in process industry also has strong temporal correlation at neighboring time points. In addition, the vanilla RNN network can only capture the dependence among the current state and previous state (i.e., the data stream flow along the forward direction). There is no doubt that one model has the access to both forward and backward direction context can capture more detailed and precious representative features [33,34]
3) Hyper-parameters tuning and over-fitting problem. Trial-and error method are widely used to tune the hyper-parameter for developed DL based soft sensor, which is time-consuming. Moreover, little existing research concerns the solution to alleviate the over-fitting problem instead of only increasing the number of testing data.

Usually, the collected unlabeled data are abandoned or seldom utilized to make contributions to established soft sensor, which would lead to the loss of massive useful process information [5,24]. Aiming to the DRIP problem, few semi-supervised methods have been proposed and utilized in fault diagnosis and detection sector, which commonly includes two main modules, i.e., unsupervised abstract feature extraction and supervised mapping correlation establishment [5,6,35,36]. These researches also shows that the feature extraction ability of DL based feature extractor especially SAE is better than traditional unsupervised feature extractor such as Gaussian mixture model and Principal Component Analysis due to its unsupervised layer-wise pretraining and supervised fine-tuning mechanism [5,37,38]. In addition, Bi-LSTM (BLSTM) networks with two separate hidden layers can be utilized to process the sequence data in two directions to capture both past and future information, respectively. Thus, the BLSTMs can more effectively learn the representative dynamic features and handle the time varying behavior such as the operating condition changes in coal preparation process. To the best of our knowledge, the BLSTM based soft sensor modeling approaches have not been reported in existing research works, especially for quality prediction of coal preparation processes.

To fill above-mentioned research gap, a deep learning based semi-supervised soft sensor modeling method is proposed. The method integrates the unsupervised SAE network with supervised BLSTM leaner to address the unlabeled data information and temporal behavior extraction problems so as to achieve more accurate and effective quality prediction for coal preparation process. In addition, the SAE based feature extractor not only can avoid the manual and separate feature extraction procedure, but can reduce the dimensionality of input data. Moreover, Taguchi experiment design based hyper-parameter tuning method and regularization technique are also proposed and utilized to improve the performance for proposed soft sensor model. Finally, the effectiveness and superiority of the proposed method are demonstrated by a real dense medium coal preparation case.

The main contributions of this work can be summarized as follow:

1) An end-to-end deep learning based soft sensor modeling approach is proposed for quality characteristic prediction oriented coal preparation process. The method can automatically extract the hierarchical representative feature and reduce the dimensionality instead of extra feature extraction and selection procedure.
2) The information hidden in extensive unlabeled data and the bidirectional temporal dynamic behavior hidden in variables data including operating parameter and quality response can be extracted simultaneously in an ensemble model by integrating the unsupervised SAE extractor and supervised BLSTM leaner.
3) Taguchi experiment design based hyper-parameter tuning method and regularization technique are exploited to efficiently improve the accuracy and generalization for developed model.

4) The ash content prediction in a real coal preparation process is used to validate the effectiveness and superiority of the proposed soft sensor modeling than other baseline methods.

The rest of the paper is organized as follows: Section 2 presents the problem formulation and briefly reviews the related deep learning methodologies. Then, the proposed soft sensor modeling approach is outlined in detail in Section 3. Subsequently, the effectiveness and feasibility of the proposed method are validated by a real industrial case in Section 4. Meanwhile, the comparison results and discussion are elaborated in this section. Finally, the conclusion and further work are given in Section 5.

2. Problem formulation and research preliminaries

Basically, two main contents are illustrated in this section. First, the studied problem and main challenges on soft sensor modeling for coal preparation process are described and formulated. Then, two utilized deep learning techniques (i.e., SAE network and BLSTM network) are introduced in brief successively.

2.1. Problem description and formulation

As depicted in Fig. 1, the endpoint product quality characteristic of coal preparation (denoted as $\mathbf{Y}$) is difficult to directly measure online under current technical and economic condition [2,39–42], but there are some easy-to-measure operating parameters (defined as $\mathbf{X}$) that can be collected by various sensors to indirectly reflect the state for $\mathbf{Y}$. Generally, these collected online variables have shorter sampling interval. Thus extensive collected dataset is unlabeled. As shown in Fig. 1, $\mathbf{S}^l$ and $\mathbf{U}^u$ represent labeled and unlabeled data set, respectively. $\mathbf{X}^l$ and $\mathbf{X}^u$ are the measured operating parameters dataset with and without corresponding outputs, respectively. U and N denote the sample number for unlabeled and labeled dataset. In practical production process, $U \gg N$. Hence, the information extraction and utilization in extensive unlabeled data is one of the main research problems in this work. In addition, the change in endpoint quality may be extremely late than the change in operating parameters for process industry. Thus, the collected data has obvious time dependency. The time dependency in this work includes two aspects: the quality characteristic at t time point, y_t, is influenced by current time inputs $\mathbf{X}_t$ and previous inputs $\mathbf{X}_{t-m}(m = 1, 2, ...,t-1)$, m is the length of time delay for operating parameters. The other aspect is that the y_t has autocorrelation with outputs at previous time points y_{t-m}. Thus, the dynamic behavior and temporal feature extraction is another core research problem which need to be addressed in this work.

It can be seen from Fig. 1, the development of soft sensor contains two core contents: the information extraction about unlabeled data and temporal dependency extraction. Thus, the quality prediction data can be denoted as a 3-tuple (**X, Y,** t), representing the multiple sensors data (input variables), quality variable, and corresponding inspection time point, respectively. After obtaining context vector $\mathbf{X}^{*l}$ which contain abstract information hidden in labeled and unlabeled data simultaneously, dataset ($\mathbf{X}^{*l}$, $\mathbf{Y}^l$, t^l) is used to further extract dynamic feature and establish final soft sensor model. l denotes the data from labeled dataset. The concrete formulation of this problem can be illustrated as follows:

Given a training set that consists of the production data $(\mathbf{X}_t^{training}, \mathbf{y}_t^{training})$, $\mathbf{X}_t^{training} \subseteq \mathbf{X}^{*l}$, $\mathbf{y}_t^{training} \subseteq \mathbf{Y}^l$, the objective is to determine the fitting correlation f and minimize the loss function L:

$$\hat{y}_t^{training} = f((\mathbf{X}_t^{training}, \mathbf{X}_{t-1}^{training}, ...,\mathbf{X}_{t-m}^{training}), (y_{t-1}^{training}, ...,y_{t-m}^{training}): \Psi) \quad (1)$$

$$L = \min_{\Psi} \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t^{training} - \hat{y}_t^{training})^2} \quad (2)$$

where $t(t = 1, 2, ...,T)$ is the time point, Ψ represents the parameter set of deep learning based soft sensor model. Eq. (1) shows the features used to develop final soft sensor model should contain the previous information for operating parameters and quality characteristic.

Subsequently, the loss function can be trained by advanced optimization algorithm to obtain satisfied result. Accordingly, the optimal parameter set, i.e., Ψ^*, can be obtained with respect to the minimal loss function. Finally, the quality metric value at $t + 1$ time point for testing set can be estimated by trained predictive model as Eq. (3).

$$\hat{y}_{t+1}^{testing} = f((\mathbf{X}_{t+1}^{testing}, \mathbf{X}_t^{testing}, ...,\mathbf{X}_{t-m}^{testing}), (y_t^{testing}, ...,y_{t-m}^{testing}): \Psi^*) \quad (3)$$

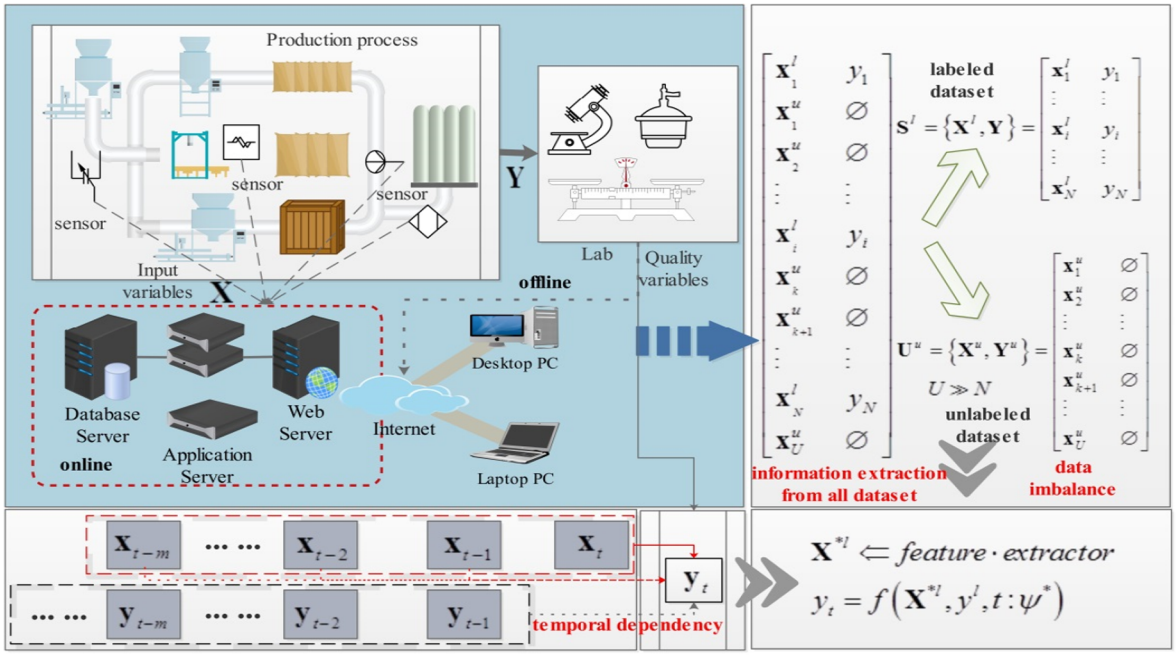

Fig. 1. Schematic diagram of the problem description.

Usually, the determination of optimal parameter set Ψ^* is based on trial-and-error or grid search methods, which are low-efficiency. Thus, rational and efficient optimal hyper-parameter estimation approach also is the important research problem in this work. Moreover, it's better that the abstract feature extraction operation can be conducted automatically during the model training process. The specific solution for each problem and the proposed soft sensor model approach are elaborated in Section3.

2.2. Related deep learning method

In this section, the theoretical basis and brief description regarding two mainly exploited deep learning modules in our proposed method, i.e., unsupervised SAE network and supervised BLSTM network are presented.

2.2.1. Stacked auto-encoder network

SAE consists of multi-layer auto-encoders (AEs) and the deep network architecture allows SAE to extract abstract feature and map complex input-output correlation hierarchically [4]. Suppose the inputs of the AE are $x = [x_{(1)}, x_{(2)}, ...,x_{(U)}]^T \in R^U$ and the hidden layer is denoted as $h = [h_{(0)}, h_{(1)}, ...,h_{d_h}]^T \in R^{d_h}$. Where U, d_h are the dimension of inputs and hidden layer, respectively. Generally, the dimension of the hidden layer d_h is less than that of inputs U for the effective features extraction and dimension reduction to alleviate over-fitting problem [43]. The encoding and decoding process from inputs to reconstructed inputs is shown as Eqs.(4) and (5).

$$h = f_{ae}(Wx + b) \tag{4}$$

$$\tilde{x} = \tilde{f}_{ae}(W'x + b'), \tilde{x} \in R^U \tag{5}$$

where $f_{ae}(x)$ is the utilized nonlinear activation function such as sigmoid function and tanh function. $\tilde{x}$ represents the reconstructed inputs. W and b are weight and bias term for this encoder. $\tilde{f}_{ae}(x)$, W', b' are activation function, weight and bias in decoder, respectively. Hence, the parameter set of an AE is $\Theta = \{W, b, W', b'\}$.

As abovementioned description, the objective of AE is to keep the reconstructed output $\tilde{x}$ as similar as possible to the initial input x. The Stacked AE means that there are L cascaded AEs stacking hierarchically. The structure of AE and SAE is depicted as Fig. 2. Two steps are performed to train the SAE model: the layer-wise unsupervised pre-training and supervised fine-tuning. During this process, the optimal parameter set Θ^* can be obtained with the loss function as shown in Eq. (6). Where, $i = 1, 2, ...,U_L$, U_L is the total number of training sample for SAE network. $\underset{\Theta}{L^{[i]}}$, $(i = 1, 2, ...L)$ represents the ith stacked AE. More details about SAE can be found in previous work [44,45].

$$\underset{\Theta}{L}(W, b, W', b') = \frac{1}{2U_L}\sum_{i=1}^{U_L} \|x - \tilde{x}\|^2 \tag{6}$$

2.2.2. Bidirectional long short-term memory network

RNN is widely used to deal with time series, which presents the support to deal with the dynamic problem. The structure of traditional RNN is illustrated in Fig. 3(a). Mathematically, the hidden state h_t and the output y_t at the tth time step are determined by current input x_t and previous hidden state h_{t-1}.

Although the RNN can handle sequential data with arbitrary lengths by using feedback neural cell, it may encounter vanishing or explosion gradient when processing extremely long sequences [33]. The LSTM model which can address these issues by memory units is an updated version of traditional RNN. The structure of an LSTM cell is presented in Fig. 3(b). The core idea is that three gates (i.e., input gate, forget gate, and output gate) are introduced into LSTM for better deciding what information to be remembered or forgotten [13] and the detailed description about operation principle of LSTM can be found in Refs. [32,33]. However, conventional LSTM only capture the dependence of current state on the previous state, which may lose some useful information when extracting temporal features [34]. Thus, the BLSTM networks which consider forward and backward LSTMs simultaneously are adopted in this work. The structure of BLSTM network is depicted in Fig. 3(c).

As shown in the BLSTM structure, the output of the network is computed by concatenating the outputs of forward and backward LSTMs. Accordingly, the mathematical equation is defined as follows:

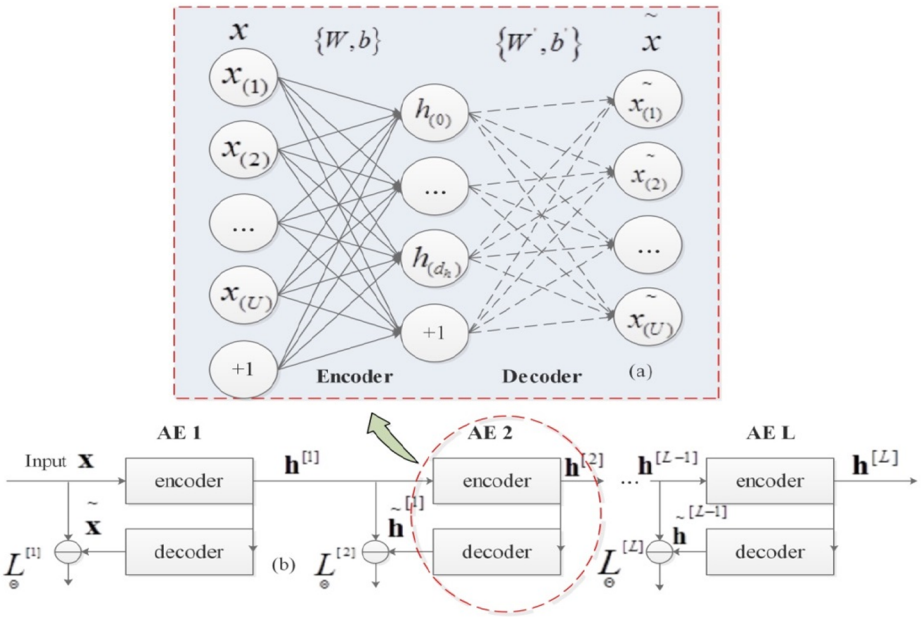

Fig. 2. The structure of SAE: (a) basic AE; (b) SAE.

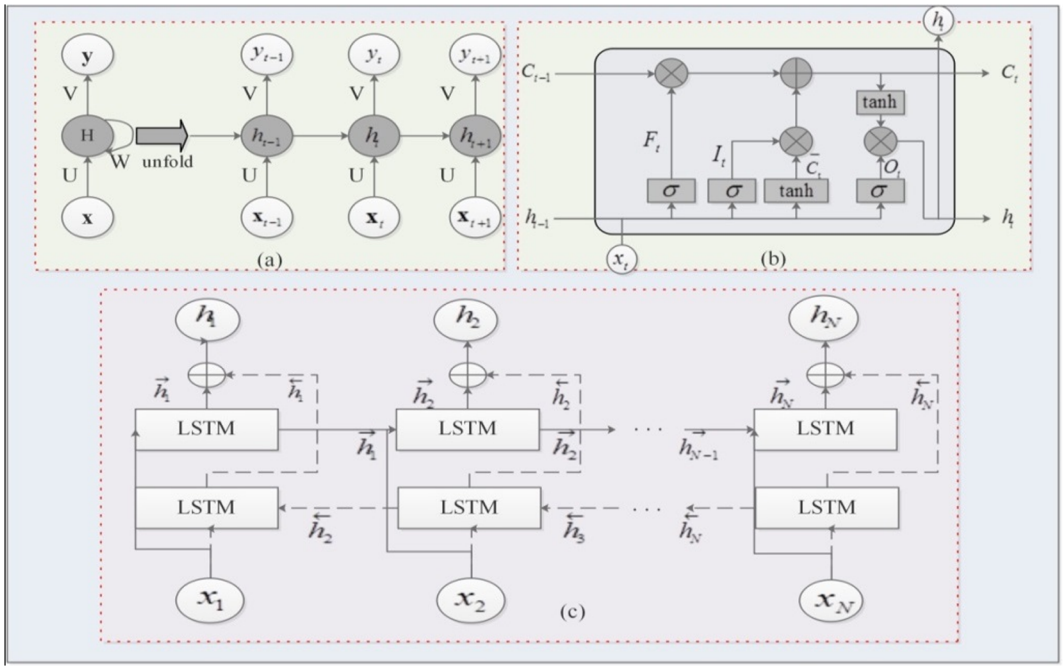

Fig. 3. The structure of RNN and its variants: (a) RNN; (b) LSTM cell; (c) BLSTM.

$$\overrightarrow{\mathbf{h}}_t = f(\mathbf{x}_t, \overrightarrow{\mathbf{h}}_{t-1}, \overrightarrow{\theta}_{BLSTM}) = \begin{cases} \overrightarrow{\tilde{\mathbf{C}}}_t = \tanh(\overrightarrow{\mathbf{W}}_c \mathbf{x}_t + \overrightarrow{\mathbf{U}}_c \overrightarrow{\mathbf{h}}_{t-1} + \overrightarrow{\mathbf{b}}_c) \\ \overrightarrow{\mathbf{I}}_t = \sigma(\overrightarrow{\mathbf{W}}_i \mathbf{x}_t + \overrightarrow{\mathbf{U}}_i \overrightarrow{\mathbf{h}}_{t-1} + \overrightarrow{\mathbf{b}}_i) \\ \overrightarrow{\mathbf{F}}_t = \sigma(\overrightarrow{\mathbf{W}}_f \mathbf{x}_t + \overrightarrow{\mathbf{U}}_f \overrightarrow{\mathbf{h}}_{t-1} + \overrightarrow{\mathbf{b}}_f) \\ \overrightarrow{\mathbf{C}}_t = \overrightarrow{\tilde{\mathbf{C}}}_t \odot \overrightarrow{\mathbf{I}}_t + \overrightarrow{\mathbf{C}}_{t-1} \odot \overrightarrow{\mathbf{F}}_t \\ \overrightarrow{\mathbf{O}}_t = \sigma(\overrightarrow{\mathbf{W}}_o \mathbf{x}_t + \overrightarrow{\mathbf{U}}_o \overrightarrow{\mathbf{h}}_{t-1} + \overrightarrow{\mathbf{b}}_o) \\ \overrightarrow{\mathbf{h}}_t = \tanh(\overrightarrow{\mathbf{C}}_t) \odot \overrightarrow{\mathbf{O}}_t \end{cases} \tag{7}$$

$$\overleftarrow{\mathbf{h}}_t = f(\mathbf{x}_t, \overleftarrow{\mathbf{h}}_{t-1}, \overleftarrow{\theta}_{BLSTM}) = \begin{cases} \overleftarrow{\tilde{\mathbf{C}}}_t = \tanh(\overleftarrow{\mathbf{W}}_c \mathbf{x}_t + \overleftarrow{\mathbf{U}}_c \overleftarrow{\mathbf{h}}_{t-1} + \overleftarrow{\mathbf{b}}_c) \\ \overleftarrow{\mathbf{I}}_t = \sigma(\overleftarrow{\mathbf{W}}_i \mathbf{x}_t + \overleftarrow{\mathbf{U}}_i \overleftarrow{\mathbf{h}}_{t-1} + \overleftarrow{\mathbf{b}}_i) \\ \overleftarrow{\mathbf{F}}_t = \sigma(\overleftarrow{\mathbf{W}}_f \mathbf{x}_t + \overleftarrow{\mathbf{U}}_f \overleftarrow{\mathbf{h}}_{t-1} + \overleftarrow{\mathbf{b}}_f) \\ \overleftarrow{\mathbf{C}}_t = \overleftarrow{\tilde{\mathbf{C}}}_t \odot \overleftarrow{\mathbf{I}}_t + \overleftarrow{\mathbf{C}}_{t-1} \odot \overleftarrow{\mathbf{F}}_t \\ \overleftarrow{\mathbf{O}}_t = \sigma(\overleftarrow{\mathbf{W}}_o \mathbf{x}_t + \overleftarrow{\mathbf{U}}_o \overleftarrow{\mathbf{h}}_{t-1} + \overleftarrow{\mathbf{b}}_o) \\ \overleftarrow{\mathbf{h}}_t = \tanh(\overleftarrow{\mathbf{C}}_t) \odot \overleftarrow{\mathbf{O}}_t \end{cases} \tag{8}$$

$$\mathbf{h}_t = \overrightarrow{\mathbf{h}}_t \oplus \overleftarrow{\mathbf{h}}_t \tag{9}$$

where two different arrow symbols: → and ← represent the forward and backward process, respectively. $\mathbf{I}$, $\mathbf{F}$, $\mathbf{O}$, $\mathbf{C}$, $\mathbf{h}$ denote the input gate, forget gate, output gate, cell state activation and hidden layer vector, respectively. θ_{BLSTM} is the parameter of BLSTM, which are shared by all time steps and determined during the training process. $\mathbf{W}$ and $\mathbf{U}$ are input and recurrent weights, respectively. $\mathbf{b}$ denotes the bias vector. σ and tanh are two different point-wise nonlinear activation function(i.e., logistic sigmoid and hyperbolic tangent). $\odot$ indicates point-wise multiplication of two vectors.

As depicted in Eq. (9), the final hidden state at tth time step is computed by augmenting both the hidden state of two different directions.

3. The proposed soft sensor modeling method

As mentioned above, SAE is capable to learn high-level abstract features with unsupervised manner and suitable to extract the information hidden in both in labeled and unlabeled data. Then, the time dependencies of labeled context vector can be captured and extracted by deep BLSTM network. Besides, all the feature extraction and selection process is automatic without external feature engineering procedure. The obtained abstract features from the hidden layer of last BLTSM will be used as inputs for further higher-level feature extraction and final mapping correlation development. Thus, an end-to end semi-supervised structure that integrates the SAE and BLSTM is proposed in our work.

3.1. Overview of the proposed framework

The Fig. 4 shows the overall framework for proposed method. It can be seen that four modules are included in the proposed soft sensor modeling method, i.e., data-processing module, core extractor and regressor development module, hyper-parameter determination modules and performance evaluation module. Notably, the feature extractor and mapping function establishment is the most important part. The structure of the core network is presented as Fig. 5, which is a sequential architecture that includes the unsupervised SAE network, supervised BLSTM unit and Full Connected layer. Thus, the encountered problems in this work can be addressed by the proposed method. The detailed training procedure will be elaborated in next subsection.

3.2. Training procedure of proposed network

Basically, Fig. 5 shows that the proposed network can be seen as a combination of unsupervised and supervised learning for collected coal preparation production data. Thus, the training procedure can be broadly divided into two steps: pre-training for unsupervised learner using all available input variables and parameters initialization for supervised networks, and parameters fine-tuning for whole network. As previously mentioned, the network parameters of SAE network or

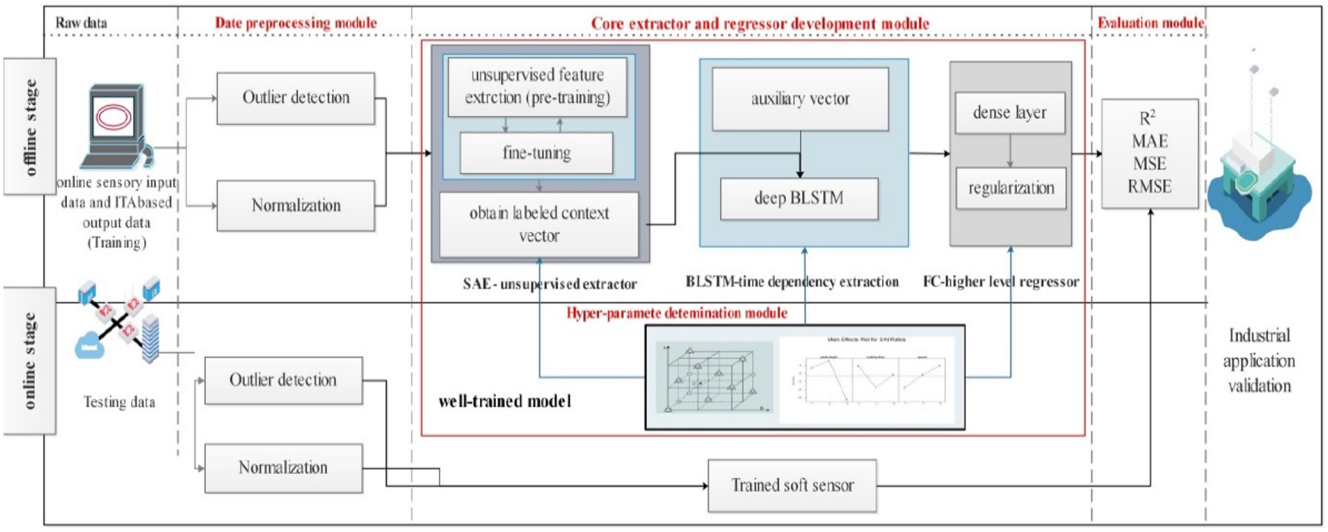

Fig. 4. The flowchart of proposed method.

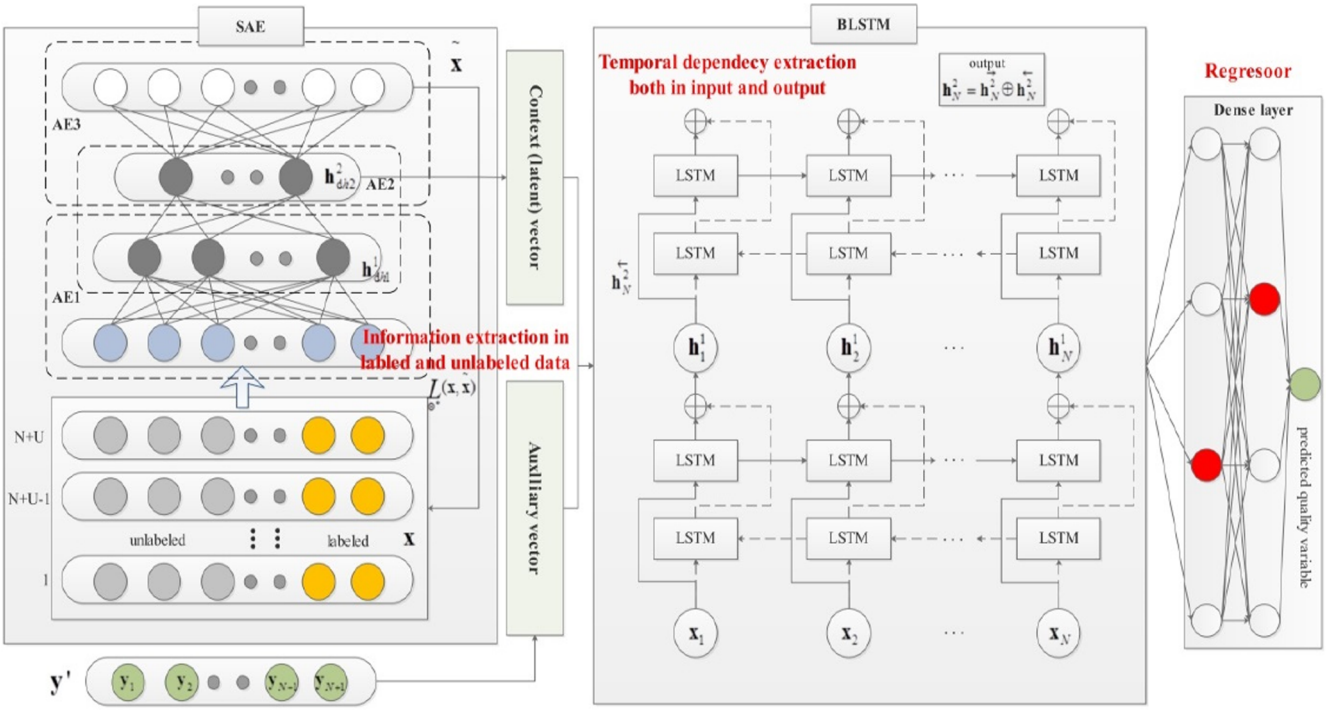

Fig. 5. The architecture of proposed deep learning network.

BLSTM unit at all N time-steps are shared. Hence, the SAE network only needs to be trained one time [13].

In the training process, the greedy layer-wise pre-training technique is used to obtain a good initialization for the weights and bias for SAE network. Concretely, the random batch gradient descent algorithm and back-propagation algorithm [4,22,43,46] are used to pre-train the SAE networks and obtain the pre-training optimal parameter set. The pre-training procedure for SAE unit can be summarized as Algorithm 1. After the SAE network has been pre-trained, the obtained optimum parameters will be set as initialized parameters for SAE when training the entire network. Moreover, the initialization procedure should also be conducted on rest structure of proposed network. Generally, Xavier uniform initializer [7] and orthogonal initializer are employed to determine the initializations of input weights and recurrent weights for BLSTMs networks, respectively. In addition, the bias of forget gate is initialized to be one and others are set as zero for BLSTMs [47,48]. Like SAE network, the parameters of fully connected layer are initialized by random initialization method [35]. In sum, the training procedure for whole proposed network can be summarized as Algorithm 2.

Algorithm 1. The pre-training algorithm for SAE

Algorithm 2. The training procedure for proposed network

3.3. Hyper-parameter determination and performance enhance solution

As can be seen in the Algorithm 1 and Algorithm 2, many learning parameters also called hyper-parameters control the function and performance of proposed network. Thus, effective and accurate determination method for these parameters plays a significant role. In this work, we utilize the Taguchi experiment design and analysis techniques to determine the optimal combination for hyper-parameter and the

---

Input: all available input variables data $\mathbf{x} = \left\{ \mathbf{x}_i^u \cup \mathbf{x}_j^l \right\}, i = 1,2,...,U, j = 1,2,...N$;

Output: the optimal parameters(weights and biases) of SAE $\Theta^{*(hi)} = \left\{ w_{hi}^*, b_{hi}^* \right\} (i = 1,2,...)$ and context vector;

Steps:

step 1: normalization and outlier removal for $\mathbf{x}$, and set the structure of SAE network;

step 2:randomly initialize the parameter set $\Theta^{(hi)}$;

step 3:learn and extract the information hidden in input data layer by layer using SAE :

$\mathbf{h}_{SAE}^{(1)} = f(\mathbf{X}^u; \Theta^{*(h1)})$, $\mathbf{h}_{SAE}^{(2)} = f(\mathbf{h}_{SAE}^{(1)}; \Theta^{*(h2)})$. And then compute the loss function;

step 4:update the parameter set by random batch gradient descent algorithm and back-propagation algorithm;

step 4:repeat step 3and 4 until convergence and obtain corresponding optimum parameter set $\Theta^{*(hi)}$ and context vector $\mathbf{c}$;

End.

Input: original dataset: D=$\left\{ \mathbf{x}, \mathbf{y}^l \right\}$,regularization coefficient: λ,look-back time step: m, batch size for SAE: g_1, batch size for BLSTM: g_2,learning rata: α, dropout rate: γ, early stopping condition: κ, activation function at various hidden layers;

Output: optimal parameter set of whole network: Ψ^*;

Steps:

step 1: preprocess the dataset: normalization and outlier removal;

step 2: determine the structure parameters for proposed network;

step 3: pre-train the SAE network using **Algorithm 1**;

step 4: initialize the parameters of remaining network layers using corresponding methods;

step 5: extract the temporal dependency using the labeled vector in obtained context vector $\mathbf{c}$ and auxiliary vector: $\mathbf{h}_N^{(2)} = \overrightarrow{\mathbf{h}}_N^{(2)} \oplus \overleftarrow{\mathbf{h}}_N^{(2)} = f_{BLSTM}\left(\left(\mathbf{h}_{SAE}^{(2)}, \mathbf{y}', \mathbf{t} \right); \overrightarrow{\theta}_{BLSTM}^{*(2)}, \overleftarrow{\theta}_{BLSTM}^{*(2)} \right)$;

step 6: The output feature vector of BLSTM networks is further used as inputs to extract higher-level and deeper features by the fully connected layer-dense layer (FCs). Finally, the linear regression layer is adopted as the top layer to merge the outputs of FCs and build the correlation between extracted features and target quality variables:

$\mathbf{o}_i = f_{DL}(\mathbf{h}_i^{(2)}; \Phi_{DL}^*) = g(\mathbf{W}_i \mathbf{h}_N^{(2)} + \mathbf{b}_i)$, $\hat{y}_i = \mathbf{W}_o \mathbf{o}_i$;

step 7: train the whole network in supervised manner and update the parameters using BPTT algorithm and Adam algorithm according to the loss function:

$L(\Psi) = \frac{1}{2(U - m + 1)} \left(\sum_{i=1}^{U-m+1} \left(y_i - \hat{y}_i \right)^2 + \lambda \left\| \Psi \right\|_2^2 \right)$. Then, the optimum parameter set Ψ^* can be obtained.

End.

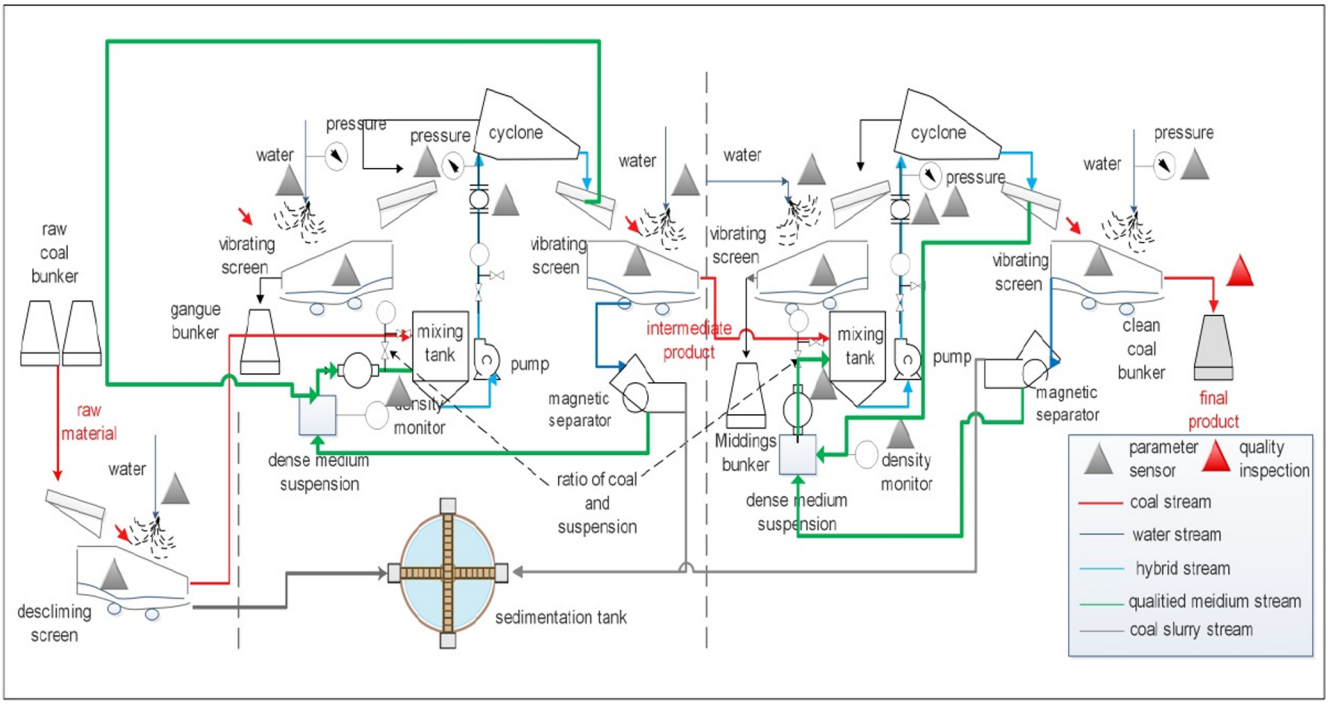

Fig. 6. The flowchart of dense medium coal preparation process.

detailed theory introduction and application steps can be found in our previous work [49].

Additionally, in order to enhance the prediction performance and alleviate the over-fitting when the number of labeled samples is limited, the feasible activation function selection and regularization techniques are design and exploited in our proposed network. More specifically, The Rectified Linear Unit (ReLU) function is employed as activation function to learn the higher-level features in dense layer due to its powerful ability in performance improvement and convergence acceleration. Moreover, the weight regularization and dropout layer are utilized to enhance the prediction performance while alleviate the over-fitting problem. The detailed introduction about these two techniques can be found in existing works [50,51]. Besides, the early stopping and is also introduced to prevent over-fitting [34]. Note that, the dropout technique is only exploited in the BLSTM units.

4. Industrial application case and results discussion

In this section, the proposed soft sensor modeling approach is verified on a real dense medium coal preparation process. First, the detailed description regarding production process and related variables about this process is outlined. Then, the established soft sensor modeling method is utilized to predict a main coal quality characteristic, i.e., ash content. Meanwhile, the effects of main hyper-parameters for proposed deep learning networks on prediction results are analyzed and discussed to determine the optimum model setting. Afterwards, the

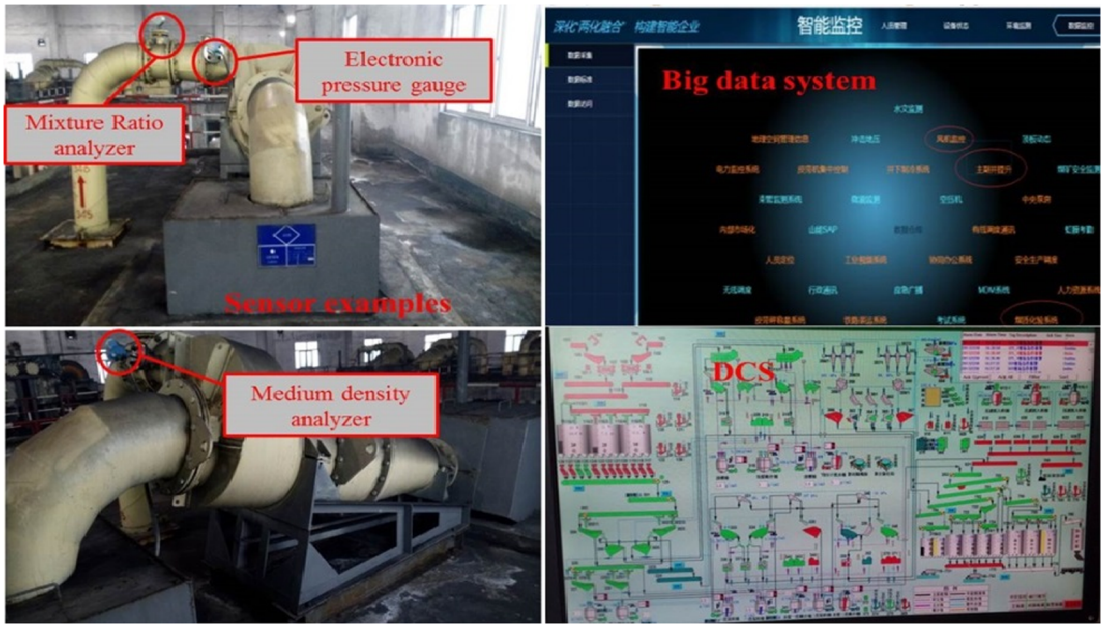

Fig. 7. The DCS and big data system.

Table 1
The definition abut input variables.

Symbol	Location	Definition
x_1	P1	circulation water density
x_2		water quantities
x_3		water pressure
x_4		feeding velocity
x_5	P2	solid/liquid ratio in low-density separator
x_6		solid/liquid ratio in high-density separator
x_7		dense medium flow
x_8	P3	the medium density in low-density tank
x_9		slime content in low-density separator
x_{10}		feeding pressure in low-density separator
x_{11}	P4	water quantities
x_{12}		water pressure
x_{13}		maximum angle of spray water
x_{14}	P5	the medium density in high-density tank
x_{15}		slime content in high-density separator
x_{16}		feeding pressure in high-density separator
x_{17}	P6	amplitude of vibrating screen
x_{18}		the thickness of processing coal
x_{19}		water quantities
x_{20}		water pressure

Notes: p_i (i = 1,2,. ..,6) represent the *ith* sub-process.

experimental results and discussion about comparison with other benchmark models are also presented to verify the effectiveness of proposed method. In this work, all algorithms utilized in proposed structure are conducted with Intel ®Core™ CPU i5-8265U 2.4 HZ, 8 GB RAM platform and Windows 10 operating system. And the python 3.7 using the Kears (2.2.4) library is employed to write the programming language.

4.1. Description of industrial process and specific problem definition

4.1.1. The introduction of dense medium coal preparation process

Dense medium beneficiation technology is well known as one of the most effective wet preparation methods, which is accomplished by exploiting the different physical property (e.g., density) between coal and impurities [1,2]. In this work, Zhengtong coal preparation plant which utilizes dense medium beneficiation technology, located in Shannxi Province, China is investigated to collect data and verify the proposed method. The flowchart of the process (clean coal stream) is elaborated in Fig. 6. It can be seen that six main processes are included in this process which are de-sliming, mixing, fist-class separation, de-watering and medium-draining, re-separation, second time dewatering and medium-draining. The detailed description about each step can be found in our previous work [27]. In addition, some online sensors are installed for key easy-to measure operating parameters in this coal preparation process such as electric pressure gauge and density meter in Fig. 6. And the advanced Distributed Control System and Coal Big Data Analysis System, as shown in Fig. 7, are also being tried to connect all the online parameters and conduct timey condition analysis and control, which give the data support for us to conducting this research.

Table 3
The experimental plan and results.

h1(SAE)	h2(SAE)	h1(BLSTM)	h2(BLSTM)	h1(dense)	h2(dense)	RMSE
10	10	10	10	5	5	0.5127
10	12	15	15	10	10	0.3742
10	14	20	20	15	15	0.4017
10	16	25	25	20	20	0.4652
10	18	30	30	30	30	0.5611
12	10	15	20	20	30	0.3511
12	12	20	25	30	5	0.3385
12	14	25	30	5	10	0.2441
12	16	30	10	10	15	0.4532
12	18	10	15	15	20	0.4684
14	10	20	30	10	20	0.2587
14	12	25	10	15	30	0.4757
14	14	30	15	20	5	0.3770
14	16	10	20	30	10	0.4707
14	18	15	25	5	15	0.3114
16	10	25	15	30	15	0.3005
16	12	30	20	5	20	0.4521
16	14	10	25	10	30	0.3673
16	16	15	30	15	5	0.3816
16	18	20	10	20	10	0.5396
18	10	30	25	15	10	0.5548
18	12	10	30	20	15	0.5152
18	14	15	10	30	20	0.2544
18	16	20	15	5	30	0.4444
18	18	25	20	10	5	0.3189

4.1.2. The problem definition and dataset description

In the coal separation process, ash content (%) (denoted as y in this work), as one of the most important quality characteristic, is usually determined by off-line industrial technology analysis at laboratory, which is time-consuming and expensive. Thus, we attempt to develop the soft sensor model and predict this key variable according the proposed method. Here, 20 online operating parameters (defined as $\mathbf{X}$) are used to prediction ash content and the detail about them is presented in Table 1.

In this work, 223-days production data including 2000 labeled and 46,000 unlabeled (total 48,000) samples are collected and stored by DCS and big data system (i.e., the sampling intervals of input variables and quality variable are 5 min and 2 h, respectively). Here, during the information extraction process with unsupervised SAE unit, 40,000 input variables are used to pre-train the SAE network and the remaining samples are used to test the pre-trained SAE unit. Besides, the number of training samples and testing samples for supervised network is1600 and 400, respectively.

4.2. Results and discussion

4.2.1. The determination about main hyper-parameter setting

Before conducting the model training procedure, some important parameters for utilized networks need to be determined. The parameters mainly include network structure parameter (i.e., the number of hidden neurons) and learning parameter of BLSTM network (i.e., look-

Table 2
The setting value of common parameters of proposed method.

Module	Hyper-parameter	Description	Set value
SAE	$g_1(x)$	activation function of first AE	sigmoid
	$g_2(x)$	activation function of second AE	sigmoid
	batch size	number of sample in one SAE training pass	1000
	learning rate	probability of descent	0.08
BLSTM	batch size	number of training samples on forward/backward pass	100
	dropout rate	probability of dropping out units in the BLSTM networks	0.2
	early stopping condition	number of time points with no significant improvement on validation error	4
Dense layer	$f_1(x)$	activation function of first dense layer	tanh
	$f_2(x)$	activation function of second dense layer	ReLu

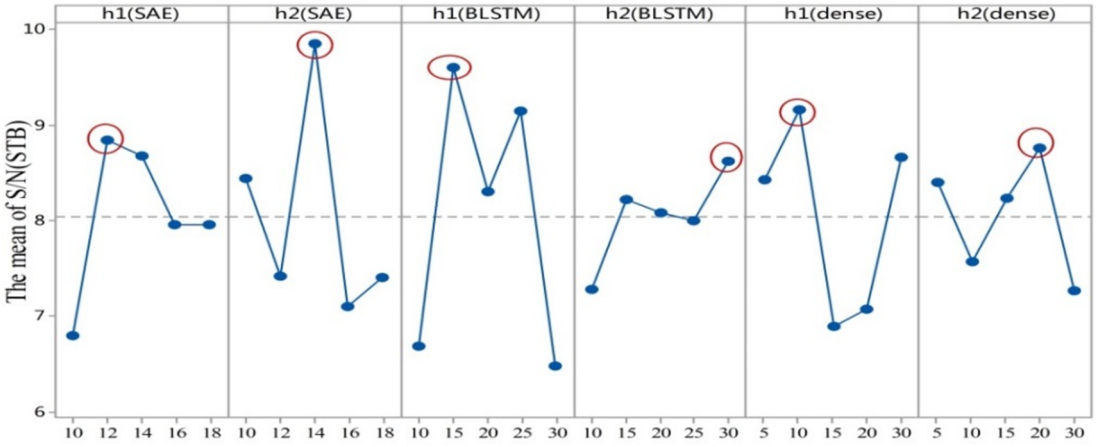

Fig. 8. The main effect plot of mean value of S/N.

back time step). The other parameters of proposed network are set according to previous research works [4,34,52,53] and trial-and error method, as shown in Table 2.

As previously description, the dimension of input and output variables is 20 and 1, respectively. Besides, total six hidden layers (two hidden layers for each module) are directly exploited in the proposed network. To determine the optimal number of hidden neurons, Taguchi orthogonal design method [54,55] is employed to explore the optimum combination of neuron number in hidden layer. We design five levels for each parameter and select L25 experimental design plan to design the experiment. Each combination is conducted 10 times and the mean RMSE is recorded in Table 3. Accordingly, the optimal number of neurons at hidden layers can be determined by Signal /Noise ratio and Taguchi analysis. The bigger value for S/N indicates the corresponding point has better prediction performance. Thus, it can be seen from Fig. 8 that the optimum combination for hidden neurons is (12, 14, 15, 30, 10, 20).

The look-back time step limits the depth of BLSTM unfolding over time and determines the maximum length of the input data [32]. Theoretically, BLSTM can automatically determine the delay time order through the gate structure in the memory cell. However, a priori and careful analysis of time step can alleviate the difficulty of training process and reduce the convergence time. In this work, fifteen time steps are discussed to determine the feasible time step in BLSTM network. As shown in Fig. 9, the proposed model has higher error (mean value in box-plot) and stronger instability (width for box-plot) when the time step less 9. And the training error has no significantly decrease after ten time steps indicating that the additional time steps are forgotten by BLSTM. In addition, although the longer time step such as 13 to 15 lead to similar mean error to model with time step 10, 11 and 12, they have poor robustness. Thus, the optimum time step is set as 10 in this work.

4.2.2. *The experimental results and discussion*

Based on the obtained optimum parameters for proposed deep learning netwok, the corresponding prediction results can be obtained. As mentioned above, the batch size for SAE network is set as 1000. The training error trends with the batch number for each AE network is described in Fig. 10. It can be observed that the used AE network not only can achieve smaller training error, but can converge very fast with a small number of training batches for each AE network, indicating the SAE network has promising ability for representative feature extraction. Furthermore, the final prediction results of whole deep learning network are depicted in Table 4. The predicted result and resudial for the testing set is presented as Fig. 11.

It can be seen that the predictive value can track well with the observed value and most of the residual lies in $[-0.05,0.0.05]$ for testing set. Additionally, the staticstical metrics in Table 4 in terms of

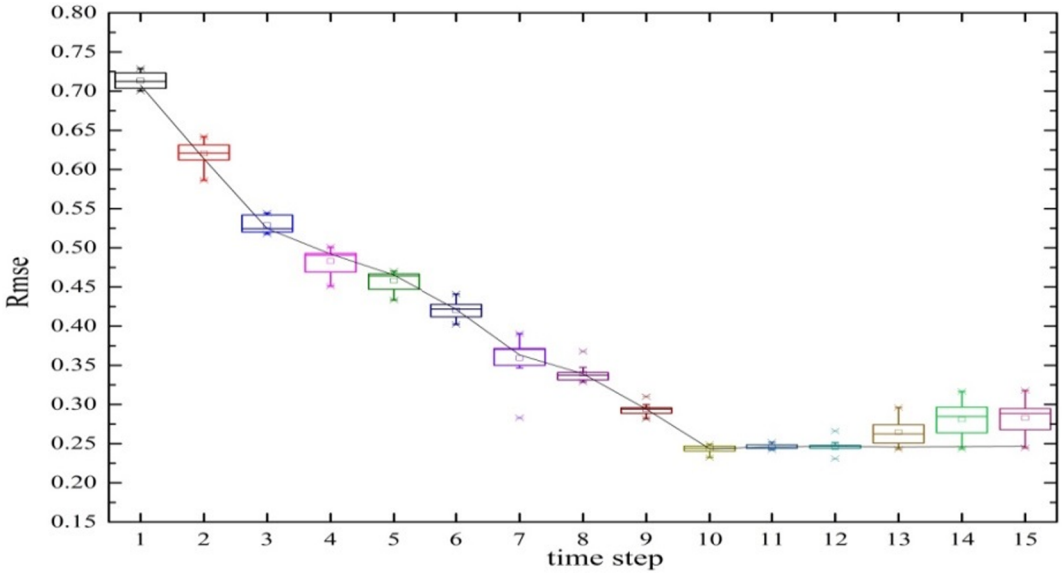

Fig. 9. The training error of BLSTM network with different look-back time step.

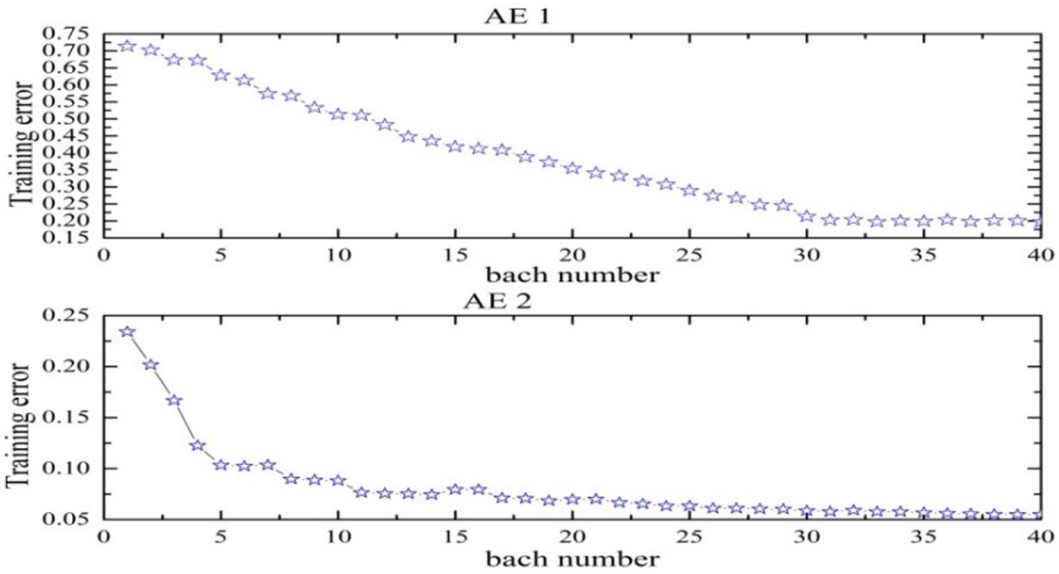

Fig. 10. The learning error of SAE network.

Table 4
The performance metrics for experimental results.

Ratio	Training set				Testing set			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
1:1	0.864	1.2071	0.6387	0.7992	0.830	1.4473	0.9822	0.9911
1:6	0.910	0.5543	0.2283	0.4778	0.887	0.6273	0.2983	0.5462
1:12	0.932	0.5192	0.1108	0.3329	0.919	0.5321	0.2032	0.4508
1:18	0.956	0.3771	0.0412	0.2027	0.933	0.3312	0.1497	0.3868
1:24	0.957	0.4021	0.0633	0. 2517	0.905	0.6117	0.2183	0.4672

R^2, MAE, MSE and RMSE also demonstrate that the proposed soft sensor modeling method has prominsing prediction performance.

To validate the effectiveness of proposed method for the encountered problems, some experiments are conducted using the collected data from the dense medium coal preparation process.

First, the advantage about utilizing extensive unlabeled data is demonstrated. In addition, the effect of ratio about labeled sample with unlabeled sample is also explored and discussed in this experiment. According to the above-mentioned sampling frequency, we select five groups dataset including 1:1(no unlabeled data), 1:6, 1:12, 1:18 and 1:24 (all unlabeled data) to analysis the prediction performance. Notably, the structure parameters and learning parameters are same for different dataset and each experiment run five times. The average evaluation metrics are outlined in Table 4. It can be observed that the prediction performance becomes better with the increase in number of unlabeled data at the early stage. Then, the performance metrics in training set began to have a little change and tend to robustness when the ratio is less than 1:18. Meanwhile, the statistical indicators for testing set shows significant decrease. For example, the determination of correlation (R^2) reduces from 0.933 to 0.905. The result shows that the utilization about unlabeled data can enhance the prediction performance comprising with supervised model while the feasible labeled data/unlabeled data ratio is also extremely significant. Excessive unlabeled data will cause information redundancy and thus lead to decreasing generalization capability and increasing time burden for the model development.

Second, five models are designed and compared to verify the effectiveness regarding temporal dependency extraction which include SAE + FC, SAE + TRNN (traditional) + FC, SAE + LSTM + FC,

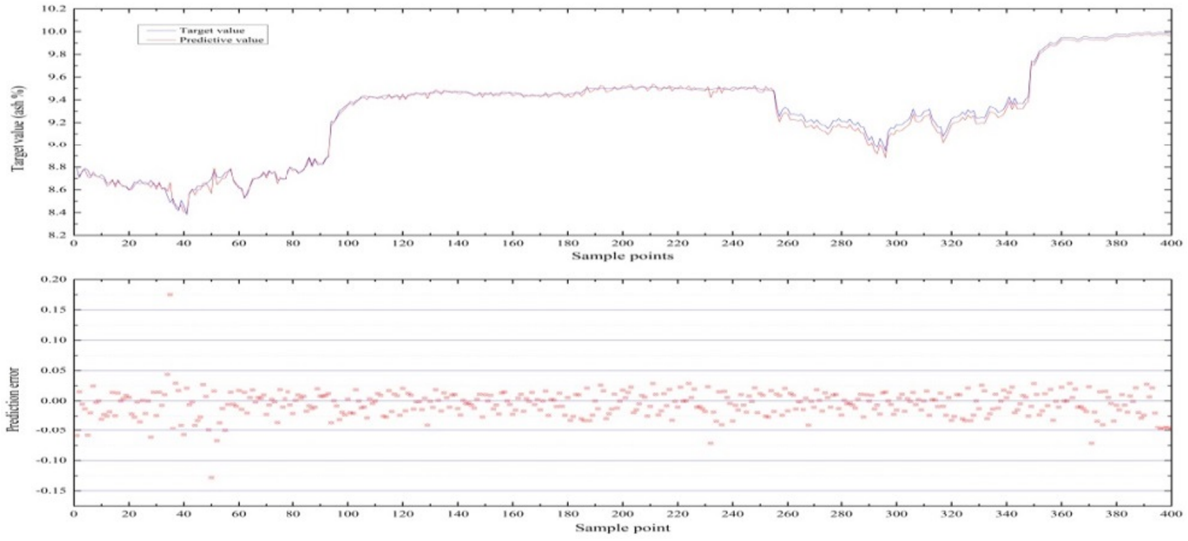

Fig. 11. Prediction result for testing set.

Table 5
The results for experiment about temporal behavior extraction.

Models	Training set				Testing set			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
SAE + FC	0.821	1.035	0.7198	0.8481	0.774	1.425	0.9946	0.9973
SAE + TRNN + FC	0.865	0.8841	0.4586	0.6772	0.819	1.023	0.6704	0.8188
SAE + LSTM + FC	0.917	0.5963	0.2206	0.4697	0.886	0.8845	0.4398	0.6632
SAE + BLSTM + FC	0.941	0.4277	0.1105	0.3324	0.901	0.5122	0.1496	0.5017
SAE + BLSTM-AV + FC*	**0.956**	**0.3771**	**0.0412**	**0.2027**	**0.933**	**0.3312**	**0.1497**	**0.3868**

Notes: * represents the proposed method in this work and the bold value is the best performance.

Table 6
The experimental result for check experiment.

Models	Training set				Testing set			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
Check model	0.961	0.3376	0.0402	0.2004	0.914	0.4223	0.2126	0.4611
proposed model	0.956	0.3771	0.0412	0.2027	0.933	0.3312	0.1497	0.3868

Table 7
The experimental result for comparison with other baseline models.

Models	Training set				Testing set			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
RBF + LSSVR	0.864	1.2071	0.6387	0.7992	0.830	1.4473	0.9822	0.9911
DBN + LSTM	0.910	0.5543	0.2283	0.4778	0.887	0.6273	0.2983	0.5462
proposed model	**0.956**	**0.3771**	**0.0412**	**0.2027**	**0.933**	**0.3312**	**0.1497**	**0.3868**

Notes: the bold value represents the best performance.

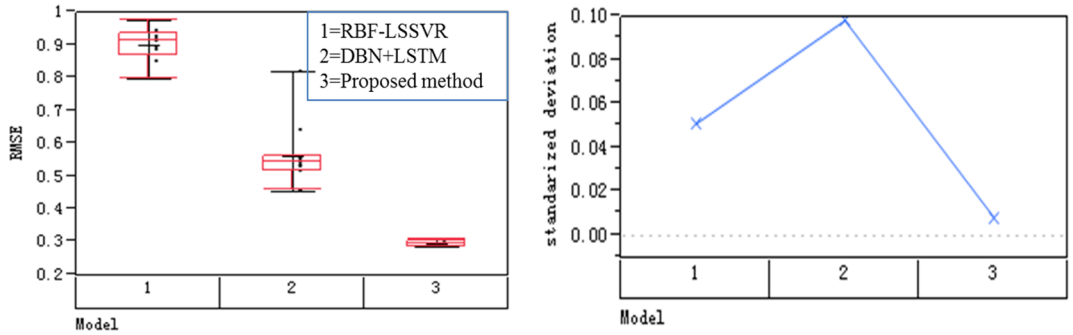

Fig. 12. The box-plot and STD of RMSE for three models.

SAE + BLSTM + FC and SAE + BLSTM-AA(auxiliary variable) + FC. It can be seen that the first model directly use the extracted context vector in SAE unit as inputs to fed into the fully connected layer to establish the mapping correlation, which can be used to evaluate the advantage about temporal feature extraction by comprising with other models. In addition, the comparison among second, third and fourth model can be used to examine whether the LSTM especially BLSTM has superiority in dynamic information extraction. Moreover, the necessarily in terms of considering auxiliary variable (time dependency for quality variable self) is also discussed by comparison analysis for last two models. Similarly, the other parameters remain same except above-mentioned difference and each model run five times. Meanwhile, in order to decrease the computation cost, the dataset with 1:18 ratio (i.e., 36,000 unlabeled samples and 2000 labeled data in which 30,000 unlabeled samples are used as training set) is exploited to develop the model. The average performance indicators are listed in Table 5.

The conclusions that can be drawn from Table 5 are summarized as follows:

1) The SAE + FC model has the worst prediction performance among these five models, which demonstrates the time dependency exists in the coal preparation operating data and its extraction is important for soft sensor modeling;
2) The models with BLSTM modules have more promising performance than other models with recurrent unit, which indicates BLSTM network used in our method can capture more representative features for time series data;
3) It also can be seen that proposed model has better performance than SAE + BLSTN-AV-FC model, which shows the consideration about temporal for output variable self can enhance the prediction performance.

Next, we will conduct check experiment to analysis whether the regularization technique used in our method can alleviate the overfitting problem and enhance the generalization capability. We design a model as same as the proposed method except regularization techniques utilization. The utilized dataset and related model parameters setting are same as last discussion section. The comparison result is

presented in Table 6. The result shows although the check model has better prediction in training set, it encounters more serious over-fitting problem. Thus, the proposed soft sensor modeling methods can effectively enhance the generalization capability, which is very import for the model utilization in the production process when the model has been established in training stage.

Finally, some representative baseline models that proposed in previous work are selected and compared with the developed method in this work to validate the effectiveness and superiority. The baseline models include shallow machine learning model LSSVR [56] and deep learning model DBN + LSTM [13]. More specifically, the Radial basis function (RBF) is used as kernel function for LSSVR and the corresponding hyper-parameters are determined with grid search method [57,58]. For fair comparison, 2 stacked Restricted Boltzmann Machine (RBM) are exploited to construct the Deep Belief Network and it has same number network neuron in hidden layer. Besides, the LSTM unit has same hidden neuron and time delay with the BLSTM unit in proposed method. Furthermore, the RBF-LSSVR is computed using Scikit-Learn package in same Python version as proposed method. And the DBN + LSTM run under the same computation environment as the proposed model. Similarly, each model also has five-times run. The experimental result is given in Table 7 and the box-plot with RSME indicator is presented in Fig. 12.

It can be seen from Table 7 and Fig. 12 that the proposed method has higher prediction accuracy and generalization capability than other two methods. Meanwhile, the proposed method is more robust comparing with the baseline models.

5. Conclusions

In this paper, a novel deep learning based semi-supervised soft sensor modeling method is developed to accurately and timely predict the quality indicator in coal preparation process combing the SAE network with BLSTM network. More specifically, the SAE unit is exploited to learn and extract representative information hidden both in labeled and unlabeled data. And then, the vector with label and the auxiliary vector (the previous state for predicted variable) are concatenated and fed into deep BLSTM unit to further extract the temporal features. Afterwards, the learned features by the last hidden layer of deep BLSTM are used as inputs for Full Connected network to capture higher-level feature and establish the mapping correlation with output variable. The experimental results based on a real dense medium coal preparation process indicate that the proposed soft senor modeling method has competitive prediction performance comprising with other benchmark methods. In addition, the proposed method, as an end-to-end model, not only can automatically extract representative features, but also can make the best of considerable unlabeled data and learn the bidirectional temporal behavior simultaneously. Meanwhile, we also find the semi-supervised model not means the prediction performance will continue increase when more and more number unlabeled data is used. Thus, the feasible ratio between labeled and unlabeled sample is significant for developed semi-supervised soft sensor. Furthermore, Taguchi experiment design based hyper-parameter determination method and some regularization techniques adopted in this work can effectively alleviate the over-fitting problem and enhance the prediction performance. Although the proposed soft sensor modeling method achieved outstanding prediction performance for complex industrial process, some issues still need to be considered in future work. For example, the modern coal preparation processes usually have long production line which contains many operating stages. This characteristic makes the time delay problem more serious. Some research works have proposed multistage modeling strategy larger-scale industry process. Hence, the possibility of combining multistage modeling strategy and deep learning techniques (specifically for LSTM network) to address the time delay problem will be explored in our further work.

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.

Acknowledgement

This work was financial supported by National Natural Science Foundation of China [grant no.71661147003, 71532008 and 71902138], and the first author was sponsored by the China Scholarship Council (CSC) under the Grant 201806250063. The authors also would like to express our sincere appreciation to Shandong Energy ZIBO Mining Group Co., LTD for their data support and related production experiment validation. The contents of this paper reflect the views of the authors and do not necessarily indicate acceptance by the sponsors.

References

[1] Z.J. Fu, J. Zhu, S. Barghi, Y.M. Zhao, Z.F. Luo, C.L. Duan, Dry coal beneficiation by the semi-industrial Air Dense Medium Fluidized Bed with binary mixtures of magnetite and fine coal particles, Fuel 243 (2019) 509–518.
[2] D. Li, D.S. Wu, F.G. Xu, J.H. Lai, L. Shao, Literature overview of Chinese research in the field of better coal utilization, J. Clean Prod. 185 (2018) 959–980.
[3] H.S. Jiang, L. Huang, Q.C. Lu, Y.M. Zhao, Z.F. Luo, C.L. Duan, L. Dong, Z.Q. Chen, B. Lv, J. Zhao, G. Huang, Separation performance of coal in an air dense medium fluidized bed at varying feeding positions, Fuel 243 (2019) 449–457.
[4] X.F. Yuan, B. Huang, Y.L. Wang, C.H. Yang, W.H. Gui, Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE, IEEE Trans. Ind. Inform. 14 (2018) 3235–3243.
[5] L. Yao, Z.Q. Ge, Scalable semisupervised GMM for big data quality prediction in multimode processes, IEEE Trans. Ind. Electron. 66 (2019) 3681–3692.
[6] L. Yao, Z.Q. Ge, Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application, IEEE Trans. Ind. Electron. 65 (2018) 1490–1498.
[7] X. Yuan, L. Li, Y. Wang, Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, IEEE Trans. Ind. Inform. (2019) 1.
[8] Z.G. Xu, Y.Z. Dang, P. Munro, Knowledge-driven intelligent quality problem-solving system in the automotive industry, Adv. Eng. Inform. 38 (2018) 441–457.
[9] S. Ibrahim, C.E. Choong, A. El-Shafie, Sensitivity analysis of artificial neural networks for just-suspension speed prediction in solid-liquid mixing systems: Performance comparison of MLPNN and RBFNN, Adv. Eng. Inform. 39 (2019) 278–291.
[10] J.-W. Han, Q.-X. Li, H.-R. Wu, H.-J. Zhu, Y.-L. Song, Prediction of cooling efficiency of forced-air precooling systems based on optimized differential evolution and improved BP neural network, Appl. Soft. Comput. 84 (2019) 105733.
[11] B.H. Si, J.G. Wang, X.Y. Yao, X. Shi, X. Jin, X. Zhou, Multi-objective optimization design of a complex building based on an artificial neural network and performance evaluation of algorithms, Adv. Eng. Inform. 40 (2019) 93–109.
[12] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S.H. Kia, H. Henao, G.A. Capolino, Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems, IEEE Trans. Ind. Electron. 66 (2019) 6331–6342.
[13] Q.Q. Sun, Z.Q. Ge, Probabilistic sequential network for deep learning of complex process data and soft sensor application, IEEE Trans. Ind. Inform. 15 (2019) 2700–2709.
[14] M. Xia, T. Li, T.X. Shu, J.F. Wan, C.W. de Silva, Z.R. Wang, A two-stage approach for the remaining useful life prediction of bearings using deep neural networks, IEEE Trans. Ind. Inform. 15 (2019) 3703–3711.
[15] J. Bedi, D. Toshniwal, Deep learning framework to forecast electricity demand, Appl. Energy 238 (2019) 1312–1326.
[16] Y. Gao, L. Gao, X. Li, X. Yan, A semi-supervised convolutional neural network-based method for steel surface defect recognition, Rob. Comput. Integr. Manuf. 61 (2020) 101825.
[17] B.T. Le, D. Xiao, Y.C. Mao, L. Song, D.K. He, S.J. Liu, Coal classification based on visible, near-infrared spectroscopy and CNN-ELM algorithm, Spectrosc. Spectr. Anal. 38 (2018) 2107–2112.
[18] Z.X. Li, K. Goebel, D.Z. Wu, Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning, J. Eng. Gas. Turbines Power-Trans. ASME 141 (2019) 10.
[19] J.J. Wang, Y.L. Ma, L.B. Zhang, R.X. Gao, D.Z. Wu, Deep learning for smart manufacturing: Methods and applications, J. Manuf. Syst. 48 (2018) 144–156.
[20] K.J. Wang, X.X. Qi, H.D. Liu, J.K. Song, Deep belief network based k-means cluster approach for short-term wind power forecasting, Energy 165 (2018) 840–852.
[21] Z.P. Zhang, J.S. Zhao, A deep belief network based fault diagnosis model for complex chemical processes, Comput. Chem. Eng. 107 (2017) 395–407.
[22] N. Chouikhi, B. Ammar, A. Hussain, A.M. Alimi, Bi-level multi-objective evolution of a Multi-Layered Echo-State Network Autoencoder for data representations,

Neurocomputing 341 (2019) 195–211.
[23] X. Wang, H. Liu, Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction, Adv. Eng. Inform. 36 (2018) 112–119.
[24] Y. Liu, C. Yang, Z.L. Gao, Y. Yao, Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes, Chemometr. Intell. Lab. Syst. 174 (2018) 15–21.
[25] X.X. Chen, X. Chen, J.H. She, M. Wu, A hybrid time series prediction model based on recurrent neural network and double joint linear-nonlinear extreme learning network for prediction of carbon efficiency in iron ore sintering process, Neurocomputing 249 (2017) 128–139.
[26] A. Sagheer, M. Kotb, Time series forecasting of petroleum production using deep LSTM recurrent networks, Neurocomputing 323 (2019) 203–213.
[27] X.H. Yin, Z. He, Z.W. Niu, Z.J. Li, A hybrid intelligent optimization approach to improving quality for serial multistage and multi-response coal preparation production systems, J. Manuf. Syst. 47 (2018) 199–216.
[28] R. Xie, K. Hao, B. Huang, L. Chen, X. Cai, Data-driven modeling based on two-stream λ gated recurrent unit network with soft sensor application, IEEE Trans. Ind. Electron. 1–1 (2019).
[29] X.L. Sun, Z.G. Cao, Y.H. Yue, Y.L. Kuang, C.X. Zhou, Online prediction of dense medium suspension density based on phase space reconstruction, Part. Sci. Technol. 36 (2018) 989–998.
[30] J.L. Chen, H.J. Jing, Y.H. Chang, Q. Liu, Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process, Reliab. Eng. Syst. Saf. 185 (2019) 372–382.
[31] Y. Qin, K. Li, Z.H. Liang, B. Lee, F.Y. Zhang, Y.C. Gu, L. Zhang, F.Z. Wu, D. Rodriguez, Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal, Appl. Energy 236 (2019) 262–272.
[32] P. Tan, B. He, C. Zhang, D.B. Rao, S.N. Li, Q.Y. Fang, G. Chen, Dynamic modeling of NOx emission in a 660 MW coal-fired boiler with long short-term memory, Energy 176 (2019) 429–436.
[33] X.L. Ma, J.Y. Zhang, B.W. Du, C. Ding, L.L. Sun, Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction, IEEE Trans. Intell. Transp. Syst. 20 (2019) 2278–2288.
[34] C. Huang, H. Huang, Y. Li, A bidirectional LSTM prognostics method under multiple operational conditions, IEEE Trans. Ind. Electron. 66 (2019) 8792–8802.
[35] C. Zhang, C. Wang, N. Lu, B. Jiang, An RBMs-BN method to RUL prediction of traction converter of CRH2 trains, Eng. Appl. Artif. Intell. 85 (2019) 46–56.
[36] A.L. Ellefsen, E. Bjorlykhaug, V. Aeligesoy, S. Ushakov, H.X. Zhang, Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture, Reliab. Eng. Syst. Saf. 183 (2019) 240–251.
[37] F.D. Frumosu, M. Kulahci, Big data analytics using semi-supervised learning methods, Qual. Reliab. Eng. Int. 34 (2018) 1413–1423.
[38] Z.Q. Ge, Z.H. Song, S.X. Deng, B. Huang, Data mining and analytics in the process industry: the role of machine learning, IEEE Access 5 (2017) 20590–20616.
[39] C. Liu, C.C. Zhou, N.N. Zhang, J.H. Pan, S.S. Cao, M.C. Tang, W.S. Ji, T.T. Hu, Modes of occurrence and partitioning behavior of trace elements during coal preparation-A case study in Guizhou Province, China, Fuel 243 (2019) 79–87.
[40] N. Wang, R.F. Shen, Z.G. Wen, D. De Clercq, Life cycle energy efficiency evaluation for coal development and utilization, Energy 179 (2019) 1–11.
[41] S. Wang, Y. Yang, X.L. Yang, Y.D. Zhang, Y.M. Zhao, Dry beneficiation of fine coal deploying multistage separation processes in a vibrated gas-fluidized bed, Sep. Sci. Technol. 54 (2019) 655–664.
[42] B. Zhang, G.Q. Zhu, B. Lv, G.H. Yan, A novel and effective method for coal slime reduction of thermal coal processing, J. Clean Prod. 198 (2018) 19–23.
[43] K. Wang, B. Gopaluni, J. Chen, Z. Song, Deep learning of complex batch process data and its application on quality prediction, IEEE Trans. Ind. Inform. (2018) 1.
[44] S. Chen, J. Yu, S. Wang, Monitoring of complex profiles based on deep stacked denoising autoencoders, Comput. Ind. Eng. 143 (2020) 106402.
[45] M. Sun, H. Wang, P. Liu, S. Huang, P. Fan, A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings, Measurement 146 (2019) 305–314.
[46] Z. Manbari, F. AkhlaghianTab, C. Salavati, Hybrid fast unsupervised feature selection for high-dimensional data, Expert Syst. Appl. 124 (2019) 97–118.
[47] H.F. Yang, Y.P.P. Chen, Hybrid deep learning and empirical mode decomposition model for time series applications, Expert Syst. Appl. 120 (2019) 128–138.
[48] J.J. Zhang, P. Wang, R.X. Gao, Deep learning-based tensile strength prediction in fused deposition modeling, Comput. Ind. 107 (2019) 11–21.
[49] X. Yin, Z. Niu, Z. He, Z. Li, D. Lee, An integrated computational intelligence technique based operating parameters optimization scheme for quality improvement oriented process-manufacturing system, Comput. Ind. Eng. 140 (2020).
[50] V. Asghari, Y.F. Leung, S.-C. Hsu, Deep neural network based framework for complex correlations in engineering metrics, Adv. Eng. Inf. 44 (2020) 101058.
[51] Z. Han, M.M. Hossain, Y. Wang, J. Li, C. Xu, Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network, Appl. Energy 259 (2020).
[52] S. Yadav, A. Ekbal, S. Saha, A. Kumar, P. Bhattacharyya, Feature assisted stacked attentive shortest dependency path based Bi-LSTM model for protein-protein interaction, Knowledge-Based Syst. 166 (2019) 18–29.
[53] S. Hochreiter, J.J.N.C. Schmidhuber, Long Short-Term Memory 9 (1997) 1735–1780.
[54] K. Ouyang, H.W. Wu, S.C. Huang, S.J. Wu, Optimum parameter design for performance of methanol steam reformer combining Taguchi method with artificial neural network and genetic algorithm, Energy 138 (2017) 446–458.
[55] G.A. Lujan-Moreno, P.R. Howard, O.G. Rojas, D.C. Montgomery, Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study, Expert Syst. Appl. 109 (2018) 195–205.
[56] M. Sharifzadeh, A. Sikinioti-Lock, N. Shah, Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression, Renew. Sust. Energ. Rev. 108 (2019) 513–538.
[57] P. Zhou, D.W. Guo, T.Y. Chai, Data-driven predictive control of molten iron quality in blast furnace ironmaking using multi-output LS-SVR based inverse system identification, Neurocomputing 308 (2018) 101–110.
[58] P. Zhou, D.W. Guo, H. Wang, T.Y. Chai, Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking, IEEE Trans. Neural Netw. Learn. Syst. 29 (2018) 4007–4021.