IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 22, 2019, accepted October 26, 2019, date of publication October 31, 2019, date of current version November 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950661

A Knowledge- and Data-Driven Soft Sensor Based on Deep Learning for Predicting the Deformation of an Air Preheater Rotor

XIAO WANG[1,2] **AND HAN LIU**[1]

[1]School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China
[2]Department of Mechanical and Electrical Engineering, Shandong Vocational College of Light Industry, Zibo 255300, China

Corresponding author: Han Liu (liuhan@xaut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China, under Grant 61973248, Grant 61833013, and Grant 61533014, in part by the Key Project of Shaanxi Key Research and Development Program, under Grant 2018ZDXM-GY-089, in part by the Research Program of Shaanxi Collaborative Innovation Center of Modern Equipment Green Manufacturing under Grant 304-210891704, and in part by the Higher Educational Science and Technology Program of Shandong Province under Grant J18KB151.

ABSTRACT In industrial processes, some important process variables cannot be measured directly by hardware sensors for technical or economic reasons. Soft sensors estimate these key variables using some other easily measured variables by building a mathematical model. A novel knowledge- and data-driven soft sensor is proposed in this paper to predict the deformation of an air preheater rotor in a thermal power plant boiler. Two submodels were constructed, including the knowledge-driven submodel, derived from all the available domain knowledge, and the data-driven submodel, constructed solely from the data. The two submodels were integrated with a mass balance model. A mathematical model based on technical expertise in predicting rotor deformation, named the Lab model, was used as the knowledge-driven submodel, and a deep learning model based on stacked autoencoders (SAE) was used as the data-driven submodel. To improve the performance of the model, the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm was adopted to optimize the SAE parameters. The experimental results demonstrate that, compared with the common knowledge-driven (KDM) and data-driven (DDM) models, the proposed Lab-stacked autoencoders (L-SAE) model is able to provide a higher predictive accuracy for the air preheater rotor deformation and inherits the advantages of both the KDM and DDM.

INDEX TERMS Deep learning, industrial process control, knowledge- and data-driven model, soft sensor.

I. INTRODUCTION

In industrial processes, there are some variables that play an important role in improving efficiency and product quality. Efficient monitoring and control of these variables are crucial for the control system of the industrial process. Unfortunately, these key variables are often difficult to measure directly using hardware sensors, which can be attributed to two main reasons. One reason is that there are no appropriate hardware sensors for some special industrial measurements due to technical or economic limitations. Another reason is that the harsh environment of industrial sites usually induces the abnormal working of hardware sensors [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Gaggero.

Soft sensors estimate the key variables in the industrial process indirectly by constructing a mathematical model. Some easily measured variables are used as the input, and the target variables are used as the output [3]. The input variables are selected from many easily measured variables, which appear to be more relevant to the target variables. Then, the key variables can be estimated by establishing the mathematical relationships with the input variables [4].

In general, there are two main types of soft sensors: knowledge-driven models (KDMs) [5], [6] and data-driven models (DDMs) [7], [8]. The KDMs occupied a dominant position for a long time and achieved good performance. The modeling process of KDMs is mainly based on industry knowledge, including the known relationships from physically based or mechanistic models. This kind of method mainly includes multivariate statistics [9], Kalman

filters [10], and clustering [11]. However, the KDMs often obtain a lower prediction accuracy in comparison with the data-driven methods for two main reasons. One reason is that the modeling process of the KDMs only considers the visible knowledge but ignores the data that reflects the real condition of the industrial process. The other reason is that the industrial processes that require soft sensor techniques always have complex physical backgrounds, making it difficult to describe the processes thoroughly with a KDM. In recent years, data-driven soft sensors have become increasingly popular due to their low cost and time savings, especially in complex industrial process modeling and control [12]. It has been proven that DDMs have high accuracy because they are based on the data collected from the industrial site, which more reliably describes the real process conditions [13]. However, because of the characteristics of the modeling method and the limitations of the process data, the prediction accuracy can only be guaranteed in the local range of the data-driven soft sensor. In addition, many industrial processes are multivariate and nonlinear and have wide operational ranges, which leads to difficulty in achieving satisfactory results. In this paper, a new knowledge- and data-driven soft sensor, named the Lab-stacked autoencoders (L-SAE) model is proposed, inspired by the idea of combining the advantages of the two abovementioned methods to estimate the rotor deformation of an air preheater rotor in a thermal power plant boiler. A knowledge-driven submodel and a data-driven submodel are integrated and balanced to create a new soft sensor.

An air preheater is used to transfer the heat from the flue gas to the combustion-supporting air in a power plant boiler. Due to the large difference between the temperature of the top and bottom of the rotor, thermal deformation occurs on the sector plates and results in deformation gaps. Consequently, air leakage happens, and the thermal efficiency of the boiler is reduced, which causes large economic losses. However, the air preheaters are typically located in harsh high-temperature corrosive environments that often cause the hardware sensors to go out of service. Therefore, a soft sensor is adopted to estimate the rotor deformation of the air preheater in this paper.

To ensure a high prediction accuracy, the selection of the two submodels in this study was quite deliberate. The most common KDM for calculating the rotor deformation, the Lab model, was used as the knowledge-driven submodel, and a deep neural network based on stacked autoencoders (SAE) was used as the data-driven submodel. The traditional data-driven methods mainly include the partial least square (PLS) [14], principal component analysis (PCA) [15], artificial neural network (ANN) [16] and combined methods such as the neural network PLS (NNPLS) [17]. Many successful applications and satisfactory results have been achieved when using these methods to estimate the key variables in industrial processes. However, there are obvious deficiencies in these methods. First, the statistics-based methods are not powerful enough when dealing with the highly nonlinear industrial data [18]. Second, the data utilization ratio of these methods is relatively low because they can only use samples that consist of both the input and output values, that is, labeled data. However, in actual industrial processes, unlabeled datasets that contain only the input samples are typical because the measurement of the target variables is difficult [19]. Third, the general methods based on ANNs are plagued by gradient diffusion and local minima [20]. In recent years, deep learning has achieved great success and has gained broad applications in many fields [21]–[23]. Its core technology is layer-by-layer unsupervised pretraining, which makes the training process of a deep neural network run smoothly and significantly improves the feature extraction and representation capability. Therefore, deep learning outperforms the traditional methods in many complex problems [24], [25]. Moreover, deep learning has also been proven to be more appropriate for soft sensor modeling [26] because it utilizes both the labeled and unlabeled data to avoid wasting the large amount of unlabeled data in comparison with the traditional soft sensors. On the other hand, deep learning performs better in revealing the underlying relationships of the nonlinear industrial data.

An autoencoder is a typical deep learning method. There are many successful applications of autoencoder-based models in speech feature parameter extraction [27], facial recognition [28] and other fields [29]. It has been verified in our practice that autoencoders work well when used in soft sensor modeling. Therefore, a deep neural network that consists of several autoencoders is adopted as the data-driven submodel in this study. Furthermore, we improve the reconstruction loss function of the autoencoder and adopt a more effective parameter optimization algorithm to achieve better performance.

Our contributions are as follows.

- A new knowledge- and data-driven soft sensor is proposed, named the L-SAE model, that integrates a knowledge-driven submodel and a data-driven submodel. The loss function is improved by combining the two parts to maximize the predictive accuracy and improve the convergence speed.
- The established model is introduced to estimate the rotor deformation of an air preheater. This is the first application of the knowledge- and data-driven model in this aspect from a review of the searchable literature.

The remainder of this paper is arranged as follows: The application background for predicting the rotor deformation of an air preheater is introduced in Section 2. The proposed L-SAE model is detailed in Section 3, including the two submodels and the overall model. In Section 4, the experiments and results analysis are presented. Finally, our work is summarized in Section 5.

II. APPLICATION BACKGROUND

In the boilers of thermal power plants, an air preheater is a heat exchange device that utilizes the heat of the flue gas to preheat the combustion-supporting air. In this way, the heat of the flue gas is recycled, the temperature of the flue gas is reduced and the boiler efficiency is improved.

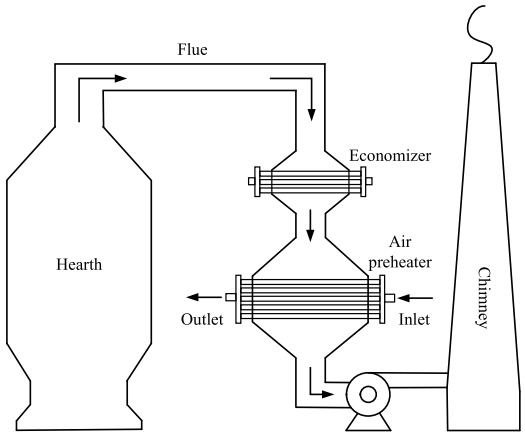


FIGURE 1. The role of the air preheater.

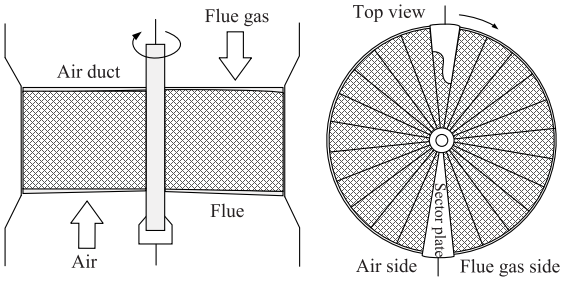


FIGURE 2. The top view of a rotary air preheater.

Meanwhile, the increase in the combustion air temperature is beneficial for fuel ignition and combustion to reduce the loss of incomplete combustion. The role of the air preheater in the boiler is shown in Fig. 1. Air preheaters fall into two categories according to the heat transfer method: tubular air preheaters and rotary air preheaters. Due to the advantages of small volume, light weight and easy installation, rotary air preheaters have gained more popularity in the large power units of thermal power plants. The top view and the stereogram of a rotary air preheater are shown in Fig. 2 and Fig. 3. A rotary air preheater consists of a fixed cylindrical housing, a cylindrical rotor, a flue, an air duct and a transmission device. The rotor is divided into a number of fan-shaped compartments, each of which is filled with heat storage panels made of undulating sheet metal. The cylindrical housing is divided into flue and air ducts to separate the flue gas and combustion-supporting air. To separate the air duct from the flue more strictly, two sector plates are installed on the bottom and the top of the air preheater.

In the rotation process of the rotor, taking a fan-shaped region of the rotor as an example, when the region rotates to the flue gas side, the hot flue gas flows through the flue from the top down and passes through the heat storage panels inside the rotor simultaneously. In this way, the heat of flue gas is transferred to the heat storage panels. As the rotor continues to rotate to the air side, the cold air flows from the bottom up through the air duct and absorbs the heat of the storage panels [30]. Once the rotor turns a full circle, the air preheater completes one heat exchange.

However, in the process of heat exchange, the temperature of the top preheater is high, while the temperature of the bottom is low. The unevenness of the heat causes a mushroom deformation of the rotor under the action of the thermal stress [31]. Some gaps appear between the rotor and the sector plates, as shown in Fig. 4. Both the flue gas and the air leak toward each other through the gaps. The hazards caused by air leakage are great. The heat transfer efficiency of the air preheater is reduced, which leads to lower boiler efficiency. Moreover, air leakage causes a large increase in auxiliary power usage that results in a direct economic loss. Taking a 600 MW coal-fired unit as an example, the generator sets consume approximately 2 million RMB of additional energy a year for each 1% rise in the air leakage rate.

Obviously, it is of great significance to measure and control the rotor deformation in real time. In practical applications, the measurement error of the rotor deformation should be less than 0.5 mm. There are two traditional methods to measure the rotor deformation in air preheaters. One is to use a mechanical probe for the measurement, and the other is to install a noncontact sensor on the sector plate. The mechanical probe has a satisfactory measurement accuracy in performing contact measurements. However, to avoid wearing the rotor, the mechanical probe can only make discrete measurements, and the measurement period is generally up to 4 hours. In addition, mechanical plugging and precision attenuation can occur on the mechanical probe after long-term use. A noncontact sensor must be installed in a confined space on the sector plate and work under severe conditions, which is costly. In the high temperature, corrosive and dusty environment where the air preheaters are located, the sensor often malfunctions. In this case, the whole air preheater must be stopped to maintain the failed sensor. The routine maintenance and overhaul of a sensor are both inconvenient.

Based on the above, it is difficult to measure the rotor deformation directly and timely. Therefore, we introduce the soft sensor technique to construct a predictive model to estimate the rotor deformation. Then, the output of the model is adopted as the control signal of the process control system and corresponding control measures can be adopted.

At present, there are few relevant studies available. In [32] and [33], mathematical models based on thermal stress theory are adopted to analyze the thermal deformation of the air preheater rotor, which are typical KDMs. First, the finite difference method is used to calculate the temperature distribution of the air preheater. Then, the finite element method is used to obtain the thermal deformation of the rotor. The authors stated that they achieved satisfactory results, but the differential problem became complicated when the temperature increased. In addition, these methods have time lag problems causing real-time prediction to be infeasible.

III. METHODS

A. KNOWLEDGE-DRIVEN SUBMODEL

There are complex lattice structures and heat storage components of various shapes and sizes inside the air preheater, and

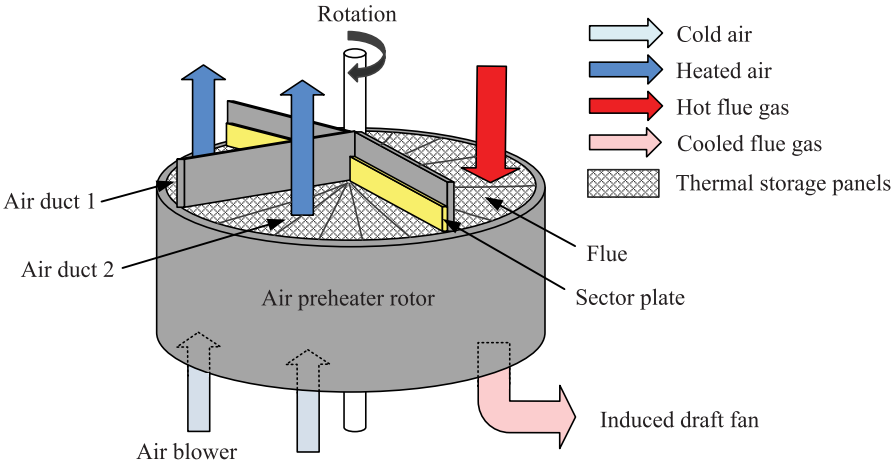


FIGURE 3. The stereogram of a trisector rotary air preheater.

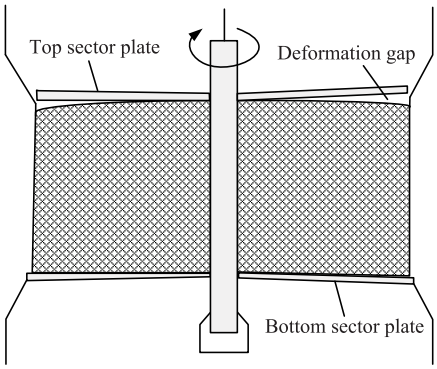


FIGURE 4. Sketch of rotor deformation in an air preheater.

the rotor deformation is a complex combination of the thermal deformations of these components. Therefore, in any working condition, the relationship between the rotor deformation and input variables is complex and nonlinear. However, air preheaters are located in harsh environments with high temperatures and corrosion, leading to a complex and unstable air preheating process. According to the existing research, there is no exact relationship between rotor deformation and other factors.

In this study, the most common design formula for predicting the rotor deformation of an air preheater, herein called the Lab model, is used as the knowledge-driven submodel. The formula is based on the knowledge of technical experts about industrial processes as follows:

$$Y = 0.006\frac{\Delta T R^2}{H} \tag{1}$$

where Y represents the maximum thermal deformation of the air preheater rotor. R denotes the radius of the rotor, and H is the height of the rotor. The temperature difference between the cold and hot ends is denoted as ΔT and is defined as follows:

$$\Delta T = \frac{T_{gi} + T_{ao}}{2} - \frac{T_{go} + T_{ai}}{2} \tag{2}$$

where T_{gi} is the inlet temperature of the gas flue, T_{ao} is the outlet temperature of the air duct, T_{go} is the outlet temperature of the flue, and T_{ai} is the inlet temperature of the air duct.

This formula is the most accurate existing KDM for predicting the deformation of an air preheater rotor. However, we found that this model behaved unsteadily in the field test. The prediction error is sometimes quite large. We need to adjust the value obtained from this model when applying it to an actual operation. Therefore, this model is not entirely suitable for dynamic prediction and control. The main limitations found in practice are listed below.

1) The air preheater rotor is an energy storage link with a complex structure, and its mathematical model must be a dynamic model rather than a simple static functional relationship.
2) In addition, the rotor deformation is also affected by the thermal deformation of the shell, main beam of the air preheater, center barrel, supporting shaft and other related mechanical components. However, equation (1) only expresses part of the steady-state characteristics of the system.
3) Over time, the coal ash that accumulates between the sealing sheets significantly increases, which increases the resistance of the air preheater, decreases the heat transfer capacity, and indirectly affects the thermal deformation law. The thermal deformation of the air preheater thus has a time-varying characteristic.

In summary, the thermal deformation model of an air preheater is a nonlinear dynamic system with inertia. Therefore, a knowledge-based model alone is insufficient to accurately predict the rotor deformation.

B. DATA-DRIVEN SUBMODEL

1) TRAINING PROCESS OF THE SAE

A deep neural network based on SAE is adopted as the data-driven submodel in this study. An autoencoder is a three-layer neural network designed to reconstruct the input data.

It consists of two parts: an encoder for extracting the features of the input data and a decoder for reconstructing the input data from the features. First, the encoder obtains the representation code h, that is, the features, as follows:

$$h = s(Wx + b) \tag{3}$$

where x is the input data. W and b are the weight and bias parameters of the encoder, respectively. $s(.)$ is the sigmoid function defined as follows:

$$s(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

Then, h is sent to the decoder as input, and the decoder yields the reconstructed data, $\hat{x}$, according to (5):

$$\hat{x} = s(W'h + b') \tag{5}$$

where W' and b' are the parameters of the decoder. Obviously, the higher the similarity is between $\hat{x}$ and x, the more important h is. Therefore, we optimized the parameter set, $\theta = \{W, b, W', b'\}$, to minimize the reconstruction error between x and $\hat{x}$. The optimal network parameters and the most important features of x are obtained when the reconstruction loss is minimal. In this paper, the loss function, L, is defined differently from the previous research as follows:

$$L(\theta) = L_{recon}(\theta) + L_{weight}(\theta) \tag{6}$$

where $L_{recon}(\theta)$ is the reconstruction loss, which is defined based on the cross-entropy as follows:

$$L_{recon}(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}log(\hat{x}_{ij}) + (1 - x_{ij})log(1 - \hat{x}_{ij})) \tag{7}$$

where m denotes the number of input samples and n denotes the input dimension. In comparison with the mean square error (MSE), cross entropy is not sensitive to outliers and behaves steadily in the convergence process, which is more suitable for designing loss functions. $L_{weight}(\theta)$ is a weight decay regularizer to ensure that L is a strictly convex function. We found in practice that the convergence process is more smooth by adding the weight decay term when optimization algorithms that are more developed than the gradient descent algorithm are adopted. In addition, it can prevent overfitting. We define $L_{weight}(\theta)$ as follows:

$$L_{weight}(\theta) = \frac{\lambda}{2}\sum_{l=1}^{2}\sum_{i=1}^{k_l}\sum_{j=1}^{k_{l+1}}(w_{ji})^2 \tag{8}$$

where k_l is the neuron number in layer l and w_{ji} denotes the weight between the jth neuron of layer $l+1$ and the ith neuron of layer l. λ is the decay coefficient.

Then, the network parameters are optimized by using the optimization algorithm. There are various existing optimization algorithms with their own advantages and drawbacks. In this paper, an advanced limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) [34] is adopted, and we will describe it in a later section.

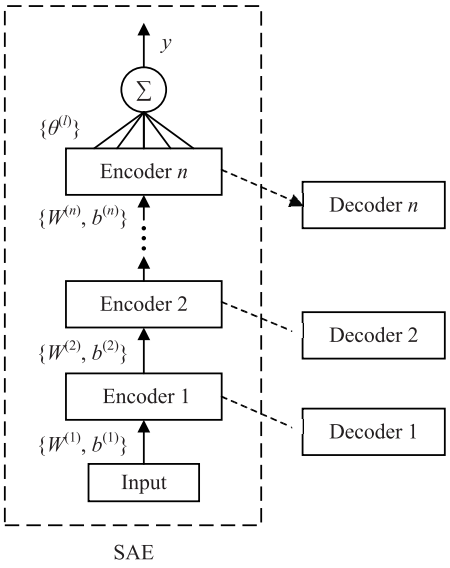


FIGURE 5. Training process of the SAE.

Obviously, the training of a single autoencoder is unsupervised since it uses only unlabeled data. When the first autoencoder is trained, the first level code of x is obtained. After that, the decoder is removed, and the code h is taken as the input of the second autoencoder. The second level code is extracted in the same way. In this way, a deep neural network is constructed by stacking autoencoders. The training process of the SAE is shown in Fig. 5. The number of network layers needs to be determined experimentally because there is no unified approach. If the SAE consists of n hidden layers, the topmost output of the hidden layer is as follows:

$$y^{(n)} = s(x, \theta^{(n)}) \tag{9}$$

where $\theta^{(n)}$ consists of the overall deep neural network parameters as follows:

$$\theta^{(n)} = \left\{W^{(1)}, b^{(1)}, \ldots, W^{(n)}, b^{(n)}\right\} \tag{10}$$

where $W^{(n)}$ and $b^{(n)}$ denote the parameters of the nth autoencoder. Considering the regression requirement of the soft sensor, we add a linear regression to the top of the model. The final output of the SAE is as follows:

$$y = \theta^{(l)T} y^{(n)} \tag{11}$$

where $\theta^{(l)}$ consists of the parameters for the regression operation which is optimized by the gradient descent method using the labeled data in the training set.

2) PARAMETER OPTIMIZATION OF THE SAE

Better parameters can improve the model accuracy. The parameter optimization of the SAE is divided into two stages. The first stage is the parameter optimization in the training process of a single autoencoder, and the second stage is the supervised fine-tuning of the whole network.

In the first stage, the L-BFGS algorithm is adopted, which is a quasi-newton algorithm with a high convergence speed

and strong global search capacity. During the training process of an autoencoder, the parameters are first initialized and then optimized through the L-BFGS algorithm. When the optimization is completed, the parameters are maintained, and the next autoencoder is trained. The steps of the L-BFGS algorithm are shown below.

Algorithm 1 Procedure of the L-BFGS Algorithm

1: Initialize θ_0 as $\theta_0^{ij} \sim [-\sqrt{6/N+M}, \sqrt{6/N+M}]$, $(\theta \in R^{N\times M})$
2: **for** k=1,2...until end conditions are satisfied **do**
3: Evaluate $g_k = \nabla f(\theta_k)$
4: $s_{k-1} = \theta_k - \theta_{k-1}, y_{k-1} = g_k - g_{k-1}$
5: $H_k^0 = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} I$
6: **for** i=$k-r,\ldots,k-1$ **do**
7: $s_i = \theta_{i+1} - \theta_i$
$y_i = g_{i+1} - g_i$
$\rho_i = \frac{1}{y_k^T s_k}$
$V_i = I - \rho_i y_i s_i^T$
8: **end for**
9: $H_k = (V_{k-1}^T \ldots V_{k-r}^T) H_k^0 (V_{k-r} \ldots V_{k-1}) + \rho_{k-r}(V_{k-1}^T \ldots V_{k-r+1}^T) s_{k-r} s_{k-r}^T (V_{k-r+1} \ldots V_{k-1}) + \rho_{k-r+1}(V_{k-1}^T \ldots V_{k-r+2}^T) s_{k-r+1} s_{k-r+1}^T (V_{k-r+2} \ldots V_{k-1}) + \ldots + \rho_{k-1} s_{k-1} s_{k-1}^T$
10: Set learning rate α_k
$\theta_{k+1} = \theta_k + \alpha_k H_k g_k$
$k \leftarrow k+1$
11: **end for**

Where H_{k+1} is the new inverse Hessian approximation obtained by updating H_k. In the L-BFGS algorithm, a modified version of H_k, $\{s_k, y_k\}$, is stored instead of the whole inverse Hessian matrix. The matrix is updated r times using the r most recent correction pairs, $\{s_i, y_i\}_{i=k-r}^{k-1}$, to reduce the cost of each iteration. Therefore, the L-BFGS algorithm has strong robustness and a high execution speed.

Based on the above, the supervised fine-tuning of the deep network is implemented using labeled data. In this stage, the stochastic gradient descent algorithm is adopted to realize end-to-end learning.

C. L-SAE MODEL

Based on the above research, a knowledge- and data-driven soft sensor is proposed, named the L-SAE model, which consists of the two submodels mentioned above to estimate the deformation of an air preheater rotor. A schematic diagram of the model is shown in Fig. 6. The knowledge-driven submodel is established based on the available domain knowledge, and the data-driven submodel is constructed solely based on data that reflect the actual conditions of the industrial installation. The two submodels are combined to capture the air preheating process more completely. The L-SAE model can be expressed by integrating the two submodels as

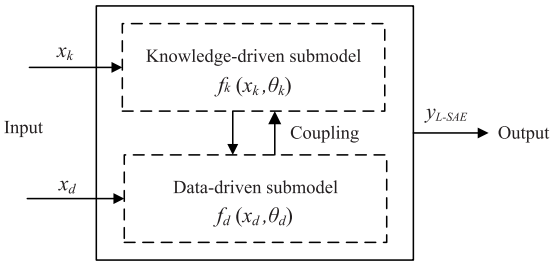


FIGURE 6. Schematic diagram of the L-SAE model.

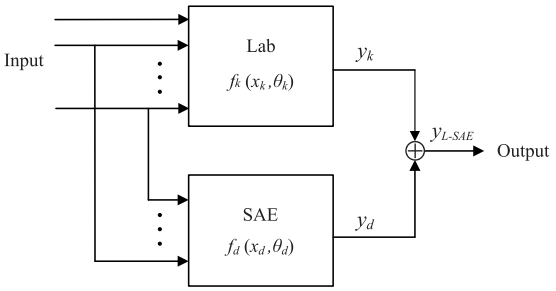


FIGURE 7. The L-SAE model with superposition coupling.

in the following mathematical formula:

$$y_{L-SAE} = f_k(x_k, \theta_k) \oplus f_d(x_d, \theta_d) \tag{12}$$

where y_{L-SAE} is the output vector of the model. f_k and f_d are the functions that are associated with the KDM and DDM submodels, respectively. θ_k and θ_d are the parameter vectors that are associated with the functions f_k and f_d, respectively. The symbol "$\oplus$" represents the coupling operation between the KDM and DDM submodels.

There are various ways of forming the coupling connections to support flexibility in combining the two submodels. Due to the diversity of different industrial knowledge, there is no generic approach for designing the coupling operation between the two submodels. For practical purposes, we set our coupling connection as a simple and commonly used superposition coupling operator through verification in practice, as shown in Fig. 7. The mathematical expression of this operator is given as follows:

$$y_{L-SAE} = \rho\,[f_k(x_k, \theta_k)] + (1-\rho)\,[f_d(x_d, \theta_d)] \tag{13}$$

where ρ is the weight coefficient that is used to balance the two parts of the L-SAE model. The optimal model performance can be obtained when the value of ρ is set appropriately. We found in practice that when the value of ρ is inappropriate, the model performance of the L-SAE is worse than each of the two submodels. But when ρ is set appropriate, the performance of the L-SAE can be improved significantly. That is, the linear superposition way of the two submodels behaves unreliable somtimes. Therefore, in the future work, we need to develop the more stable and reliable coupling connection way of the two submodels. Here, we set ρ to 0.25 based on repeated experiments. Together with

(1) and (11), equation (13) can be simplified as follows:

$$y_{L-SAE} = \rho Y + (1-\rho)y \tag{14}$$

This is the final output of the overall model.

IV. EXPERIMENTS

In this section, to verify the feasibility and effectiveness of the L-SAE, the established soft sensor is applied to the rotor deformation prediction of an air preheater in a thermal power plant boiler. Then, the model performance of the L-SAE is compared to that of other three common KDM and DDM approaches. The experiments are implemented on a PC with an Intel (R) Core (TM) i7 3.60 GHz processor with 8 GB RAM using MATLAB 2018a.

A. EXPERIMENTAL SETTINGS

1) SETTINGS OF THE SAE

In the process of soft sensor modeling, the selection of the input variables has an important impact on the prediction performance of the model. Using input variables that have no or low correlation with the target variable will negatively affect the model performance. Therefore, among the variables available, we need to choose those that have a strong correlation with the target variable as the input variables. Currently, the selection of the soft sensor input variables is generally based on mechanism knowledge and analysis experience. In our experiment, the alternative variables include the rotation speed of the rotor, the inlet and outlet temperatures of the flue and the air duct, the pressure and flow rate of the flue gas and the air. The rotor rotation is transverse, so there is no correlation between the rotation speed and the thermal deformation. The pressure and the flow rate of the flue gas and the air are correlated with the rotor deformation, but they are also correlated with the temperatures. Finally, we choose four temperature variables that are most correlated with the rotor deformation as the input variables of the SAE, including the inlet temperature of the air duct, T_{ai}; the outlet temperature of the air duct, T_{ao}; the inlet temperature of the flue, T_{gi}; and the outlet temperature of the flue, T_{go}.

In addition, the hyperparameters of the SAE are determined experimentally, and the model structure is set to 4-2-2-1. The final model structure of the L-SAE model is shown in Fig. 8.

2) DATA PREPROCESSING

All the historical data used in the experiment came from the real processes of a thermal power plant in western China. The input temperature samples are collected by sensors positioned at the inlet and outlet of the air duct and the flue. The target rotor deformation samples are collected by the noncontact sensor mounted on the sector plate.

To eliminate unit limitations, all the samples are normalized to [0,1] as follows:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{15}$$

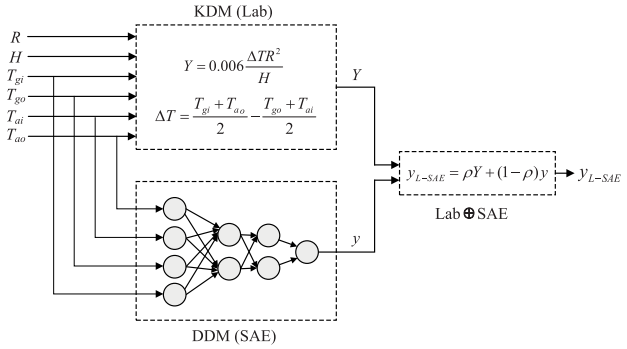


FIGURE 8. The structure of the L-SAE model.

In addition, considering the negative effect on the model performance of the outliers in the original dataset, the Hampels method [35], which outperforms the popular 3σ method in our practice, is adopted to eliminate the outliers. Samples satisfying (16) are flagged as outliers and removed.

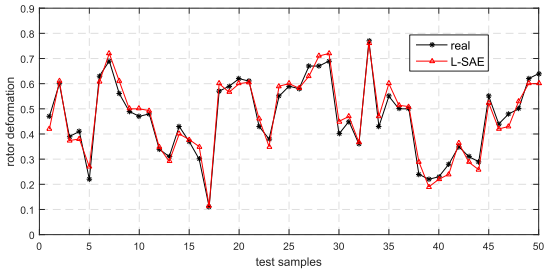
$$\frac{|x_i - x_{0.5}|}{1.4826 \times median(|x_i - x_{0.5}|)} > 3 \tag{16}$$

where x_i denotes the ith input sample and $x_{0.5}$ denotes the median value of x_i.

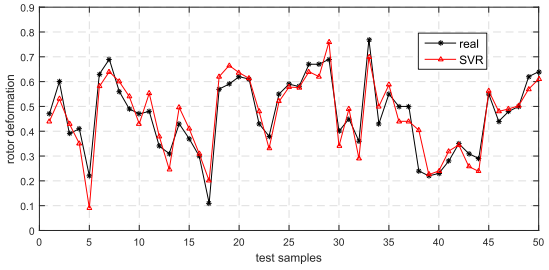
It is worth noting that when deep learning is used for image processing or speech recognition, it involves processing huge amount of data. But in the industrial case used in this paper, collecting samples is very difficult due to complex reasons such as high temperature and dust, which causes the data set size is relatively small. To ensure the high quality of the samples, only 500 samples are selected after the preprocessing. That is, the dataset used for experiment consists of 400 unlabeled samples and 100 labeled samples, which is divided into the training dataset and test dataset. The training dataset contains 400 unlabeled samples and 50 labeled samples, while the remaining 50 labeled samples serve as the test dataset.

B. RESULTS AND DISCUSSION

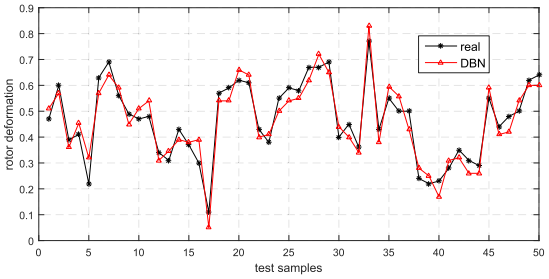
To verify the superiority of the proposed method, the other five commonly used soft sensor models were used to predict the dynamic rotor deformation of the air preheater for comparison, including the support vector machine regression (SVR), the NNPLS integrating PLS with a neural network, the plain deep neural network (DNN) with two hidden layers, the deep beliefe network (DBN) with two hidden layers and the Lab model. The root mean square error (RMSE) is used to measure the predictive errors of the soft sensors. The same dataset is used for the training and test of the six different models to ensure a fair comparison, and the final error of each model is the average of five tests. To highlight the contrast effect, the comparative results of four of the six methods are shown in Fig. 9, and the RMSE results of all the methods are shown in Table 1. In Fig. 9, the vertical axis represents the rotor deformation, of which the unit is millimeter in the original samples. Because the data has been normalized to



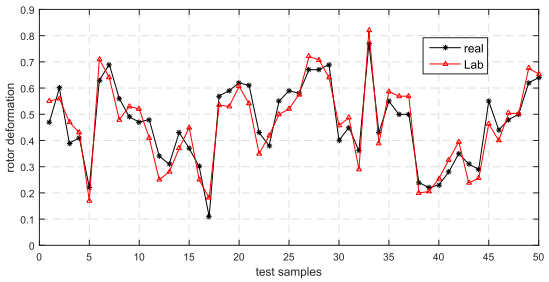
(a)



(b)



(c)



(d)

FIGURE 9. Comparison of the prediction results of four different soft sensors.

[0,1] in the data preprocessing stage, the unit of the vertical axis is not marked.

We can see from Fig. 9 and Table 1 that the errors of all the models are far less than the required 0.5 mm. The output data of the L-SAE more closely resembles the real data compared to that of the other models, which means a better nonlinear fitting ability. Second to the L-SAE, the SVR behaves best due to the strong capability of nonlinear approximation. The performance of the DBN and the DNN are not satisfied at short of training data. As mentioned above, the dataset size in this study is small, which is far from enough for the training of deep neural networks. In contrast, the L-SAE performs better under the same condition. The NNPLS model has the largest test error, showing that the PLS approach is not sufficiently powerful when dealing with highly nonlinear data. In addition, the L-SAE has a better ability to track extreme changes in the data. When the samples vary greatly, the L-SAE accommodates the changes better, and maintains a good performance.

TABLE 1. Experimental results of the six different models.

	Lab	SVR	NNPLS	DNN	DBN	L-SAE
Training RMSE	-	0.046	0.062	0.066	0.054	0.033
Test RMSE	0.071	0.057	0.082	0.081	0.067	0.045
Time(s)	-	33.7	82.8	288	134	117

From Table 1, according to the training and test errors of the six models, we can see that the L-SAE achieves the lowest error, while the NNPLS achieves the highest test error rather than the Lab model. This shows that not all the DDMs work better than the KDMs when used for soft sensor modeling. From the table, the L-SAE model outperforms the others. Second to the L-SAE, the SVR shows the best performance. The DBN performs medially, which shows that the pure deep learning methods are more suitable for modeling based on big data, but less powerful when handling small datasets. The performance of the Lab model is poor for the reasons detailed in the first part of Section 3. It is noteworthy that the DNN has the second highest errors after the NNPLS, showing the low efficiency when training the deep neural networks with more than four layers directly.

The experiment time shown in Table 1 contains training and test time of each method. It can be seen that the serious time lag and precision attenuation problems that cannot be solved by the traditional method are solved by the soft sensors. In total, the methods based on deep learning consume much more time than other methods, which shows that the deep learning methods have the slow convergence problem. However, we can see that the convergence speed of the L-SAE is much more faster than the DBN and the DNN due to the improvements of the training process and the optimization method. To clarify, the industrial process in this experiment is typical slow process with low sensitivity to time. Moreover, the soft sensors are trained off-line. Hence, compared with the training time, the test time, that is, the forward computation time of the models, is more important for soft sensor application. According to the experiment results, the test time of all the six methods is less than five seconds, which is quite satisfying.

Based on the above analysis, it is obvious that the L-SAE is feasible and performs better than the other common KDMs and DDMs. In our previous research [36], we established a soft sensor based on SAE with an SVR method to estimate the rotor deformation, which was a typical DDM based on deep

learning. The experiment results are slightly different because the datasets used in the two papers are not the same one. However, the results of the L-SAE model are good enough, proving the effectiveness of the knowledge- and data-driven approach for soft sensor modeling. In the future work, we will focus on improving the coupling way of the two submodels and modeling based on bigger datasets.

In summary, soft sensors, which effectively combine a KDM and DDM, can balance the advantages and disadvantages of the two methods. The industry process information is exploited more completely, and the performance of the soft sensor is improved.

V. CONCLUSION

In this paper, a new modeling method for soft sensors, named the L-SAE, based on knowledge and data is proposed. It is a hybrid approach that integrates a KDM and a DDM as two submodels. In this way, both the domain knowledge and industrial data are considered. In this paper, the Lab model is used as the knowledge-driven submodel, and a deep neural network based on SAE is used as the data-driven submodel. To verify the superiority of the proposed method, the established L-SAE model is applied to predicting the deformation of air preheater rotors in thermal power plant boilers. The L-SAE model achieves a higher predictive accuracy compared to that of the other five soft sensors, including the most commonly used KDM and DDM methods. Furthermore, this study developed the first knowledge and data-driven soft sensor for estimating the deformation of air preheater rotors in thermal power plant boilers. Our main work in the future is improving the coupling connection method between the two submodels and improving our approach with larger datasets.

REFERENCES

[1] Y. Liu, C. Yang, K. Liu, B. Chen, and Y. Yao, "Domain adaptation transfer learning soft sensor for product quality prediction," *Chemometrics Intell. Lab. Syst.*, vol. 192, p. 103813, Sep. 2019.

[2] J. F. de Canete, P. D. Saz-Orozco, R. Baratti, M. Mulas, A. Ruano, and A. Garcia-Cerezo, "Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network," *Expert Syst. Appl.*, vol. 63, pp. 8–19, Nov. 2016.

[3] W. Shao, S. Chen, and C. J. Harris, "Adaptive soft sensor development for multi-output industrial processes based on selective ensemble learning," *IEEE Access*, vol. 6, pp. 55628–55642, 2018.

[4] X. Yuan, Z. Chen, and Y. Wang, "Probabilistic nonlinear soft sensor modelingbased on generative topographic mappingregression," *IEEE Access*, vol. 6, pp. 10445–10452, 2018.

[5] K. Joosten, S. X. Cohen, P. Emsley, W. Mooij, V. S. Lamzin, and A. Perrakis, "A knowledge-driven approach for crystallographic protein model completion," *Acta Crystallographica*, vol. 64, no. 4, pp. 416–424, 2010.

[6] Y. Liu, C.-P. Chou, J. Chen, and J.-Y. Lai, "Active learning assisted strategy of constructing hybrid models in repetitive operations of membrane filtration processes: Using case of mixture of bentonite clay and sodium alginate," *J. Membrane Sci.*, vol. 515, pp. 245–257, Oct. 2016.

[7] D. Wang, J. Liu, and R. Srinivasan, "Data-driven soft sensor approach for quality prediction in a refining process," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 11–17, Feb. 2010.

[8] B. O. S. Teixeira, W. S. Castro, A. F. Teixeira, and L. A. Aguirre, "Data-driven soft sensor of downhole pressure for a gas-lift oil well," *Control Eng. Pract.*, vol. 22, no. 1, pp. 34–43, 2014.

[9] Y. Masuda, H. Kaneko, and K. Funatsu, "Multivariate statistical process control method including soft sensors for both early and accurate fault detection," *Ind. Eng. Chem. Res*, vol. 53, no. 20, pp. 8553–8564, 2014.

[10] J. Chai, C. Wu, C. Zhao, H. L. Chi, X. Wang, W. K. Ling, and K. L. Teo, "Reference tag supported RFID tracking using robust support vector regression and Kalman filter," *Adv. Eng. Informat.*, vol. 32, pp. 1–10, Apr. 2017.

[11] T. H. Pan, S. H. Wong, and S. S. Jang, "Development of a novel soft sensor using a local model network with an adaptive subtractive clustering approach," *Ind. Eng. Chem. Res.*, vol. 49, no. 10, pp. 4738–4747, 2010.

[12] X. Yuan, O. Chen, Y. Wang, C. Yang, and W. Gui, "Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE," *Neurocomputing*, to be published.

[13] B. Bidar, J. Sadeghi, and F. Shahraki, "Data-driven soft sensor approach for online quality prediction using state dependent parameter models," *Chemometrics Intell. Lab. Syst.*, vol. 162, pp. 130–141, Mar. 2017.

[14] S. Y. Park and C. H. Han, "A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns," *Comput. Chem. Eng.*, vol. 24, nos. 2–7, pp. 871–877, 2000.

[15] P. Zhu, X. Liu, Y. Wang, and X. Yang, "Mixture semisupervised Bayesian principalcomponent regression for soft sensor modeling," *IEEE Access*, vol. 6, pp. 40909–40919, 2018.

[16] W. Z. Sun and J. S. Wang, "Elman neural network soft-sensor model of conversion velocity in polymerization process optimized by chaos whale optimization algorithm," *IEEE Access*, vol. 5, pp. 13062–13076, 2017.

[17] B. Liang, X. Yuan, and Z. Ge, "Co-training partial least squares model for semi-supervised soft sensor development," *Chemometrics Intell. Lab. Syst.*, vol. 147, pp. 75–85, Oct. 2015.

[18] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1489–1496.

[19] J. Wang, Z. Cao, and B. Yang, "A mothed of improving identification accuracy via deep learning algorithm under condition of deficient labeled data," in *Proc. Chin. Control Conf.*, 2017, pp. 2281–2286.

[20] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, no. 9, pp. 60–88, Dec. 2017.

[21] F. Ugo, D. S. Alfredo, P. Francesca, Z. Paolo, and P. Francesco, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Inf. Sci.*, vol. 479, pp. 448–455, Apr. 2019.

[22] Z. Zhou and J. Gong, "Automated residential building detection from airborne LiDAR data with deep neural networks," *Adv. Eng. Informat.*, vol. 36, pp. 229–241, Apr. 2018.

[23] M. Havaei, A. Davy, D. Wardefarley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[24] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 30–42, Jan. 2012.

[25] A. Mnih and G. E. Hinton, "A scalable hierarchical distributed language model," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2008, pp. 1081–1088.

[26] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *J. Process Control*, vol. 24, no. 3, pp. 223–233, 2014.

[27] J.-Y. Kim, S.-J. Bu, and S.-B. Cho, "Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders," *Inf. Sci.*, vols. 460–461, pp. 83–102, Sep. 2018.

[28] R. Huanga, L. Chang, G. Lic, and J. Zhoub, "Adaptive deep supervised autoencoder based image reconstruction for face recognition," *Math. Problems Eng.*, vol. 2016, no. 5, pp. 1–14, 2016.

[29] W. Wei, H. Yan, Y. Wang, and W. Liang, "Generalized autoencoder: A neural network framework for dimensionality reduction," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 496–503.

[30] A. M. Duinea and P. M. Mircea, "Model in absolute units of the air preheater for operation simulation of high power steam generator," in *Proc. Int. Conf. Appl. Theor. Electr.*, 2014, pp. 1–4.

[31] H. Y. Wang, X. L. Bi, L. L. Zhao, Q. T. Zhou, H. T. Kim, and Z. G. Xu, "A study on thermal stress deformation using analytical methods based on the temperature distribution of storage material in a rotary air-preheater," *Appl. Thermal Eng.*, vol. 29, nos. 11–12, pp. 2350–2357, 2009.

[32] L. Wang, D. Lei, C. Tang, F. Qiang, C. Wang, and D. Che, "Thermal deformation prediction based on the temperature distribution of the rotor in rotary air-preheater," *Appl. Thermal Eng.*, vol. 90, pp. 478–488, Nov. 2015.

[33] S. Xie and F. Huang, "Prediction on thermal deformation of 800H alloy based on GRNN neural network," *Hot Work. Technol.*, vol. 90, no. 12, pp. 45–47, 2014.
[34] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1, pp. 503–528, 1989.
[35] L. Davies and U. Gather, "The identification of multiple outliers," *J. Amer. Stat. Assoc.*, vol. 88, no. 423, pp. 782–792, 1993.
[36] X. Wang and H. Liu, "Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction," *Adv. Eng. Inform.*, vol. 36, pp. 112–119, Apr. 2018.

XIAO WANG received the B.S. degree in electronic information from the Shandong University of Science and Technology, Shandong, China, in 2005, and the M.S. degree in electronic engineering from the Hubei University of Technology, Hubei, China, in 2011. She is currently pursuing the Ph.D. degree in control theory and control engineering with the Department of Automation and Information Engineering, Xi'an University of Technology, Xi'an, China. Her current research interests include soft sensor, deep learning, and neural network compression.

HAN LIU received the B.S. degree in automation from the Shannxi Mechanical Institute, Xi'an, China, in 1993, and the M.S. and Ph.D. degrees in control science and engineering from the Xi'an University of Technology, Xi'an, in 1996 and 2003, respectively. He is currently a Professor and the Dean of the School of Automation and Information Engineering, Xi'an University of Technology. His current research interests include industrial artificial intelligence, machine learning, pattern recognition, intelligent information processing, and nonlinear system control.