Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

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

Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE

Xiaofeng Yuan, Chen Ou, Yalin Wang*, Chunhua Yang, Weihua Gui

School of Automation, Central South University, Changsha 410083, China

ARTICLE INFO

Article history:
Received 31 March 2018
Revised 20 October 2018
Accepted 24 November 2018
Available online xxx

Key Words:
Deep learning
Soft sensor
Feature representation
Stacked autoencoder (SAE)
Hybrid variable-wise weighted SAE (HVW-SAE)

ABSTRACT

Soft sensors have been extensively used to predict difficult-to-measure quality variables for effective modeling, control and optimization of industrial processes. To construct accurate soft sensors, it is significant to carry out feature extraction for massive high-dimensional process data. Recently, deep learning has been introduced for feature representation in process data modeling. However, most of them cannot capture deep quality-related features for output prediction. In this paper, a hybrid variable-wise weighted stacked autoencoder (HVW-SAE) is developed to learn quality-related features for soft sensor modeling. By measuring the linear Pearson and nonlinear Spearman correlations for variables at the input layer with the quality variable at each encoder, a corresponding weighted reconstruction objective function is designed to successively pretrain the deep networks. With the constraint of preferential reconstruction for more quality-related variables, it can ensure that the learned features contain more information for quality prediction. Finally, the effectiveness of the proposed HVW-SAE based soft sensor method is validated on an industrial debutanizer column process.

1. Introduction

In industrial processes, it is crucial to carry out real-time process monitoring and control of product quality for safe and stable production [1–4]. However, it is difficult to measure the quality variable online for most processes due to lack of measuring devices, low analyzer reliability, large measurement delay and expensive cost, etc. Therefore, soft sensors have been largely designed to predict the difficult-to-measure quality variable through those easy-to-measure ones by some predictive mathematical models. In this way, soft sensors can provide real-time feedback information for process operators to take necessary actions for monitoring, control and optimization [5–13].

There are mainly two categories of soft sensors, which are first-principle models (FPMs) [14] and data-driven models (DDMs) [6]. FPMs are developed based on deep knowledge of process physicochemical backgrounds. However, it is often very cost, laborious and time-consuming to obtain process background knowledge. Sometimes, it is even impossible to gain accurate FPMs for modern large-scale industrial processes. With the development of computer science and data technology, a large number of sensor data can be collected from real operating plants for process data analysis and mining. Thus, data-driven soft sensors have been largely developed completely based on historical data collected within processes. In the past decades, data-driven soft sensors have gained more and more attention for process modeling since they do not need much process background knowledge. By far, a number of data-driven methods like principal component regression (PCR) [15], partial least squares (PLS) [16], gaussian process regression (GPR) [17], support vector regression (SVR) [18] and artificial neural network (ANN) [19], have been successfully applied to petrochemical, nonferrous metal and pharmaceuticals industries, etc. [20,21].

Usually, there are strong correlations and high redundancies between the abundant data in industrial plants. It is desirable and necessary to carry out feature learning from data for regression models. There are many feature extractors to capture intrinsically data structures, like linear principal component analysis (PCA) and partial least squares (PLS), nonlinear kernel PCA [22], kernel PLS [23] and ANN [24]. However, most of them are shallow learning architectures with no more than one hidden layer. Shallow learning is effective for simple data patterns but limited to describe complex data structures. Comparatively, multi-layer deep networks usually have more powerful representation for those complex problems. In 2006, Hinton et al. proposed the popular deep learning techniques with greedy layer-wise unsupervised pretraining for deep networks [25], which has raised increasing attention and achieved state-of-the-art results in many areas like image pattern recognition, speech recognition, natural language

* Corresponding author.
E-mail address: yuanxf@csu.edu.cn (X. Yuan).

https://doi.org/10.1016/j.neucom.2018.11.107
0925-2312/

processing, etc. Since then, many deep learning networks like stacked autoencoder (SAE) [26], deep belief network (DBN) [27] and convolutional neural network (CNN) [28] have been intensively researched in academia and extensively used in industrial areas. With the hierarchical layer structure of many nonlinearities, deep learning can learn deep abstract features for many kinds of complex tasks like classification and regression problems, which is more effective and efficient than shallow networks. Taking SAE for example, an asymmetric stacked autoencoder was proposed to design unequal number of encoders and decoders to improve the classification capacity of deep networks [29]. An stacked convolutional denoising autoencoders was proposed as an unsupervised deep network for feature representation, which can deal with the difficulty of obtaining labeled data [30]. Also, a coupled stacked denoising tensor autoencoder model was designed for feature representation in action recognition, which can solve the spatiotemporal corruption problem simultaneously [31]. For soft sensor applications, the deep abstract features can be directly used as predictors in regression models for the quality variable. In this way, deep learning can make soft sensors more robust and improve the prediction performance largely. However, it is not the best way to directly apply these deep learning networks for soft sensor modeling. Though high-level features can be learned for the raw observed input data with these deep learning networks, they cannot ensure the relevance of extracted features with the output variable. There may be much irrelevant information with the quality variable in the deep features. To solve this problem, a novel deep learning model was proposed for quality-related feature representation based on variable-wise weighted stacked autoencoder (VW-SAE) in our previous work [32]. In the pretraining of VW-SAE, the Pearson correlation coefficients are first calculated between the variables at the input layer with the quality variable in each autoencoder. Then, a variable-wise weighted reconstruction objective function is designed to train the AE with different weights for variables according to their correlations with the quality variable. By paying more attention to prior reconstructing variables with large weights, the learned features are more quality-relevant and suitable for prediction.

As can be seen, the success of VW-SAE mainly lies in the measure of variable correlations with the quality variable in each autoencoder in order to learn quality-related features. However, the main limitation of VW-SAE is that Pearson coefficient is a linear correlation measurement. It can only measure the linear correlation between two variables. Usually, there exist both linear and nonlinear relationships between variables in industrial processes. Thus, it is not sufficient to describe complex variable correlations. Also, the variable weights cannot be determined merely by their linear correlations with the quality variable in each autoencoder. Spearman correlation coefficient is a nonparametric measure of correlation, which is appropriate for process variables with complex nonlinear relationships. It is effective for nonlinear correlation analysis, and the measuring accuracy will greatly decrease for linear relationship. Since most industrial process variables usually have complex linear and nonlinear relationships with the quality variable due to specific physical and chemical mechanism, a single correlation measure cannot describe the complex data relationship accurately.

In this way, it is necessary to develop models that can help handle complex data correlations in VW-SAE. In this paper, a novel hybrid variable-wise weighted stacked autoencoder (HVW-SAE) algorithm is proposed for complex hierarchical quality-related feature representation. In HVW-SAE, both Pearson and Spearman correlation coefficients are exploited to measure linear and nonlinear correlations between input or feature variables with the quality variable at each autoencoder, respectively. Then, the two coefficients are combined to a hybrid one that can better describe variable relationships. After that, the hybrid coefficient is used to design the weighted reconstruction objective for pretraining of each autoencoder. As the hybrid coefficient take both linear and nonlinear variable correlations into consideration, the weighted reconstruction objective is more reasonable and proper to measure different importance of variables. Hence, HVW-SAE is able to learn deep features that contain more quality-related information from raw observed input data. Compared to the original VW-SAE, the design of the hybrid correlation is especially significant due to the hierarchical structure and the layer-wise pretraining of deep learning networks. This is because if the variable relationship is not fully considered in one autoencoder, the related information will be lost and cannot be recovered in the pretraining of the subsequent remaining autoencoders. In VW-SAE, only linear variable correlations are considered, in which the complex nonlinear relationships are missed. Hence, in the pretraining stage from low layers to high layers, there is important nonlinear quality-related information loss in each autoencoder. Due to the stacked network structure and layer-wise pretraining trick, the quality-related information loss will be accumulated from the input layer to the top feature layer in VW-SAE. Different from that, both linear and nonlinear variable correlations are sufficiently considered for feature learning at each autoencoder in HVW-SAE. Hence, quality-related features can be appropriately learned layer by layer with the proposed HVW-SAE network. For soft sensor applications, the extracted deep features are more suitable for building the predictive model as predictors. To validate the effectiveness of the proposed HVW-SAE, it is used for soft sensor modeling for an industrial debutanizer column process.

The remaining parts of this paper are organized as follows. In Section 2, we briefly review the related preliminaries of related works. Then, the hybrid variable-wise weighted stacked autoencoder is described in detail in Section 3. Also, the procedure of deep learning based soft sensor modeling is introduced in this section. Following this, the proposed HVW-SAE is validated on an industrial debutanizer column process in Section 4. Finally, conclusions are given in Section 5.

2. Preliminaries

2.1. Autoencoder

Autoencoder is a kind of unsupervised three-layer neural network with an encoder and a decoder, in which the target values at the output layer are set to be equal to its inputs. Suppose the inputs of the AE is $x = [x_{(1)}, x_{(2)} \ldots, x_{(d_x)}]^T \in R^{d_x}$, where d_x is the dimension of the inputs. The input x is mapped to the hidden representation $h \in R^{d_h}$ by function f

$$h = f(x) = s_f(Wx + b) \tag{1}$$

where W is a $d_h \times d_x$ weight matrix and $b \in R^{d_h}$ is the bias vector. In the decoder, the hidden representation h is mapped to the output layer of $\tilde{x} \in R^{d_x}$ by mapping function $\tilde{f}$

$$\tilde{x} = \tilde{f}\,(x) = s_{\tilde{f}}\left(\tilde{W}\,x + \tilde{b}\right) \tag{2}$$

where $\tilde{b} \in R^{d_x}$ is the bias vector term and $\tilde{W}$ is a $d_x \times d_h$ weight matrix for the output layer. Usually, the activation function s_f and s_g can be the commonly used sigmoid function or other functions. Hence, the parameter set of an autoencoder is $\theta = \{W, \widetilde{W}, b, \tilde{b}\}$. AE tries to learn a function $g_\theta(x) = \tilde{f}(f(x)) \approx x$. Denote the raw observed input dataset as $x_i \in \{x_1, x_2, \ldots x_N\}$. To obtain the model parameters, the reconstructed loss function is minimized by calculating the objective function as

$$J\left(W,\tilde{W},b,\tilde{b}\right) = \frac{1}{2N}\sum_{i=1}^{N}\|\tilde{x}_i - x_i\|^2 = \frac{1}{2N}\sum_{i=1}^{N}\|g_\theta(x_i) - x_i\|^2 \tag{3}$$

2.2. Stacked autoencoder

SAE can be constructed by hierarchically stacking multiple autoencoders, in which the raw input data are transmitted to the input layer of the whole SAE network. Also, the hidden layer of the first autoencoder is retained as the first hidden layer the SAE network. Then, the first hidden feature variables are served as the inputs of the second AE. After pretraining of the second AE, it hidden layer is kept as the second hidden layer of the SAE network. In a progressively way, the whole SAE can be constructed layer by layer. The last layer of the deep architecture obtains the final feature representation of the raw input data that can be used for complex prediction tasks.

2.3. VW-SAE

Variable-wise weighted stacked autoencoder (VW-SAE) is an improved variant of the traditional stacked autoencoder [13]. For the first autoencoder in SAE, Pearson correlation analysis is carried out for the raw input variables with the output variable, important quality-relevant variables are identified from other ones. To prior reconstruct quality-related variables, a variable-wise weighted reconstruction objective function is designed for first-level feature learning. As the first-level features are learned with the constraint that they should reconstruct quality-related input variables as good as possible, they contain more quality-relevant information. Meanwhile, irrelevant information is reduced. After the first VW-AE is constructed and trained, its feature variables are transmitted to the input layer of the second autoencoder. In a similar way, the second-layer quality-relevant features can be learned by prior reconstructing more important first-layer feature variables. The procedure can be carried out layer by layer until the top-layer features are learned.

3. Hybrid variable-wise weighted SAE

The original VW-SAE is able to capture quality-related features for regression application. However, it only focuses more on the linear variable relationship at each autoencoder since Pearson coefficient is a linear correlation measurement. The nonlinear variable correlations may be neglected and not properly retained at each layer of VW-SAE. To deal with this problem, a novel hybrid variable-wise weighted stacked autoencoder algorithm is proposed for better representation of output-related features. Different from the original VW-SAE, Hybrid correlation coefficient is designed by simultaneously measuring the Pearson and Spearman variable correlations, which can capture more complex data relationship between different variables.

3.1. Pearson correlation coefficient

The Pearson correlation coefficient is a linear measure between two variables. It is applicable to: (1) The linear relationship between two variables is continuous data. (2) The population of two variables is normal distribution, or near-normal unimodal distribution. (3) The observed values of two variables are paired, and each pair is independent of each other.

Assume the paired labeled training data as $\{X_h, Y_h\} = \{(x_1,y_1),(x_2,y_2),\ldots(x_j,y_j),\ldots(x_{N_h},y_{N_h})\}$, where $x_j = [x_{j(1)}, x_{j(2)}\ldots, x_{j(d_x)}]^T$. The Pearson correlation coefficient of the dth variable is calculated by the labeled training data as

$$\rho_d\left(X_{h(d)}, Y_l\right) = \frac{\sum_{j=1}^{N_h}\left(x_{j(d)} - \bar{x}_{(d)}\right)\left(y_j - \bar{y}\right)}{\sqrt{\sum_{j=1}^{N_h}\left(x_{j(d)} - \bar{x}_{(d)}\right)^2}\sqrt{\sum_{j=1}^{N_h}\left(y_j - \bar{y}\right)^2}} \tag{4}$$

where $\bar{x}_{(d)}$ and $\bar{y}$ are the means of the dth input and the target variables, respectively; $X_{h(d)}$ is the collection set of the dth input variable of the labeled data. That is to say, $X_{h(d)} = \{x_{1(d)}, x_{2(d)}, \ldots, x_{N_h(d)}\}$.

3.2. Spearman correlation coefficient

The Spearman correlation coefficient is a nonparametric measure of correlation and is appropriate for process variables with complex nonlinear relationship. It does not assume a normal distribution of data or any definite relationship between the data points [33]. Here, the Spearman correlation coefficient of the dth input variable with the quality variable is calculated as

$$\gamma_d\left(X_{h(d)}, Y_h\right) = 1 - \frac{6\sum c_{j(d)}^2}{N_h\left(N_h^2 - 1\right)} \tag{5}$$

$$c_{j(d)} = rg\left(x_{j(d)}\right) - rg\left(y_j\right) \tag{6}$$

where $rg(x_{j(d)})$ and $rg(y_j)$ are the ranks of each observation in the sample set $X_{h(d)}$ and Y_h ; $c_{j(d)}$ is the difference between the two ranks of each observation.

3.3. Hybrid correlation coefficient

Pearson correlation can only describe the linear relation of two variables while Spearman correlation coefficient is able to measure the nonlinear correlation between variables. To better capture the complex variable relationship, the hybrid correlation coefficient is designed as shown in formula (7). Here, k is a tuning parameter to adjust the ratio between linear and nonlinear correlation coefficient, which can vary between 0 and 1. If $k = 1$, only linear Pearson coefficient is used to describe variable relationship. Contrarily, nonlinear coefficient is used if k=0. Compared with the original correlation coefficient, the hybrid correlation coefficient can better reflect the real variable relation.

$$\lambda_{(d)}\left(X_{h(d)}, Y_h\right) = k|\rho_d| + (1-k)|\gamma_d| \tag{7}$$

3.4. Hybrid variable-wise weighted SAE

The traditional autoencoder aims to minimize the reconstruction error $\|x - \tilde{x}\|^2$ at each of the d_x dimensions equally. The objective function can be expressed as:

$$\|x - \tilde{x}\|^2 = \sum_{d=1}^{d_x}\left(x_{(d)} - \tilde{x}_{(d)}\right)^2 \tag{8}$$

It is easily seen that the reconstruction of each dimension should be kept very accurate to keep the overall reconstruction error small. Hence, all input variables occupy the same importance with each other in the reconstruction objective of autoencoder. However, these variables are very different since they have different correlations with the quality variable. If all variables are accurately reconstructed, the obtained feature usually contains information that is irrelevant for output prediction. To solve this problem, it is natural to learn features with the hope that the autoencoder can pay more importance reconstruct variables that are highly related for output prediction. Hence, it is necessary to measure the correlations of different input variables with the quality variable at each AE. Then, different weights can be assigned on different dimension of variables in the reconstruction error object. VW-AE can handle part of the problems since Pearson correlation is analyzed for different input variables. However, it is only a single linear correlation measurement, which is not sufficient for complicated data relationship. Thus, A hybrid VW-AE (HVW-AE) is further proposed in this part. In HVW-AE, Pearson and Spearman correlation are utilized to simultaneously measure the linear and nonlinear relationship. Then, they are integrated to a hybrid coefficient. A large hybrid coefficient represents more relevant of the corresponding input variable with the quality variable. Hence, it can be used to represent the importance of different input variables in AE. Then, a variable weighted objective function can be designed to pretraining the AE to learn nonlinear quality-related features. Depending on the types of training dataset, the AE can be trained in a supervised or semi-supervised way.

If all the training data are labeled samples, the following reconstruction objective function is designed for training of AE model.

$$J_\lambda(W,b) = \frac{1}{2N_h}\sum_{n=1}^{N_h}\sum_{d=1}^{d_x}\lambda_{(d)}\left(x_{n_{(d)}} - \tilde{x}_{n_{(d)}}\right)^2 = \frac{1}{2N_h}\sum_{n=1}^{N_h}(x_n - \tilde{x}_n)^T\Delta(x_n - \tilde{x}_n) \tag{9}$$

where Δ is a $d_x \times d_x$ diagonal matrix with its dth diagonal elements being $\lambda_{(d)}, d = 1, 2, \ldots, d_x$.

In many cases, the training dataset often includes unlabeled data that only has input part additional with labeled data. Denote the unlabeled data as $\{X_u\} = \{x_1, x_2, \ldots x_j, \ldots x_{N_u}\}$, where N_u is the number of unlabeled data samples. Then, HVW-AE is trained in a semi-supervised way by minimizing the variable-wise weighted reconstruction error on both labeled and unlabeled input data as

$$J_\lambda(W,b) = \frac{1}{2(N_h+N_u)}\sum_{n=1}^{N_h+N_u}\sum_{d=1}^{d_x}\lambda_{(d)}\left(x_{n_{(d)}} - \tilde{x}_{n_{(d)}}\right)^2 = \frac{1}{2(N_h+N_u)}\sum_{n=1}^{N_h+N_u}(x_n - \tilde{x}_n)^T\Delta(x_n - \tilde{x}_n) \tag{10}$$

For the pretraining, the gradient descent method can be used to minimize the objective function by iteratively updating parameter W and b, which is calculated as

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha\frac{\partial}{\partial W_{ij}^{(l)}}J_\lambda(W,b) \tag{11}$$

$$b_i^{(l)} = b_i^{(l)} - \alpha\frac{\partial}{\partial b_i^{(l)}}J_\lambda(W,b) \tag{12}$$

where α is the learning rate, $W_{ij}^{(l)}$ is the weight associated with the jth neuron in the lth layer and the ith neuron in the $(l+1)$th layer, $b_i^{(l)}$ is the corresponding bias from the lth layer to the ith neuron in the $(l+1)$th layer.

The key step is to compute the partial derivatives in the above equations. In detail, for a single training example $(x, \tilde{x})$ the back-propagation algorithm in an AE is depicted as:

(1) Perform a feedforward pass to compute the activations for the hidden layers l_2 and the output layers l_3 of the AE.

(2) For each output neuron i in layer l_3, calculate the following derivative terms

$$\delta_i^{(3)} = \frac{\partial}{\partial z_i^{(3)}}J_\lambda(W,b;x,\tilde{x}) = -\lambda_{(i)}\left(\tilde{x}_i - a_i^{(3)}\right)\cdot f'\left(z_i^{(3)}\right) \tag{13}$$

where $z_i^{(3)}$ is the weighted sum of the inputs to neuron i in l_3 layer, and $a_i^{(3)}$ is the activated value of the ith neuron of the l_3 layer.

(3) For each node i in layer l_2, compute the derivatives

$$\delta_i^{(2)} = \left(\sum_{j=1}^{s_3}W_{ji}^{(2)}\delta_j^{(3)}\right)f'\left(z_i^{(2)}\right) \tag{14}$$

where s_3 is the number of neurons in l_3 layer.

(4) Calculate the partial derivatives:

$$\frac{\partial}{\partial W_{ij}^{(l)}}J_\lambda(W,b;x,\tilde{x}) = a_j^{(l)}\delta_i^{(l+1)} \tag{15}$$

$$\frac{\partial}{\partial b_i^{(l)}}J_\lambda(W,b;x,\tilde{x}) = \delta_i^{(l+1)} \tag{16}$$

Once aforementioned formulas are obtained, derivatives of the overall cost function $J_\lambda(W, b)$ can be computed as:

$$\frac{\partial}{\partial W_{ij}^{(l)}}J_\lambda = \frac{1}{2(N_h+N_u)}\sum_{n=1}^{N_h+N_u}\frac{\partial}{\partial W_{ij}^{(l)}}J_\lambda(W,b;x_n,\tilde{x}_n) \tag{17}$$

$$\frac{\partial}{\partial b_i^{(l)}}J_\lambda = \frac{1}{2(N_h+N_u)}\sum_{n=1}^{N_h+N_u}\frac{\partial}{\partial b_i^{(l)}}J_\lambda(W,b;x_n,\tilde{x}_n) \tag{18}$$

Thus, back propagation algorithm can be iteratively carried out to train the network parameters.

3.5. Hybrid variable-wise weighted SAE

Multiple HVW-AEs can be stacked layer by layer to construct the deep HVW-SAE network, in which the hidden layer of the previous HVW-AE is used as the input layer of the latter one. Fig. 1 gives the structure of HVW-SAE. The main difference between the proposed hybrid variable-wise weighted SAE and the original SAE is that in each AE, hybrid variable weights is additionally calculated by the labeled data. Then, a hybrid variable weighted loss function is used to train this AE so that the extracted features are more relevant to the output variable. Assume HVW-AE l is pre-trained, then its hidden layer feature h^l will send to the input layer of the next HVW-AE $(l+1)$ to obtain its hidden feature h^{l+1}. To train this HVW-AE, the hybrid correlation coefficients are first calculated for each variable of h^l between the quality variable on the labeled data. Then, the coefficients are used to design the variable weighted reconstruction objective to pretrain HVW-AE $(l+1)$ as

$$J_\lambda^{l+1}(W,b) = \frac{1}{2N_h}\sum_{n=1}^{N_h}\sum_{d=1}^{d_l}\lambda_{(d)}^{l+1}\left(h_{n(d)}^{l+1} - h_{n(d)}^{l}\right)^2 \tag{19}$$

where d_l is the number of neurons in the h^l layer.

In a similar way, HVW-AE $(l+1)$ can be trained in a semi-supervised way by minimizing the hybrid variable-wise weighted reconstruction error on both labeled and unlabeled input data as

$$J_\lambda^{l+1}(W,b) = \frac{1}{2(N_h+N_u)}\sum_{n=1}^{N_h+N_u}\sum_{d=1}^{d_l}\lambda_{(d)}^{l+1}\left(h_{n(d)}^{l+1} - h_{n(d)}^{l}\right)^2 \tag{20}$$

As can be seen, different feature variables are assigned with different weights according to their hybrid correlations with the output variable at each layer. The weights are used to design the new hybrid weighted loss function, which pays more attention in accurate reconstruction for important variables. As the whole network is constructed and pre-trained layer by layer, the important quality-related information will be further retained and strengthened from low levels to high levels. Meanwhile, the

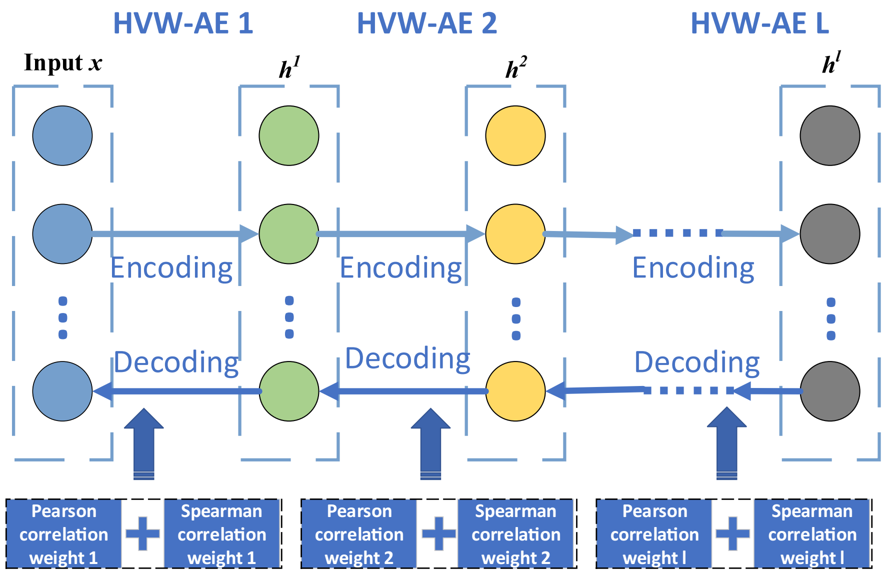

Fig. 1. The structure of HVW-SAE.

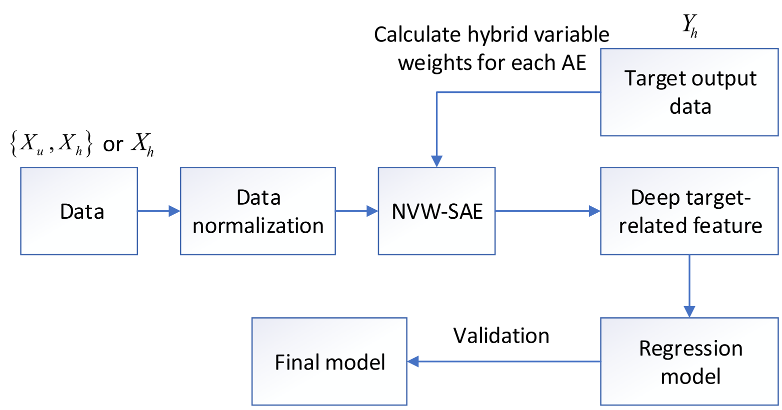

Fig. 2. HVW-SAE based soft sensor.

irrelevant information is reduced or even eliminated gradually from low layers to high layers in HVW-SAE.

3.6. Hybrid VW-SAE based soft sensor

As hybrid variable-wise weighted SAE is capable to learn high-level output-relevant features, it is very suitable for soft sensor modeling. Fig. 2 provides the basic framework for the proposed HVW-SAE based soft sensor. First, collect training input and output dataset from industrial processes $[X, Y_h]$. $X = [X_h, X_u]$ is composed of the labeled input data X_h and unlabeled input data X_u. Second, Normalize the data set $[X, Y_h]$. Then, both labeled and unlabeled data, or only labeled data are used to train the hybrid variable-wise weighted SAE in a semi-supervised or supervised manner. After that, deep output-related features are extracted for the raw input data. Next, the deep output-related features can be used construct the regression model for the quality variable. Hence, the regression model is just a two-layer neural network. That is to say, an output layer is added to the top layer of HVW-SAE for the quality variable. Then, we employ fine-tuning to slightly modify the features of the HVW-SAE according to the difference between the features and the labeled output data. At last, the trained HVW-SAE based soft sensor network can be used to predict the quality variable for new data samples.

Usually, the root mean squared error (RMSE) index is used to validate the effectiveness of soft sensor algorithms. It is defined as

$$\mathrm{RMSE} = \sqrt{\sum_{n=1}^{N_T} \left(y_n - \hat{y}_n\right)^2 / (N_T - 1)} \tag{21}$$

where y_n and $\hat{y}_n$ are the labeled and predicted output values of the nth testing sample, respectively; N_T is the number of testing samples. Another widely used index is the coefficient of determination R^2, which represents a squared correlation between the actual output and estimated output. R^2 can give information about how much of the total variance in the output variable data can be explained by the model. R^2 index is defined as

$$R^2 = 1 - \sum_{n=1}^{N_T} \left(y_n - \hat{y}_n\right)^2 / \sum_{n=1}^{N_T} (y_n - \bar{y})^2 \tag{22}$$

where $\bar{y}$ is the mean of labeled output values in the testing dataset.

4. Case study

In this section, case study on an industrial debutanizer column process is carried out to evaluate the performance of the proposed hybrid variable-wise weighted SAE. First, description about debutanizer column process is given in details. Then, HVW-SAE based

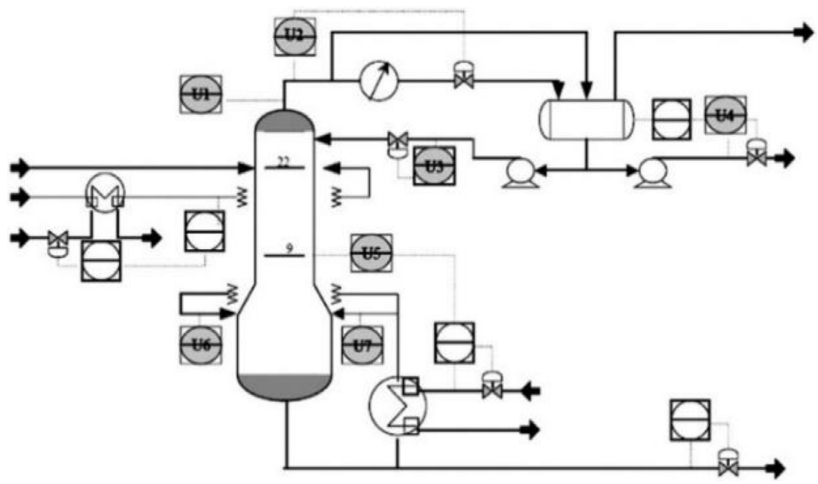

Fig. 3. The flowchart of the debutanizer column [35].

Table 1
Variable description for the debutanizer column.

Input variables u	Variable description
u_1	Top temperature
u_2	Top pressure
u_3	Reflux flow
u_4	Flow to next process
u_5	6th tray temperature
u_6	Bottom temperature A
u_7	Bottom temperature B
Output variable y	Butane (C4) content in the debutanizer column bottom

Table 2
The prediction performance of the five methods.

Method	RMSE	R^2
NN	0.0542	0.8918
SAE	0.0450	0.9192
VW-SAE-ρ	0.0389	0.9438
VW-SAE-γ	0.0336	0.9580
Hybrid VW-SAE	0.0308	0.9615

soft sensor modeling is carried out for debutanizer column product concentration prediction.

4.1. Description of the debutanizer column

The debutanizer column is a part of important processing units in a refinery process. It is used to remove the propane (C3) and butane (C4) from the naphtha stream. The flowchart of this debutanizer column is shown in Fig. 3. For product quality and process control, the butane content in the bottom product should be minimized. This requires real-time measurement of butane content in the bottom [34]. However, the gas chromatograph is installed in the overhead of the sequential deisopentanizer column, which is located some distance away from the debutanizer column. This often introduces a large measuring delay. To deal with the measuring problem, a soft sensor can be used to estimate the concentration of butane in the bottom in real time. To this end, seven process variables are selected for data modeling. Table 1 gives the detailed description of the seven input variables and the quality output variable.

4.2. Results and discussions

To construct the HVW-SAE model, process dynamics are taken into consideration. In reference [34], the authors have done a lot of experiments to obtain the optimal raw input variables in a trial and error way, additionally guided by expert knowledge and physical insight. Finally, the raw input variable vector for the HVW-SAE network is designed as

$$\begin{bmatrix} u_1(k), u_2(k), u_3(k), u_4(k), u_5(k), u_5(k-1), \\ u_5(k-2), u_5(k-3), (u_6(k)+u_7(k))/2, \\ y(k-1), y(k-2), y(k-3), y(k-4) \end{bmatrix}^T \tag{23}$$

This data set is shared by Fortuna et al. [16]. A total of 2000 input and output data samples are used in this study. For HVW-SAE network construction and testing, 1000 samples are used as training samples and the remaining ones are used for testing dataset.

The architecture of the HVW-SAE is as follows: 13 input nodes, three AEs with hidden neuron numbers as 10,7 and 4. The activation function is chosen as sigmoid function. Each HVW-AE model is trained with random batch gradient descent algorithm. The adjustable parameter k is set as 0.5. The batch size is set as 20 samples. After the HVW-SAE is layer-wise pre-trained, an output layer is added to the top layer of the HVW-SAE network for fine tuning. Then, the trained HVW-SAE based soft sensor model can be applied to the testing dataset for output prediction. Finally, the prediction RMSE and R^2 on the testing dataset are 0.0308 and 0.9615, respectively. The detailed prediction results are shown in Fig. 4. As can be seen, the predicted output values can generally track well with the real output values on the testing dataset. The main prediction errors lie in those data samples with extreme small or big output values.

To validate the effectiveness and flexibility of the proposed method, the original VW-SAE, NVW-SAE and SAE with the same network structure are also adopted for soft sensor modeling. Additionally, a multi-layer neural network (NN) with structure of [13 10 7 4 1] is trained to predict the output variable, in which the weights and bias are initialized with random values. Table 2 gives the prediction results of multi-layer NN, SAE, VW-SAE, NVW-SAE and HVW-SAE on the testing dataset. As can be seen from Table 2, NN gives the worst prediction results on the testing dataset due to its random initialization. For the four SAE based methods, they achieve better prediction accuracy since pretraining technique is utilized for feature learning from raw input data. However, VW-SAE outperform SAE since quality-relevance are utilized for feature learning. Moreover, HVW-SAE gives the best prediction performance as hybrid correlations are considered for quality-related feature learning.

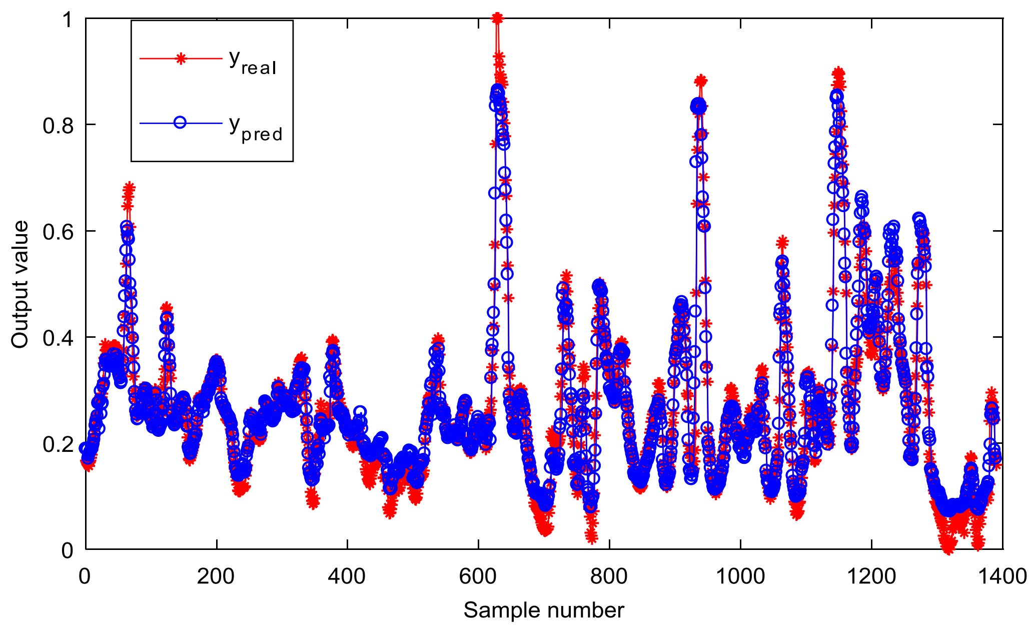

Fig. 4. The prediction performance on the testing dataset for HVW-SAE.

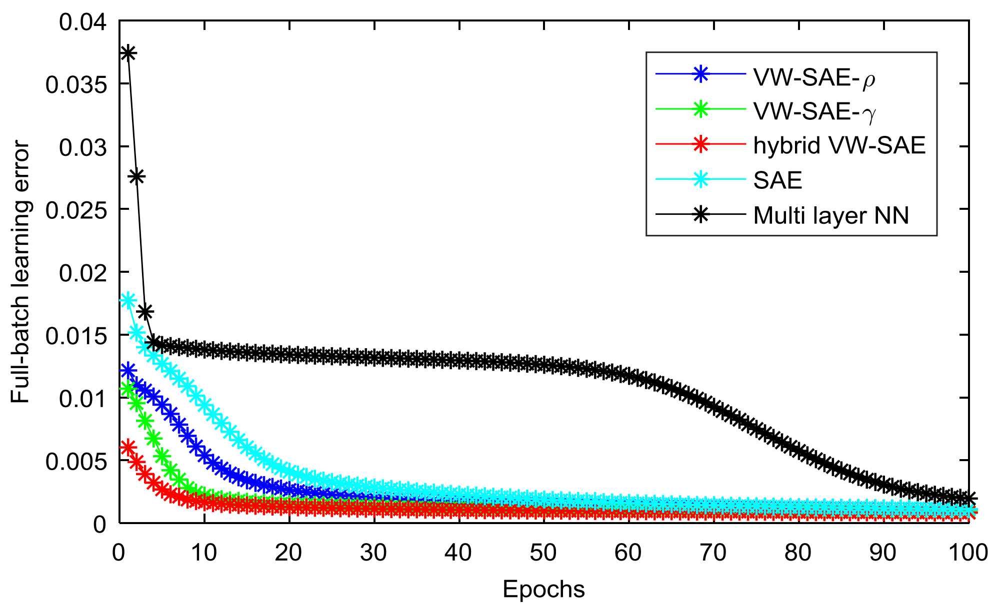

Fig. 5. Fine-tuning procedure of the five methods.

Furthermore, we have investigated the learning procedure of the five neural network-based soft sensor methods. Fig. 5 shows the full batch training errors for the five methods. From Fig. 5, it can be seen that the initial training error of HVW-SAE is much lower than the other methods. Moreover, HVW-SAE based methods can reach the convergent state much faster than other methods. Therefore, HVW-SAE is able to learn deep complex quality-relevant features and is suitable for soft sensor modelling.

5. Conclusion

In this paper, deep learning has been introduced for feature representation in soft sensor applications. Since traditional deep learning only focuses on feature representation for the raw input data, it cannot capture quality-related features for regression modeling. Hence, a novel hybrid variable-wise weighted stacked autoencoder algorithm is proposed for hierarchical complex quality-related feature representation layer by layer for industrial processes. In HVW-SAE, the hybrid correlation coefficient is utilized to measure both linear and nonlinear relationships of input and feature variables with the quality variable at each autoencoder. Then, a corresponding weighted reconstruction objective function is designed to learn nonlinear quality-related features. This method is compared with other four neural networks for soft sensor prediction, in which the results show its effectiveness and superiority over the others.

Conflict of interest

None.

Acknowledgement

This paper is supported in part by National Natural Science Foundation of China (NSFC) (61703440), in part by Major Project of the National Natural Science Foundation of China (61590921), in part by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (61621062), in part by the Natural Science Foundation of Hunan Province of China (2018JJ3687), and in part by Innovation-driven plan in Central South University (2018CX011).

Reference

[1] X. Yuan, Y. Wang, C. Yang, Z. Ge, Z. Song, W. Gui, Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes, IEEE Trans. Ind. Electron. 65 (2018) 1508–1517.
[2] Z. Ge, J. Chen, Plant-wide industrial process monitoring: a distributed modeling framework, IEEE Trans. Ind. Inform. 12 (2017) 310–321.
[3] L. Zhou, J. Zheng, Z. Ge, Z. Song, S. Shan, Multimode process monitoring based on switching autoregressive dynamic latent variable model, IEEE Trans. Ind. Electron. 65 (2018) 8184–8194.
[4] Q. Jiang, S. Ding, Y. Wang, X. Yan, Data-driven distributed local fault detection for large-scale processes based on GA-regularized Canonical Correlation analysis, IEEE Trans. Ind. Electron. 64 (2017) 8148–8157.
[5] S. Khatibisepehr, B. Huang, S. Khare, Design of inferential sensors in the process industry: a review of Bayesian methods, J. Process Control 23 (2013) 1575–1596.
[6] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, Comput. Chem. Eng. 33 (2009) 795–814.
[7] X. Yuan, L. Li, Y. Wang, Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, IEEE. T. Ind. Inf. (2019), doi:10.1109/TII.2019.2902129.
[8] X. Yuan, Z. Ge, Z. Song, Y. Wang, C. Yang, H. Zhang, Soft sensor modeling of nonlinear industrial processes based on weighted probabilistic projection regression, IEEE Trans. Instrum. Meas. 66 (2017) 837–845.
[9] Z. Ge, Z. Song, S.X. Ding, B. Huang, Data mining and analytics in the process industry: the role of machine learning, IEEE Access 5 (2017) 20590–20616.
[10] Q. Jiang, X. Yan, B. Huang, Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and bayesian inference, IEEE Trans. Ind. Electron. 63 (2016) 377–386.
[11] L. Zhou, G. Li, Z. Song, S.J. Qin, Autoregressive dynamic latent variable models for process monitoring, IEEE Trans. Control Syst. Technol. 25 (2017) 366–373.
[12] Z. Chen, S.X. Ding, T. Peng, C. Yang, W. Gui, Fault detection for non-gaussian processes using generalized canonical correlation analysis and randomized algorithms, IEEE Trans. Ind. Electron. 65 (2018) 1559–1567.
[13] X. Yuan, Y. Wang, C. Yang, W. Gui, L. Ye, Probabilistic density-based regression model for soft sensing of nonlinear industrial processes, J. Process Control 57 (2017) 15–25.
[14] B. Huang, Y. Qi, A.M. Murshed, Dynamic modelling and predictive control in solid oxide fuel cells: First principle and data-based approaches, John Wiley & Sons, 2012.
[15] X. Yuan, Z. Ge, B. Huang, Z. Song, Y. Wang, Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR, IEEE Trans. Ind. Inform. 13 (2017) 532–541.
[16] M. Ma, S. Khatibisepehr, B. Huang, A Bayesian framework for real-time identification of locally weighted partial least squares, AIChE J. 61 (2014) 518–529.
[17] X. Yuan, Z. Ge, B. Huang, Z. Song, A probabilistic just-in-time learning framework for soft sensor development with missing data, IEEE Trans. Control Syst. Technol. 25 (2017) 1124–1132.
[18] Y. Liu, J. Chen, Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes, J. Process Control 23 (2013) 793–804.
[19] J. Gonzaga, L. Meleiro, C. Kiang, R. Maciel Filho, ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process, Comput. Chem. Eng. 33 (2009) 43–49.
[20] X. Yuan, J. Zhou, Y. Wang, C. Yang, Multi-similarity measurement driven ensemble just-in-time learning for soft sensing of industrial processes, J. Chemom. 32 (2018) e3040.
[21] N. Chen, J. Dai, X. Yuan, W. Gui, W. Ren, H.N. Koivo, Temperature prediction model for roller kiln by ALD-based double locally weighted kernel principal component regression, IEEE Trans. Instrum. Meas. 67 (2018) 2001–2010.
[22] X. Yuan, Z. Ge, Z. Song, Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes, Ind. Eng. Chem. Res. 53 (2014) 13736–13749.
[23] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing kernel Hilbert space, J. Mach. Learn. Res. 2 (2002) 97–123.
[24] L. Fortuna, P. Giannone, S. Graziani, M.G. Xibilia, Virtual instruments based on stacked neural networks to improve product quality monitoring in a refinery, IEEE Trans. Instrum. Meas. 56 (2007) 95–101.
[25] Y. Bengio, P. Lamblin, D. Popovici, H. Larocchelle, Greedy layer-wise training of deep networks, in: Proceedings of the International Conference on Neural Information Processing Systems, 2006, pp. 153–160.
[26] J. Yu, C. Hong, Y. Rui, D. Tao, Multi-task autoencoder model for recovering human poses, IEEE Trans. Ind. Electron. 65 (2018) 5060–5068.
[27] S. Lee, J.H. Chang, Oscillometric blood pressure estimation based on deep learning, IEEE Trans. Ind. Inform. 13 (2016) 461–472.
[28] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, K.R. Muller, Evaluating the visualization of what a deep neural network has learned, IEEE Trans. Neural Netw. Learn. Systems. 28 (2016) 2660–2673.
[29] A. Majumdar, A. Tripathi, Asymmetric stacked autoencoder, in: Proceedings of the International Joint Conference on Neural Networks, 2017, pp. 911–918.
[30] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, D. Tao, Stacked convolutional denoising auto-encoders for feature representation, IEEE Trans. Cybern. 47 (2016) 1017–1027.
[31] C. Jia, M. Shao, S. Li, H. Zhao, Y. Fu, Stacked denoising tensor auto-encoder for action recognition with spatiotemporal corruptions, IEEE Trans. Image Process. (2018) 1878–1887.
[32] X. Yuan, B. Huang, Y. Wang, C. Yang, W. 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.
[33] J. Hauke, T. Kossowski, Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data, Quaest. Geogr. 30 (2011) 87–93.
[34] L. Fortuna, S. Graziani, M.G. Xibilia, Soft sensors for product quality monitoring in debutanizer distillation columns, Control Eng. Pract. 13 (2005) 499–508.
[35] L. Fortuna, S. Graziani, A. Rizzo, M.G. Xibilia, Soft Sensors for Monitoring and Control of Industrial Processes, Springer Science & Business Media, 2007.

Xiaofeng Yuan received the B.Eng. and Ph.D. degrees from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2011 and 2016, respectively.

He was a visiting scholar with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, from November 2014 to May 2015. He is currently an Associate Professor with the School of Automation, Central south University. His research interests include deep learning and artificial intelligence, machine learning and pattern recognition, industrial process soft sensor modeling, process data analysis, etc.

Chen Ou is a postgraduate student at the School of Automation, Central South University, Changsha, China. He received his B.Eng. degree in School of Electrical and Information Engineering from Hunan Institute of Engineering in 2017. His research interests include deep learning, machine learning, soft sensor modeling, process data mining, etc.

Yalin Wang received the B.Eng. and Ph.D. degrees from the Department of Control Science and Engineering, Central South University, Changsha, China, in 1995 and 2001, respectively.

Since 2003, she has been with the School of Information Science and Engineering, Central south University, where she was at first an Associate Professor and is currently a Professor. Her research interests include the modeling, optimization and control for complex industrial processes, intelligent control, and process simulation.

Chunhua Yang received the M.Eng. degree in automatic control engineering and the Ph.D. degree in control science and engineering from Central South University, Changsha, China, in 1988 and 2002, respectively. She was with the Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium, from 1999 to 2001. She is currently a Full Professor with Central South University. Her current research interests include modeling and optimal control of complex industrial process, intelligent control system, and fault-tolerant computing of real-time systems.

Weihua Gui received the degree of the B.Eng. and the M.Eng. at Department of Control Science and Engineering from Central South University, Changsha, China, in 1976 and 1981, respectively. From 1986 to 1988 he was a visiting scholar at Universität-GH-Duisburg, Germany. He has been a full professor in the School of Information Science & Engineering, Central South University, Changsha, China, since 1991. His main research interests include modeling and optimal control of complex industrial processes, distributed robust control, and fault diagnoses. He was elected as an academician of Chinese Academy of Engineering in 2013.

Please cite this article as: X. Yuan, C. Ou and Y. Wang et al., Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE, Neurocomputing, https://doi.org/10.1016/j.neucom.2018.11.107