Hindawi
Journal of Control Science and Engineering
Volume 2020, Article ID 6347625, 9 pages
https://doi.org/10.1155/2020/6347625

Research Article

Soft Sensor Modeling of Key Effluent Parameters in Wastewater Treatment Process Based on SAE-NN

Yousuf Babiker M. Osman[1,2] **and Wei Li**[1,2]

[1]*College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China*
[2]*Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou 730050, China*

Correspondence should be addressed to Wei Li; liwei@lut.cn

Received 3 December 2019; Revised 20 February 2020; Accepted 14 May 2020; Published 29 May 2020

Academic Editor: Radek Matušů

Real-time measurements of key effluent parameters play a highly crucial role in wastewater treatment. In this research work, we propose a soft sensor model based on deep learning which combines stacked autoencoders with neural network (SAE-NN). Firstly, based on experimental data, the secondary variables (easy-to-measure) which have a strong correlation with the biochemical oxygen demand (BOD5) are chosen as model inputs. Moreover, stochastic gradient descent (SGD) is used to train each layer of SAE to optimize weight parameters, while a strategy of genetic algorithms to identify the number of neurons in each hidden layer is developed. A soft sensor model is studied to predict the BOD5 in a wastewater treatment plant to evaluate the proposed approach. Interestingly, the experimental results show that the proposed SAE-NN-based soft sensor has a better performance in prediction than the current common methods.

1. Introduction

Recently, water pollution has been one of the most serious and ongoing problems facing our world. Key variables in wastewater treatment need to be evaluated in order to control pollution and ensure that water emission calculations are up to international standards.

Several methods have been used to calculate the key variables in the treatment of wastewater. However, in the wastewater treatment system, there are a large number of variables that are difficult to measure online, such as BOD5, which is calculated by a normal 5-day off-line delay. This makes it inappropriate for real-time measurement and may lead to effluent quality violations. Soft sensor technology provides a good solution to these problems [1–3]. Soft measurement estimates variables that are difficult to measure by correlating them with available variables that are easy to measure. Soft sensors can be categorized into two separate classes, namely, model-driven and data-driven, which can be differentiated on a very general level. The first-principle models are most often soft sensors model-driven family [4]. The model-driven (white-box model) is built based on the deep knowledge of process mechanism backgrounds. But, due to the complicated physical backgrounds and harsh conditions of industrial plants, when using first-principle approaches, it is difficult to model the entire process when that phase is considered. Nevertheless, data-driven (black-box models) are built on historical data that can be obtained from industrial processes and built without any operational experience or prior knowledge, making it an acceptable choice for soft sensor modeling of complex processes [5]. For the development of data-driven soft sensors, an abundance of multivariate statistical methods and machine learning methods such as Partial Least Squares (PLS), Principal Component Analysis (PCA), Fuzzy Logic, Support Vector Regression (SVR), and Artificial Neural Network (ANN) have been used [6].

The data-driven model is highly sensitive to dimensionality accompanied by a high degree of correlation, which leads to nontrivial correlation since the data are available in a large amount; however, it lacks the robustness. Such an issue can result in poor robustness and instability of soft sensor algorithms, besides a fall in performance prediction. Therefore, extracting the most useful information for soft sensor models is a crucial step [7]. The most famous linear algorithms feature representation for discovering data models from two different perspectives are PLS and PCA. In addition, machine learning algorithms such as Support Vector Machine (SVM) and ANN, due to their ability to cope with nonlinearity, have been commonly used in soft sensor modeling. However, with one hidden layer of model structures, these algorithms are considered as shallow learning methods. Shallow learning can be useful for simple processes and can cope with problems due to time-consuming, cost, or technical limitations with the use of a few samples and labeled data that include both input and target values. Thus, these approaches are often unsuitable for modern applications when facing highly complex processes. More potential solutions should be developed to deal with these problems. Consequently, in comparison with shallow ones, deep learning with multilayer architectures has better performance in these complex processes.

Deep learning has been widely implemented in natural language processing, image processing, speech recognition, etc. over the past few years [8–11]. In order to optimize the weight of the deep network, Hinton suggested a greedy layer-wise unsupervised pretraining learning process, making it a good solution and attracting more attention and rapidly developing [12, 13]. Recently, in many fields, deep neural networks have been proposed and undoubtedly achieved success. In those complex issues that conventional neural networks cannot properly solve, deep neural networks have shown remarkable performance. It can create more complex features successfully; meanwhile, when learning deep architectures directly, it can prevent the gradient vanishing and exploding problems, causing the gradient-based backpropagation to be unable to run the lower layers in the network [14, 15]. Deep learning has also been shown to be particularly appropriate for modeling soft sensor as it is more descriptive than conventional soft sensor models. Qiu et al. used the stacked autoencoders soft sensor to predict BOD5 in the wastewater treatment process. They showed that when compared to shallow neural networks, a deep neural network can achieve better prediction and generalization efficiency [16]. Wang et al. proposed a data-driven soft sensor model that integrates stacked autoencoders with support vectors regression (SAE-SVR) to estimate the rotor deformation of air preheaters in thermal power plant boiler. They used the Broyden–Fletcher–Goldfarb–Shanno Limited Memory (L-BFGS) algorithm to optimize weight parameters and the GA to achieve optimum SVR parameters [17]. Yan et al. proposed a deep learning-based soft sensor modeling that integrates denoising autoencoder with a neural network (DAE-NN) to estimate flue gas oxygen content in ultrasuperficial units of 1000 MW. They used improved gradient descent to update the parameters of the model [18]. Yuan et al. proposed a novel variable-wise weighted stacked autoencoder (VW-SAE) soft sensor for high-level output-related feature extraction on an industrial debutanizer column process to estimate product concentration prediction [8]. Liu et al. proposed a stacked autoencoder based deep neural network for achieving gearbox fault diagnosis [19].

In this present research work, we propose a novel soft sensor prediction modeling approach for key parameters of online measurement in wastewater treatment, which combines a deep neural network SAE-NN and the GA. The main contributions of this paper are duly to be summarized as follows. (i) The SAE, which integrates autoencoder (AE) with neural network (NN), has been used for predictive modeling of key BOD5 effluent parameter for on-line monitoring. In order to obtain the SAE, the multilayer AEs achieve coarse tuning through unsupervised learning; then the SAE achieves fine-tuning through supervised learning BP. The problem of nonlinear mapping between the auxiliary variables and the primary variables was better solved. (ii) GA was implemented to determine the number of neurons in each hidden layer, aiming at the issue that the deep neural network structure was difficult to optimize. Consequently, the accuracy of the prediction model was improved by optimizing the network structure. (iii) In order to further raise the performance of the model, the original data set was augmented by resampling and polynomial interpolation, which improved the completeness of the data. The problem of overfitting of the model was alleviated. Our approach is employed for the modeling and prediction of BOD5 in WWTPs. The experimental results showed a better prediction performance by using the proposed soft sensor modeling method based on the combination of SAE-NN and GA for on-line wastewater monitoring.

2. Problem Statement

2.1. Stacked Autoencoders. Autoencoder (AE) is an unsupervised machine learning neural network that aims to turn inputs into outputs with as little distortion as possible; namely, target variables are the matching as input variables. The dimension of the output layer is, therefore, set to be equivalent to the dimension of the input layer. The main differences between AE and multilayer NN are as follows: (i) AE merely requires data input and it will unsupervisely evaluate the output data, while multilayer NN is subject to strict supervision, which means labeled data are needed; (ii) AE is based on dimensionality reduction. This is important if input components include, or are highly correlated, a lot of redundancy. AE is a decoder and an encoder. Figure 1 depicts the AE model's basic structure. Assume that AE inputs are $x = [x_{(1)}, x_{(2)}, \ldots, x_{(d_x)}]^{\mathrm{T}} \in R^{d_x}$, where d_x stands for the dimension of the input.

The inputs x are mapped to the hidden layer $h \in R^{d_x}$ by function f as

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

where h_d stands for the hidden variable vector dimension and W stands for $h_d \times d_x$ weight matrix and $b \in R^{d_h}$ soft sensor in the decoder; the hidden representation h is mapped to the output layer of $\tilde{x} \in R^{d_x}$ by mapping function $\tilde{f}$.

$$\tilde{x} = \tilde{f}(h) = r_{\tilde{f}}(\tilde{w} + \tilde{b}), \tag{2}$$

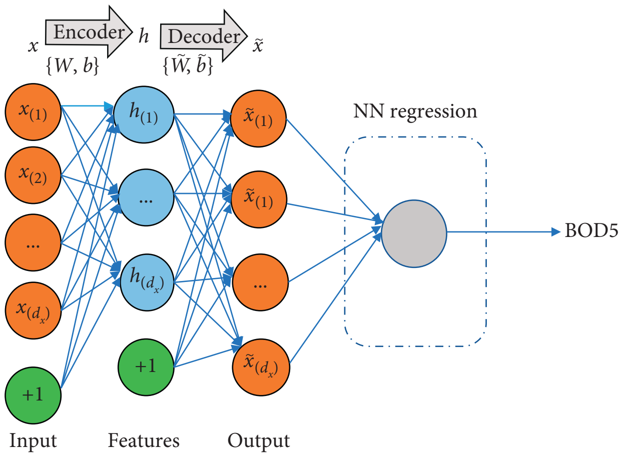


Figure 1: The structure of autoencoder.

where $\tilde{w}$ stands for a $h_d \times d_x$ weight matrix and $\tilde{b} \in R^{d_x}$ stands for the output layer bias function. The nonlinear activation functions of r_f and $r_{\tilde{f}}$ represent the rectified linear units (ReLU) and can be described as

$$r_f(x) = \max(0, x). \tag{3}$$

An AE's parameter set is $\theta = \{w, \tilde{w}, b, \tilde{b}\}$. The initial input x of AE is used to be as similar to the reconstructed output $\tilde{x}$ as possible. That is the task that AE attempts to learn $g_\theta(x) = \tilde{f}(f(x)) \approx x$. Assume the set of input data $X = \{x_1, x_2, \ldots, x_N\}$, where N represents an overall number of samples in training. By calculating the mean square error, the reconstructed loss function is reduced to obtain the model parameters as

$$j(w, \tilde{w}, b, \tilde{b}) = \frac{1}{2N}\sum_{i=1}^{N}\left\|\tilde{x} - x_i\right\|^2 = \frac{1}{2N}\sum_{i=1}^{N}\left\|g_\theta - x_i\right\|^2. \tag{4}$$

By SGD, the autoencoder parameters can be optimized.

Within the stacked autoencoder, there are multiple AEs connected layer by layer that can be trained through supervised fine-tuning and layer-wise unsupervised pretraining. Using raw input data, an unsupervised pretraining is used to train the first AE and obtain the trained function vector. The former layer's function vector is used as the output of the next layer, and this process can be repeated by layer pretraining before training the entire SAE layer. The output layer will be applied to the top of the SAE after training all the hidden layers and backpropagation (BP) using the labeled training set to minimize cost function and update weights to achieve supervised fine-tuning.

2.2. SAE Parameter Optimization Algorithm. By using an optimization algorithm, we must optimize weights in the pretraining process for each layer of the AE network; then these are the initial parameters used for a deep AE network. The BP algorithm is the most common method, but training deep backpropagation AE network typically results in lower quality of generalization. That is, the top-layer parameters will simply adapt to fit as much as possible the training data sets, regardless of the lower-layer parameters estimate. We adopt the SGD algorithm in this research work in order to optimize the initial parameters. It is a method of optimization for unconstrained problems with optimization. In SGD, for each iteration, several samples are selected randomly instead of the entire data set [19–25]. The process of conducting one iteration in each sample update basically depends on random shuffling, and hence the updated model's parameters are estimated by

$$\text{for } i \text{ range}(m)\text{: } \theta_J = \theta_J - \alpha\left(\tilde{y}^i - y^i\right)x_J^i. \tag{5}$$

At each iteration, we figure out the cost function gradient of a single example rather than the sum of the cost function gradient of all the examples, so the SGD algorithm has a high speed of execution and can also be used for online learning.

2.3. Model Structure Identification Using a Genetic Algorithm (GA). How to assess network architecture is one of the critical aspects that need to be dealt with for a neural network. In other words, the appropriate number of neurons should be selected for each hidden layer. In this present work, GA is employed to identify the number of neurons in each hidden layer, which is a process of searching and an optimization method that is driven based on the process of natural selection. Generally, it is widely utilized for finding the near-optimal solution for optimization problems with large parameter space. When employing GA, there are two preconditions that have to be realized, a defining chromosome or solution representation and a fitness function to evaluate the solutions. In this work, the root means squared error (RMSE) acts as a fitness value.

3. Main Results

3.1. Soft Sensor Modeling of Key Effluent Parameter BOD5 Based on SAE-NN. The main objective of the SAE-NN based on the soft sensor is to take the unlabeled raw data in the soft sensor modeling and to benefit the critical information behind process data.

The SAE-NN-based soft sensor structure is shown in Figure 2. First of all, the original data set from the wastewater treatment plant will be analyzed. Then, the secondary BOD5 related variables are selected (including labeled data and unlabeled data) that are used to pretrain SAE to have improved neural network initialization trained on labeled data y. Finally, the prediction values of BOD5 are obtained by SAE-NN. The soft sensor proposed in this work has two main parts: supervised learning (classical neural network NN) and an unsupervised pretraining layer (SAE). In applications of large-scale data set such as wastewater treatment data set, three layers of SAE can be utilized through the following steps. Firstly, an AE will be trained to acquire features. Second, learn secondary features after using the primary features as raw input to the next AE. Such procure is repeated to reach the last level of AE. After building the stacking of the encoders, the acquired features are used as raw input to NN regressor. And finally, the data labels are mapped after applying the training process. Eventually, all

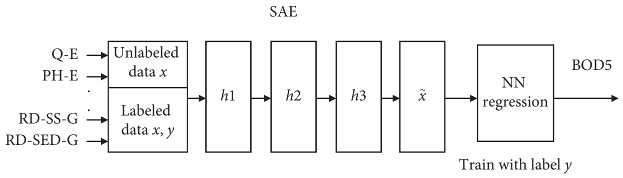


Figure 2: SAE-NN's BOD5 prediction structure.

layers are merged into stacked autoencoder and a final layer of NN regressor which is capable of regressing the BOD5 key effluent parameter.

The SAE-NN soft sensor modeling procedure is summarized as follows:

Step 1. Select secondary variables based on process knowledge and data collection and divide them into a training set, validation set, and testing set.

Step 2. Data preprocessing: resampling and interpolating the data set involves changing the time series observation rate and applying data normalization so that all observations are within 0 and 1.

Step 3. Define the deep SAE structure, train the individual AE in an unsupervised pretraining layer and use the SGD algorithm to obtain the optimized weight values, and use the genetic algorithm to determine the optimum number of neurons in each hidden layer.

Step 4. Primarily, the initiation of the weight of SAEs will be considered to launch the supervised neural network and train it on the criterion of supervised training.

Step 5. Test a SAE-NN's performance based on a soft sensor.

3.2. Case Study. In this section, in order to predict the BOD5 in a wastewater treatment plant, the soft sensor model proposed is applied to an actual WWTP. BOD5 is determined by standard, off-line 5-day delay, which plays an essential role in controlling the key effluent indicator and in preventing water body eutrophication. Soft sensor technology provides a good solution for dealing with these problems. Compared to other data-driven modeling approaches, the proposed soft sensor has shown a better performance of prediction.

3.2.1. Case Description. In the WWTP, basically intended to remove organic matter and nutrients, an activated wastewater treatment plant is commonly used. The influential rate, the performance, and the number of species of microorganisms vary over time, and the process information is very restricted. Therefore, due to its climate sensitivity and seasonal changes, an online analyzer appears to be unavailable. Based on the above, the complexity and fluctuations result in deterioration or even failure of the online analyzer performance. The proposed wastewater processing plant [26] is shown in Figure 3 which consists of four essentials: pretreatment, primary settlers, aeration tanks, and secondary settlers.

Firstly, wastewater is processed after primary settlers in the bioreactor tank where the microorganisms decrease the level of the substrate. Secondly, for biomass sled settlement, the sewage water is moved to secondary settlers.

Thus, at the top of the settlers, there is clean water and the sewage processing plant is performed. To retain a sufficient level of biomass, a fraction of the sludge is added to the input of the aeration tank to allow the organic matter to be oxidized and the remaining sludge to be purged. The plant primarily treats sewage flow of 35,000 m^3/d. A series of device variables, 8 of which are performance indicators, are calculated at several plant locations with the regular measurement of a sensor, giving a set of 38 values per day, 9 of which are percentages of performance. In this work, the behavior of the plant along of 527 days has been considered; individually it involves 38 process variables. However, the data set reduction was done to cope with any missing values for attributes. In other words, all rows with any missing data were removed from the data set, thereby resulting in a data set with 381 instances. It is necessary to select the correct secondary variables in order to achieve high performance because irrelevant variables will deteriorate the soft sensor's prediction performance. Figure 4 shows Pearson's linear correlation. Nineteen process variables were chosen to predict BOD5, including local settlers performance based on SS/COD/BOD5, suspended solids (SS), sediments, biochemical demand for oxygen (BOD), unstable suspended solids, chemical demand for oxygen (COD), input, and global plant performance based on BOD/COD/SS input. Then, these nineteen variables were employed as the model soft sensor input, and BOD5 was employed as the model soft sensor output. Table 1 shows the details of secondary variables.

3.2.2. Augmentation Processing and Data Preprocessing. Because of overfitting, training a deep SAE with small data set would deteriorate SAE's performance; that is, the network is working well on the training set, but the testing set is worth it [27, 28]. Augmentation data are used to solve this problem in order to expand the data set and reduce the problem of overfitting [29, 30]. The number of samples is increased in the data augmentation method by applying sampling polynomial interpolation to the data set, where we increased the frequency of data set from days to hours and used an interpolation scheme to fill in the new hourly frequency.

To eliminate different scales of data set, all data set is scaled to (0, 1) by min-max scaler according to the following equation:

$$x_i^* = \frac{x - \min(x_i)}{\max(x_i) - \min(x_i)}, \tag{6}$$

where x_i^* refers to normalized variables and i refers to the dimension of the data set.

3.2.3. Setting Parameters of the Deep Neural Network. The performance of the soft sensor is governed by a neuron's number in each hidden layer; this process was not modeled

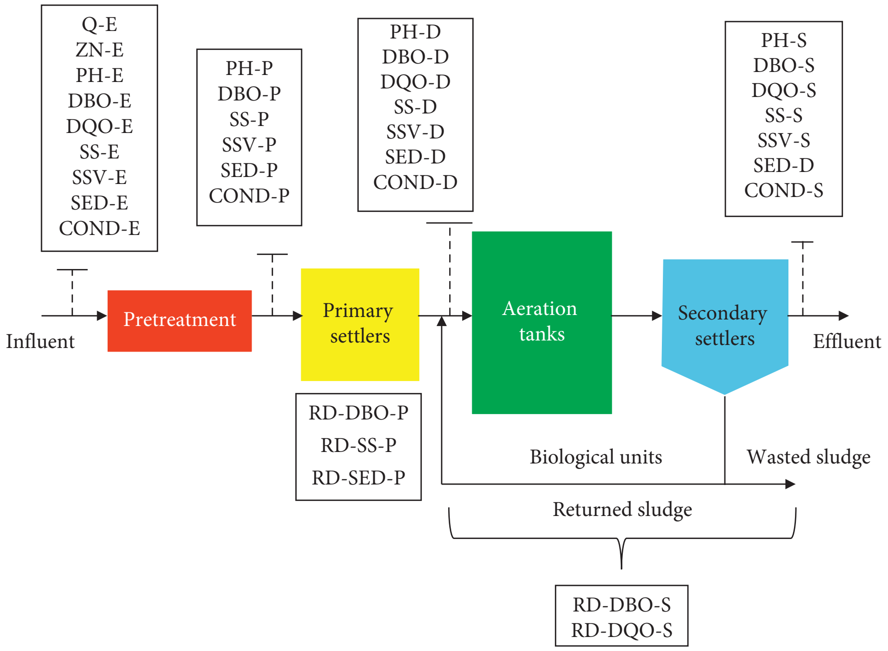


FIGURE 3: The schematic of the wastewater plant.

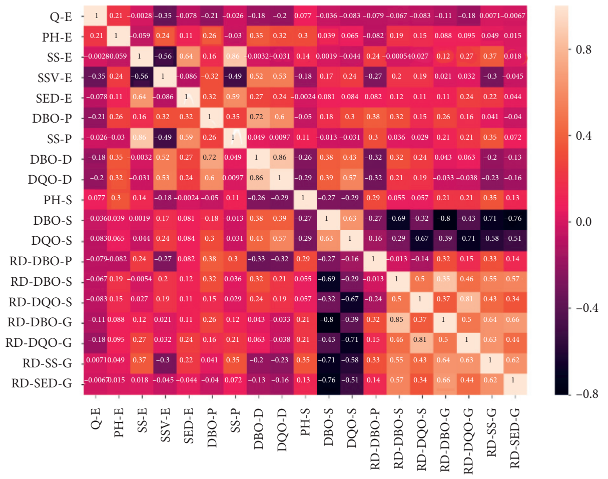


FIGURE 4: Pearson's linear correlation.

up to date. The genetic algorithm used in this work is to choose the number of neurons in the hidden layer. To evaluate the soft sensor model performance, the root means squared error (RMSE) and the correlation coefficient (R^2) are used.

The calculation of RMSE is

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{y_{i,\text{predict}} - y_{i,\text{real}}}{N}}, \quad (7)$$

where $y_{i,\text{predict}}$ and $y_{i,\text{real}}$ are, respectively, the forecast value and the real value for the example; and in the given data set, N refers to the total number of examples.

(1) R^2 is calculated by

$$R^2 = \frac{1 - \sum_{i=1}^{N} \left(y_{i,\text{predict}} - y_{i,\text{real}}\right)^2}{\sum_{i=1}^{N} \left(y_{i,\text{predict}} - \overline{y}\right)^2}, \quad (8)$$

where $\overline{y}$ stands for the average of the test set's output values.

4. Simulation Experiment and Result Analysis

The proposed soft sensor was validated in this study and compared to three conventional soft sensors: SVR (there are a number of core functions like linear kernel, polynomial kernel, sigmoid kernel, RBF kernel, etc., gamma = scale (hard no limit on iterations within solver) max_iter = −1), PCA-SVR combining PCA (number of components = 10) and support vector regression, and NN with three hidden layers (activation = relu, optimizer = sgd, momentum for gradient descent update = 0.9, and initial learning rate = 0.001). The data set itself will be used to model training for comparison purposes in order to ensure a fair comparison. The number of neurons in each hidden layer is determined experimentally as 13, 13, 13, and the regularization parameters C and ε are 5 and 0,022, respectively, based on the linear kernel function of SVR, using the GA. 12749 samples have been utilized as samples for training (the initial training samples are divided into new training samples and validation samples, of which 30% were initial training samples) and 3188 are used as testing samples. In order to obtain the SAE, an unsupervised pretraining layer-wise method is utilized to achieve a good initialization for the weights and bias of each AE. Every AE is equipped with the SGD algorithm for batch normalization to speed up training, and Dropout is used for regularization, setting the batch size as 512 samples. Every AE is trained iteratively by 50 epochs, in addition to the supervised backpropagation-trained fine-tuning. The size of the batch is set as 128 samples and

Table 1: Variables selected for the soft sensor.

Attribute no.	Comments	Attributes
1	Input flow to the plant	Q-E
2	pH input to the plant	PH-E
3	Suspended solids input to the plant	SS-E
4	Volatile input suspended solids into the plant	SSV-E
5	Plant input sediments	SED-E
6	Biological oxygen demand input for primary settlers	BDO-P
7	Suspended solids input into primary settlers	SS-P
8	Biological oxygen demand input for secondary settlers	DBO-D
9	Chemical oxygen demand input for secondary settlers	DQO-D
10	pH output	pH-S
11	Biological oxygen demand output	DBO-S
12	Chemical oxygen demand output	DQO-S
13	Input performance biological oxygen demand in primary settlers	RD-DBO-P
14	Input performance biological oxygen demand in secondary settler	RD-DBO-S
15	Input performance chemical oxygen demand in secondary settlers	RD-DQO-S
16	The output of global quality biological oxygen demand	RD-DBO-G
17	The output of global quality chemical oxygen demand	RD-DQO-G
18	The output of global quality suspended solids	RD-SS-G
19	The output of global quality sediments	RD-SED-G

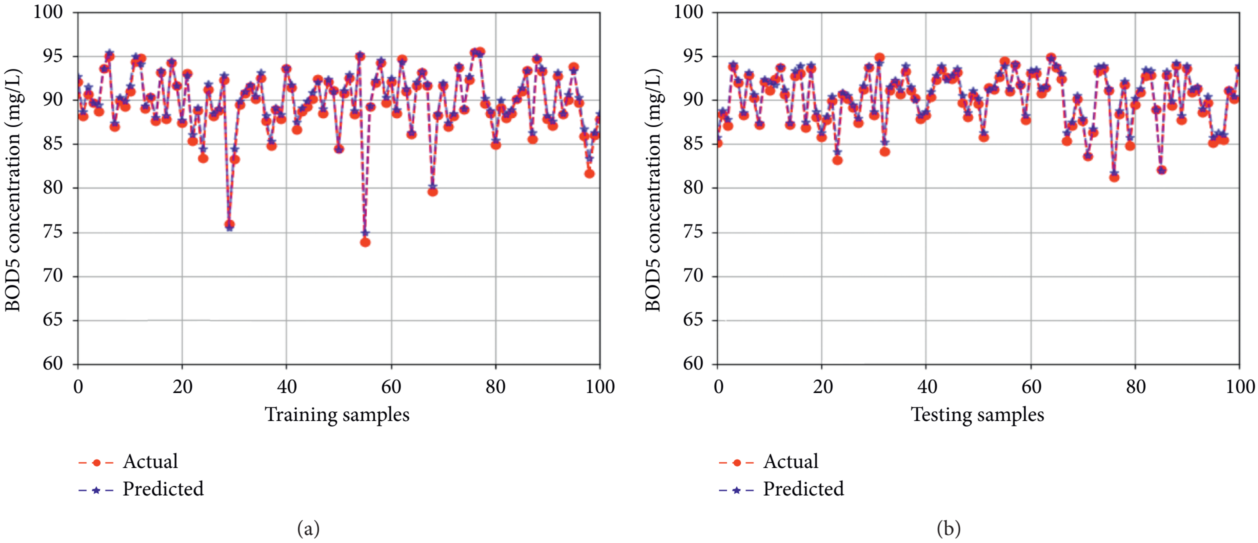
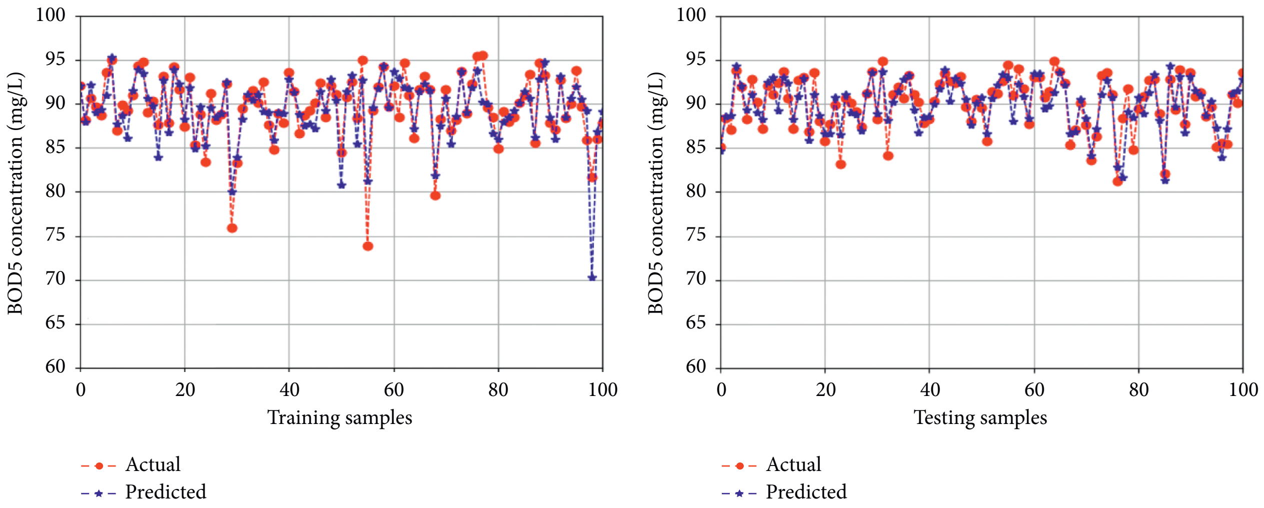


Figure 5: Soft sensor performance prediction based on SAE-NN.

iteratively as 150 epochs. As can be seen in Figure 5, RMSE and R^2 of 0.0051 and 0.989, respectively, were predicted by the proposed soft sensor based on SAE-NN. It can be seen that the predicted values of the BOD5 match real values and when faced with drastic variations it performs better. The figures show only 100 training samples and 100 testing samples because the training and testing data set is too large to show. Table 2 describes the predictive performance of the soft sensor based on different approaches. The results of PCA-SVR, SVR, multilayer NN, and the proposed soft sensor based on SAE-NN are predicted in Table 2 on the training and testing data set. As can be seen, the SAE-NN-based soft sensor has much better performance in learning and generalization than other conventional soft sensors and given a fairly satisfactory estimate of BOD5, while PCA-SVR, SVR, and multilayer NN obtained relatively poor results. The SVR had the worst results because it was unable to adequately describe the nonlinear structure data. The PCA-SVR model achieved slightly better predictive results as PCA can eliminate input data from noise and data redundancy to improve predictive performance. More accurately than PCA-SVR and SVR,

Table 2: Results of different models of soft sensors.

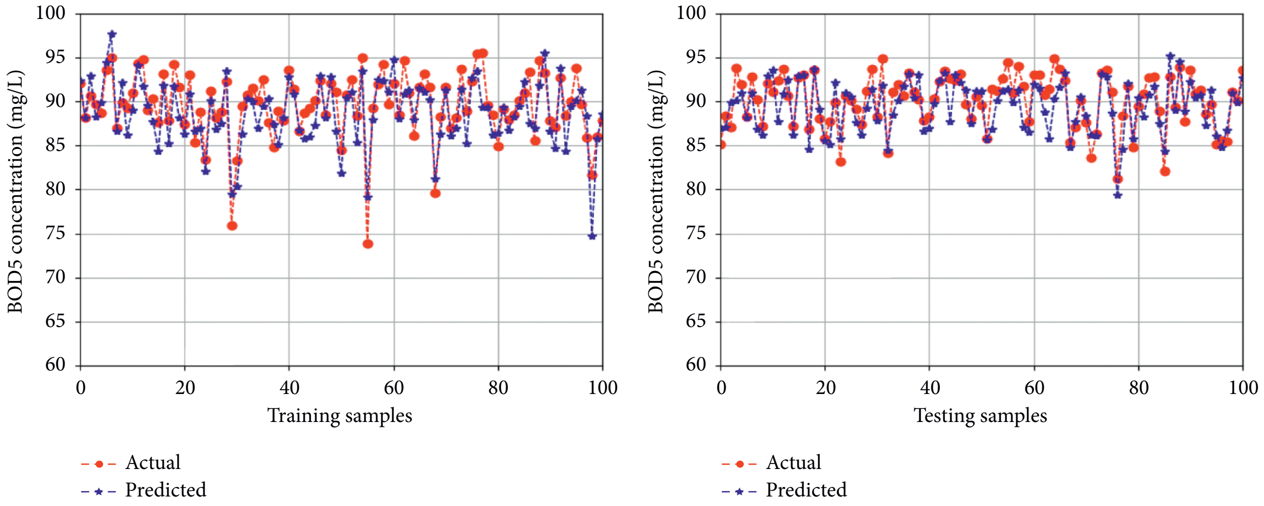
	SAE-NN	PCA-SVR	SVR	NN (3 hidden layers)
Training RMSE	0.0051	0.0246	0.0283	0.0202
Testing RMSE	0.0041	0.0249	0.0287	0.0203
Training R^2	0.989	0.876	0.837	0.917
Testing R^2	0.987	0.849	0.799	0.899



(a)

(b)

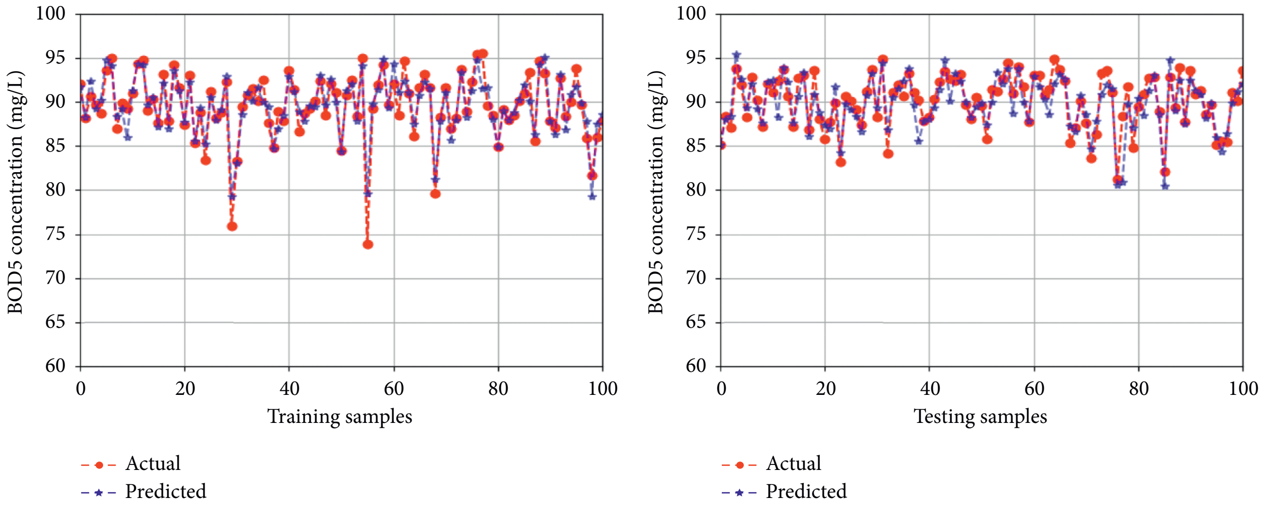
Figure 6: Soft sensor performance prediction based on PCA-SVR.



(a)

(b)

Figure 7: Soft sensor performance prediction based on SVR.



(a)

(b)

Figure 8: Soft sensor performance prediction based on NN.

traditional multilayer NN can approximate the complex data relationship. Nonetheless, NN with 3 hidden layers does not provide great predictive performance compared to SAE-NN.

The network parameters for the multilayer NN have been randomly initialized and local optima are easily disposed to it. That is, training NN with BP results in a slow rate of convergence and difficulties in deciding an appropriate architecture to achieve a minimum. In contrast, layer by layer can be extracted from high-level abstract features in SAE. Therefore, for tasks of prediction, these features are much more structured. That is, the performance of SAE-NN is better than that of multilayer NN. The comparative results of the other traditional soft sensor are shown in Figures 6–8. It can be observed from Figures 6–8 that from estimating the BOD5 in the WWTP plant, this explicitly shows the fact that soft sensor based on SAE-NN has good performance. That is, from tracking the BOD5's varying trend, SAE-NN performs well. The prediction errors of SAE-NN are smaller than those of the other, as can be seen. That is, the SAE-NN forecast easily shifts without significant variations and displays greater robustness than models of shallow architecture predictions. The experiments were performed on a PC using Intel® Core™ i5-8250U CPU @1.60 GHz (8 CPUs)~1.8 GHz, 4 GB RAM using the Keras Python deep learning library (using TensorFlow backend) 2.2.4. [31].

5. Conclusions

In this paper, in a wastewater treatment plant, SAE-NN based on the data-driven soft sensor is proposed and implemented to estimate the BOD5. The stacked AEs have been trained for the supervised NN to obtain initialization weights, which resulted in the best generalization of the NN system and avoiding the issue of overfitting. In addition, in each hidden layer, GA was developed to determine the appropriate number of neurons. Generally, the soft sensor output is estimated by approximating the real values of BOD5. In most cases, the SAE-NN-based soft sensor outperforms all additional tools regarding the soft sensors. Deep learning in many industrial process applications is superior to shallow learning when faced with complex situations and is a promising approach for modeling soft sensors. Automatically selecting the appropriate parameter values to improve the performance of the deep network will be the focus in future work. Further future work will also extend our approach to a pretraining layer-wise manner which is supervised or semisupervised.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant nos. 61763027 and 61364011).

References

[1] H. Haimi, M. Mulas, F. Corona, and R. Vahala, "Data-derived soft-sensors for biological wastewater treatment plants: an overview," *Environmental Modelling & Software*, vol. 47, pp. 88–107, 2013.

[2] H. Han, S. Zhu, J. Qiao, and M. Guo, "Data-driven intelligent monitoring system for key variables in wastewater treatment process," *Chinese Journal of Chemical Engineering*, vol. 26, no. 10, pp. 2093–2101, 2018.

[3] V. Gopakumar, S. Tiwari, and I. Rahman, "A deep learning based data driven soft sensor for bioprocesses," *Biochemical Engineering Journal*, vol. 136, pp. 28–39, 2018.

[4] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009.

[5] C. M. Thürlimann, D. J. Dürrenmatt, and K. Villez, "Soft-sensing with qualitative trend analysis for wastewater treatment plant control," *Control Engineering Practice*, vol. 70, pp. 121–133, 2018.

[6] H. Kaneko and K. Funatsu, "Application of online support vector regression for soft sensors," *AIChE Journal*, vol. 60, no. 2, pp. 600–612, 2014.

[7] S. I. Abba and G. Elkiran, "Effluent prediction of chemical oxygen demand from the astewater treatment plant using artificial neural network application," *Procedia Computer Science*, vol. 120, pp. 156–163, 2017.

[8] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3235–3243, 2018.

[9] W. Zhu, Y. Ma, Y. Zhou, M. Benton, and J. Romagnoli, "Deep learning based soft sensor and its application on a pyrolysis reactor for compositions predictions of gas phase components," in *13th International Symposium on Process Systems Engineering (PSE 2018)*, pp. 2245–2250, Elsevier, Amsterdam, Netherlands, 2018.

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

[11] X. Yuan, C. Ou, 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*, vol. 396, pp. 375–382, 2019.

[12] J. Zhu, Z. Ge, and Z. Song, "Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1877–1885, 2017.

[13] Hamedmoghadam H., N. Joorabloo, and M. Jalili, Australia's long-term electricity demand forecasting using deep neural networks, arXiv preprint arXiv:1801.02148, 2018.

[14] R. Zhang and J. Tao, "Data-driven modeling using improved multi-objective optimization based neural network for coke furnace system," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3147–3155, 2017.

[15] Y. Bengio, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems*, pp. 153–160, MIT Press, Cambridge, MA, USA, 2007.

[16] Y. Qiu, Y. Liu, and D. Huang, "Date-driven soft-sensor design for biological wastewater treatment using deep neural networks and genetic algorithms," *Journal of Chemical Engineering of Japan*, vol. 49, no. 10, pp. 925–936, 2016.

[17] X. Wang and H. Liu, "Soft sensor based on stacked autoencoder deep neural network for air preheater rotor deformation prediction," *Advanced Engineering Informatics*, vol. 36, pp. 112–119, 2018.

[18] W. Yan, Di Tang, and Y. Lin, "A data-driven soft sensor modeling method based on deep learning and its application," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4237–4245, 2016.

[19] G. Liu, H. Bao, and B. Han, "A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis," *Mathematical Problems in Engineering*, vol. 2018, Article ID 5105709, 10 pages, 2018.

[20] B. Schölkopf, J. Platt, and T. Hofmann, "Greedy layer-wise training of deep networks," in *International Conference on Neural Information Processing Systems*, pp. 153–160, MIT Press, Cambridge, MA, USA, 2006.

[21] X. Yuan, Y. Wang, C. Yang, Z. Ge, Z. Song, and W. Gui, "Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1508–1517, 2018.

[22] J. Yu, C. Hong, Y. Rui, and D. Tao, "Multi-task autoencoder model for recovering human poses," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 5060–5068, 2017.

[23] Y. Ju, J. Guo, and S. Liu, "A deep learning method combined sparse autoencoder with SVM," in *Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, Xi'an, China, September 2015.

[24] S.-Z. Li, B. Yu, W. Wu, S.-Z. Su, and R.-R. Ji, "Feature learning based on SAE-PCA network for human gesture recognition in RGBD images," *Neurocomputing*, vol. 151, pp. 565–573, 2015.

[25] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[26] C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, University of California, Oakland, CA, USA, 1998.

[27] S. Feng, H. Zhou, and H. Dong, "Using deep neural network with small dataset to predict material defects," *Materials & Design*, vol. 162, pp. 300–310, 2019.

[28] C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pp. 562–570, San Diego, CA, USA, May 2015.

[29] K. Xu, X Shen, T Yao, X Tian, and T Mei, "Greedy layer-wise training of long short-term memory networks," in *Proceedings of the 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, San Diego, CA, USA, July 2018.

[30] S. Khan, N. Islam, Z. Jan, I. Ud Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, 2019.

[31] Chollet F., Keras, https://keras.io, 2015.