Research Article

Modeling NOx Emissions with an Intelligent Combinatorial Algorithm

XiaoJuan Chen,1 Haiyang Zhang,1 Xiaoxue Xing,2 and Hongwu Qin2

1School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun 130022, China
2School of Electronic Information Engineering, Changchun University, Changchun 130022, China

Correspondence should be addressed to XiaoJuan Chen; cxj001@cust.edu.cn

Received 19 November 2020; Revised 4 January 2021; Accepted 6 March 2021; Published 19 March 2021

Academic Editor: Waldemar T. Wójcik

Copyright © 2021 XiaoJuan Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coal combustion is considered to be the key source of nitrogen oxide (NOx) emissions in thermal power plants. Methods for effective reduction in these emissions are critically sought on the national and global levels. Such methods typically achieve this goal through accurate modeling and prediction. However, such modeling process is difficult because of the complexity of the NOx emission mechanisms and the influence of many factors. Furthermore, real-operation data of power plants tend to be centralized in some local areas because of working condition experiment so that no single model can deal with the complicated and changeable boiler production processes. In this paper, we address this problem and propose a model intelligent combinatorial algorithm (MICA). First, the actual production data are preprocessed by a wavelet denoising algorithm, and the model input variables are selected based on a random forest algorithm. Then, several models for NOx emission prediction are constructed by various data-driven algorithms. Finally, a C4.5 algorithm is applied to intelligently combine these models. The experimental results indicate that the proposed algorithm can construct an accurate prediction model for NOx emissions based on actual operating data. The mean absolute percentage errors are within 1%. Moreover, a correlation of 0.98 between predicted and measured values was obtained by applying the MICA model.

1. Introduction

About 70% of China’s electricity requirement comes from coal-based power plants, which are predicted to stay as the main electricity source for a long future span [1]. Nevertheless, nitrogen oxide (NOx) emissions caused by coal burning are responsible for several phenomena, including acid rain, photochemical smog, and global warming. These phenomena have resulted in serious detrimental effects on air quality and human health [2]. With the adoption of a new national policy in China for coal-fired boiler emissions, NOx emissions are now limited to be below 50 mg/Nm³ at 6% O₂ [3]. Therefore, numerous novel approaches have been proposed in order to cut the NOx emissions and meet the aforementioned emission level.

Some of the techniques that have been proposed to alleviate the negative impacts of NOx emissions include selective catalytic and noncatalytic reduction, employing low NOx burners, excess air reduction, adsorption, and flue gas recirculation [4]. Furthermore, boiler combustion optimization has received special attention as a way to minimize NOx emissions in coal-based power plants. In fact, this optimization approach is more cost effective and easier to implement than equipment retrofit [5]. Generally, the combustion optimization approach has two major stages: modeling and optimization. In the first stage, a model for the relationship of NOx emissions and the operating parameters is constructed. This stage involves the selection of the modeling approach and the operating parameters. In the second stage, the operating parameters of the constructed model are optimized.

Indeed, creating an adequate model for NOx emission prediction is quite challenging, especially because of the complexity of the combustion processes. Numerous approaches have been introduced to address this challenge and accurately model NOx emissions. In particular, conventional mechanism-based algorithms have been employed in the prediction of NOx emissions of coal-powered boilers; for
example, Diez et al. [6] carried out an investigation based on computational fluid dynamics (CFD) for the reduction in NOx emissions with overfire air. Belosevic et al. [7] used a differential equation model to estimate NOx emissions. Nevertheless, those models are highly complex and computationally demanding as they require full knowledge of the operating conditions and involve a large number of parameters [8]. Moreover, the degradation of the boiler equipment overtime will gradually increase the inaccuracy of these models. Meanwhile, data-driven models for NOx emission prediction have received growing attention. For instance, Zhou et al. [9] have employed artificial neural networks (ANNs) for modeling NOx emissions caused by a 600 MW coal-powered tangentially fired boilers. Ilamathi et al. [10] trained an ANN model to predict NOx emissions under full-load conditions for a 210 MW coal-powered boiler. However, ANNs usually suffer from overfitting and poor generalization, particularly with limited training data. Recently, the focus on ANNs in machine learning has been shifted to support vector regression (SVR) and one of its variants, namely, least-square support vector machines (LSSVMs). In fact, SVR and LSSVM have emerged as promising tools for NOx emission modeling. For example, Zhou et al. [11] modeled NOx emissions with SVR and compared the SVR outcomes with those of the back-propagation neural networks (BPNs) and the generalized regression neural networks (GRNNs). Lv et al. [8, 12] also applied LSSVM to model NOx emissions. Moreover, extreme learning machines (ELMs) have been also used for modeling the NOx emissions caused by pulverized coal-fired power plants. Tan et al. [2] introduced a novel ELM to investigate the correlation between the boiler NOx emissions and the operating parameters. Similarly, Li et al. [13] modeled the NOx emissions of a boiler with an ELM and demonstrated better regression precision. Besides, Elman networks were also employed for NOx emission prediction [14]. Although each prediction model performs well for certain production conditions, no single model can deal with the complicated and changeable boiler production processes.

In the second stage, real-time combustion optimization needs the optimization algorithm with rapid convergence speed and high-quality solutions [2]. A new hybrid jump particle swarm optimization (PSO) algorithm has been used to tune control parameters of the boiler-turbine unit control [15]. Differential evolution algorithm (DE) has been proposed to optimize LSSVM parameters for different problems [16]. Ant colony optimization (ACO) algorithm has been proposed to select the parameters for SVR to model NOx emission [11]. Moreover, genetic algorithm (GA) and its variants have been employed to accomplish the operational parameter optimization task [10, 17, 18].

In this paper, a model intelligent combinatorial algorithm is introduced to address the aforementioned challenges in NOx emission modeling.

Boiler combustion involves complex physical and chemical processes, where the parameters of each process are strongly correlated. Information redundancy in these processes could increase the model complexity and negatively affect the model performance. Therefore, proper selection of the input parameters is critical for maintaining a high model accuracy. The training of our models was carried out using historical operation data which was collected from a 1000 MW ultra-supercritical once-through boiler. Mechanism analysis was used to select the initial operating parameters. Then, a random forest (RF) algorithm was employed to decrease the input parameter count.

Our work seeks essentially to construct a model intelligent combinatorial algorithm for predicting NOx emissions caused by a coal-powered boiler. Our proposed approach has several stages. First of all, the model input parameters for predicting the NOx emissions are selected based on mechanism analysis and the RF algorithm. Subsequently, the performance of different data-driven algorithms is tested, and the prediction results of different models are obtained. Finally, according to the analysis results, an intelligent combinatorial algorithm was designed. In particular, the C4.5 algorithm was used for intelligent model selection and obtaining the final prediction algorithm. Then, the model intelligent combinatorial algorithm (MICA) was exploited to create the NOx emission prediction model based on data from real operations. A performance comparison was made against the conventional single model. The rest of our paper is arranged as follows. A brief description is given in Section 2 for the boiler employed in the present study. Section 3 introduces the proposed modeling algorithm. Results and their discussion are detailed in Section 4, while conclusions are highlighted in Section 5.

2. Boiler Description

The data employed in creating our model pertain to a 1000 MW ultra-supercritical once-through boiler with variable pressure and octagonal-inverse double tangential firing. The furnace has a height of 66.4 m and a cross-sectional area of 32.084 m × 15.670 m. Forty-eight low NOx burners were divided based on their elevations into six groups, and the burners within each group were placed at the eight corners of the furnace. For each burner group, a mill with a medium speed was used to feed pulverized coal. One layer of overfire air (OFA) and four layers of additional air (AA) were distributed alternately in a vertical direction in order to effectively decrease NOx emissions and improve combustion efficiency. The furnace has a rectangular structure, and eight reverse double tangential-swing burners are arranged in each layer with four burners placed on the front wall and the other four on the rear wall. Due to this structure, coking can be effectively prevented in the furnace.

A schematic diagram of the layouts of the furnace and the burner is demonstrated in Figure 1. Coal is pulverized within the mills using the primary air flow into the furnace and through the burners. To ensure proper fuel burning, secondary air flow (heated by an air preheater) is blown through OFA ports and an AA port into the furnace. The pulverized coal is fully burned in the fuel air to produce heat. Then, the produced heat is used to convert water into steam, which subsequently drives turbines that generate electricity.
A power plant typically has a distributed control system (DCS), which reveals useful information on the operation of the plant. Based on mechanism analysis and suggestions made by experts, 112 parameters were initially picked as potential inputs for data-driven models. In fact, while models can be tailored for specific types of coal, the coal characteristics were roughly kept fixed, as listed in Table 1.

### 3. Model Development for NOx Emission Prediction

The proposed approach for NOx emission prediction consists of three major stages: data preprocessing, feature selection, and MICA modeling. Data preprocessing involves data denoising and data normalization, as detailed in Section 3.1. Feature selection methods seek the process variables with the greatest influence. As shown in Section 3.2, those variables were found in our work using mechanism analysis and the RF algorithm. Finally, a MICA-based prediction model for NOx emissions is established.

#### 3.1. Data Preprocessing

A dataset of more than 5000 real-operation data samples was collected from the DCS of an operational power plant with a sampling interval of 1 minute and a boiler load varying from 700 MW to 1000 MW. The data were preprocessed to remove noise arising from data acquisition and transmission. Firstly, each outlier sample was identified and replaced by the average of its neighboring samples. Subsequently, wavelet denoising [18–20] was used to simultaneously reduce the noise and analyze the data in the time and frequency domains. Wavelet denoising consists of three steps, namely, decomposition, thresholding, and reconstruction. In this paper, a 3-layer Haar wavelet was used for decomposition, and a soft-threshold function was adopted for thresholding. Raw and denoised data samples for NOx emissions are shown in Figure 2. All data samples were normalized into the interval [−1, 1], and the thresholded wavelet coefficients were transformed back into the time domain.

#### 3.2. Feature Selection

Typically, data-driven models are sensitive to input feature selection, which can seriously affect the modeling accuracy. Indeed, excessive information redundancy can lead to inaccurate predictive models. As mentioned earlier in Section 2, NOx emissions of power stations are particularly influenced by numerous variables. In this work, a random forest (RF) algorithm is used to select the input variables and hence improve the accuracy of the NOx emission prediction model. Based on feature importance, 31 input parameters were selected as model input variables, as shown in Table 2.

#### 3.3. Model Intelligent Combinatorial Algorithm

The accuracy of an NOx emission prediction model based on a single predictive method is typically limited due to weaknesses of...
the learning algorithms. In this paper, a variety of learning algorithms (including the BP, Elman, ELM, and LSSVM algorithms) are used to address the NOx emission prediction problem. An analysis of the experimental results shows that the C4.5-based MICA achieves high-precision prediction of the boiler NOx emissions. A flowchart of the MICA-based NOx emission model is depicted in Figure 3. The steps of the MICA-based modeling approach are as follows:

Step 1: build a data-driven basic model. Subdivide the given dataset into three subsets P1, P2, and P3 for model training, testing, and validation, respectively. The training subset, P1, is used to construct a basic model based on the four learning methods.

Step 2: build a sample classification model. For each learning method, the relative error of the corresponding basic model is obtained for the test subset P2. Then, each testing sample is assigned the label of the optimal predictive model: “1” for BP, “2” for the Elman network, “3” for ELM, and “4” for LSSVM. Partial classification results are shown in Table 3. The results show that 121, 73, 87, and 168 samples were associated with the BP, Elman, ELM, and LSSVM methods, respectively. Also, the results show that a single method cannot obtain satisfactory results in all the cases. Finally, the C4.5 method is employed to construct a predictor based on the new testing set with labels.

The learning algorithms. In this paper, a variety of learning algorithms (including the BP, Elman, ELM, and LSSVM algorithms) are used to address the NOx emission prediction problem. An analysis of the experimental results shows that the C4.5-based MICA achieves high-precision prediction of the boiler NOx emissions. A flowchart of the MICA-based NOx emission model is depicted in Figure 3. The steps of the MICA-based modeling approach are as follows:

Step 1: build a data-driven basic model. Subdivide the given dataset into three subsets P1, P2, and P3 for model training, testing, and validation, respectively. The training subset, P1, is used to construct a basic model based on the four learning methods.

Figure 2: Results of wavelet denoising.

Table 2: Process variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Unit</th>
<th>Variation range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boiler load</td>
<td>WM</td>
<td>[700, 1000]</td>
</tr>
<tr>
<td>Total air flow</td>
<td>t/h</td>
<td>[2728, 3789]</td>
</tr>
<tr>
<td>Total coal flow</td>
<td>t/h</td>
<td>[260, 382]</td>
</tr>
<tr>
<td>Water coal ratio</td>
<td>%</td>
<td>[6.44, 7.81]</td>
</tr>
<tr>
<td>O2 content</td>
<td>%</td>
<td>[2.29, 5.54]</td>
</tr>
<tr>
<td>Feed water flow</td>
<td>t/h</td>
<td>[1923, 2713]</td>
</tr>
<tr>
<td>Feed water temperature</td>
<td>°C</td>
<td>[297, 317]</td>
</tr>
<tr>
<td>Exhaust gas temperature</td>
<td>°C</td>
<td>[119, 138]</td>
</tr>
<tr>
<td>Overfire air valve</td>
<td>%</td>
<td>[56, 183]</td>
</tr>
<tr>
<td>Additional air (A, B, C, D)</td>
<td>t/h</td>
<td>[45, 183]</td>
</tr>
<tr>
<td>Hot primary air valve opening (B, C, D, E, F)</td>
<td>%</td>
<td>[10, 99]</td>
</tr>
<tr>
<td>Cold primary air valve opening (B, C, D, E, F)</td>
<td>%</td>
<td>[0, 99]</td>
</tr>
<tr>
<td>Secondary air valve opening (AB, BC, CC, DD, DE, EF, FF)</td>
<td>%</td>
<td>[23, 82]</td>
</tr>
<tr>
<td>Outlet pressure of blower</td>
<td>kPa</td>
<td>[0.97, 2.91]</td>
</tr>
<tr>
<td>NOx emission in flue gases</td>
<td>mg/Nm³</td>
<td>[113, 388]</td>
</tr>
</tbody>
</table>
Step 3: validate the model. For a new test sample, classification is performed to find the best learning model firstly and then predict the NOx emissions for that sample. This is a basic model based on the four learning methods.

4. Results and Discussion

The real-operation data samples employed in this work cover about 4 consecutive days with a sampling interval of 1 minute and a boiler load ranging from around 700 MW to 1000 MW. These samples are divided based on the sampling time into three subsets T1, T2, and T3.

4.1. Performance Indices. For evaluating the performance of the proposed NOx emission prediction model, we use five performance indices, namely, the mean absolute error (MAE), the root-mean-square error (RMSE), the mean absolute percentage error (MAPE), the coefficient of determination ($R^2$), and the classification accuracy (AC). Those indices are, respectively, defined as follows [16, 21]:

$$MAE = \frac{1}{M} \sum_{j=1}^{M} |y_j - \hat{y}_j|,$$

$$MAPE = \frac{1}{M} \sum_{j=1}^{M} \left| \frac{y_j - \hat{y}_j}{y_j} \right|,$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{j=1}^{M} (y_j - \hat{y}_j)^2},$$

$$R^2 = 1 - \frac{\sum_{j=1}^{M} (y_j - \hat{y}_j)^2}{\sum_{j=1}^{M} (y_j - \bar{y})^2},$$

$$AC = \frac{A}{A + B} \times 100\%,$$

where $M$ denotes the number of samples, $y_j$ and $\hat{y}_j$ denote, respectively, the actual and predicted NOx emission levels, and $\bar{y}$ denotes the mean value of the actual measurements. The symbols $A$ and $B$ denote the numbers of correctly and incorrectly classified samples, respectively.

4.2. Prediction Results. We compare the prediction results of the MICA-based NOx emission model with those of the basic model with four learning methods (including BP, Elman, ELM, and LSSVM). Figure 4 and Table 4 summarize the results of those five models. Figure 4 shows that all prediction curves closely match the real-data measurements, and that all five algorithms can be used to predict the NOx emissions. Nevertheless, the errors of the MICA-based model clearly vary within a smaller range. The worst results are provided by the Elman algorithm as the predicted curve does not reflect the true data. Besides, relative errors of different models will be given in Figure 4(d), from which a much easier and more detailed comparison can be made. It can be discovered that the majority of relative errors of all models are less than 5%, which provides necessary conditions for combustion optimization. Meanwhile, the majority of relative errors of MICA model are less than 1%. Specifically, Figure 4(d) shows that the relative errors of the MICA-based model are less than those of the other models, and thus, the MICA-based model performs better than the other models. Specifically, Table 4 shows that the proposed model has superior NOx prediction performance with MAE = 2.02, RMSE = 3.87, and MAPE = 0.9% for the data subset T1. Among all of the models used in this study, the performance of Elman model and ELM model is relatively poor. The RMSE of the Elman reaches 10.737, and MAE is up to 7.9483. Also, the RMSE of the ELM reaches 9.2816, and MAE is up to 6.4739. The BP model and LSSVM model exhibit moderate performance. In
general, the proposed MICA model gives a better performance for NOx emission prediction compared with single model.

4.3. Data Preprocessing Analysis. As Table 5 shows, the models based on denoised data give better prediction accuracies with less RMSE, MAPE, and MAE than the models based on raw data in the case of T1. This confirms the importance of data denoising for improving the prediction performance.

4.4. Feature Selection Analysis. Figure 5 compares the true and predicted NOx emission levels on test samples with and without the RF feature selection algorithm in the case of T2. Obviously, the data are diagonally distributed along the perfect-fit line where the true and predicted values are equal. This means that the NOx emission predictions are highly accurate for all datasets. For a given model with RF-based feature selection, the determination coefficient ($R^2$) is higher than that of the model constructed with no feature selection; for example, the $R^2$ of the BP-based model with RF is 0.87, and the model without RF is 0.76. The $R^2$ of the Elman-based model with RF is 0.79, and the model without RF is 0.65. The $R^2$ of the ELM-based model with RF is 0.91, and the model without RF is 0.77. The $R^2$ of the LSSVM-based model with RF is 0.93, and the model without RF is 0.92. The $R^2$ of the MICA-based model with RF is 0.98, and the model without RF is 0.95. The RMSE, MAPE, $R^2$, and MAE values of the five models are also easy to be calculated and will be listed in Table 6. The models constructed with
Table 4: Comparative results with different evaluation criteria.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Performance index</th>
<th>BP</th>
<th>Elman</th>
<th>ELM</th>
<th>LSSVM</th>
<th>MICA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>0.021922</td>
<td>0.037648</td>
<td>0.030312</td>
<td>0.017279</td>
<td>0.0093577</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>6.7636</td>
<td>10.737</td>
<td>9.2816</td>
<td>5.7475</td>
<td>3.8705</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>4.6308</td>
<td>7.9483</td>
<td>6.4739</td>
<td>3.7111</td>
<td>2.0177</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>0.96564</td>
<td>0.91276</td>
<td>0.93255</td>
<td>0.97412</td>
<td>0.98427</td>
</tr>
<tr>
<td>T1</td>
<td>MAPE</td>
<td>0.016702</td>
<td>0.034262</td>
<td>0.029356</td>
<td>0.0099125</td>
<td>0.0061889</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>6.97</td>
<td>9.888</td>
<td>8.8721</td>
<td>5.3717</td>
<td>4.4285</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>3.5601</td>
<td>7.1228</td>
<td>6.0779</td>
<td>2.1552</td>
<td>1.3058</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>0.97127</td>
<td>0.93843</td>
<td>0.94887</td>
<td>0.98133</td>
<td>0.98729</td>
</tr>
<tr>
<td>T2</td>
<td>MAPE</td>
<td>0.018918</td>
<td>0.049245</td>
<td>0.029788</td>
<td>0.012914</td>
<td>0.0074429</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>9.1382</td>
<td>17.444</td>
<td>14.03</td>
<td>8.3502</td>
<td>6.5442</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>4.6028</td>
<td>11.752</td>
<td>7.2196</td>
<td>3.2876</td>
<td>1.9638</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>0.96319</td>
<td>0.91324</td>
<td>0.96893</td>
<td>0.98097</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Results of data preprocessing.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>MAPE</th>
<th>RMSE</th>
<th>MAE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>Raw</td>
<td>0.042456</td>
<td>12.77</td>
<td>8.8601</td>
<td>0.87336</td>
</tr>
<tr>
<td></td>
<td>Denoised</td>
<td>0.021922</td>
<td>6.7636</td>
<td>4.6308</td>
<td>0.96564</td>
</tr>
<tr>
<td>Elman</td>
<td>Raw</td>
<td>0.05083</td>
<td>14.941</td>
<td>10.53</td>
<td>0.82287</td>
</tr>
<tr>
<td></td>
<td>Denoised</td>
<td>0.037648</td>
<td>10.737</td>
<td>7.9483</td>
<td>0.91276</td>
</tr>
<tr>
<td>ELM</td>
<td>Raw</td>
<td>0.04226</td>
<td>13.382</td>
<td>8.9012</td>
<td>0.91276</td>
</tr>
<tr>
<td></td>
<td>Denoised</td>
<td>0.030312</td>
<td>9.2816</td>
<td>6.4739</td>
<td>0.93255</td>
</tr>
<tr>
<td>LSSVM</td>
<td>Raw</td>
<td>0.033494</td>
<td>11.15</td>
<td>7.0784</td>
<td>0.90009</td>
</tr>
<tr>
<td></td>
<td>Denoised</td>
<td>0.017279</td>
<td>5.7475</td>
<td>3.7111</td>
<td>0.97412</td>
</tr>
<tr>
<td>MICA</td>
<td>Raw</td>
<td>0.019025</td>
<td>7.7413</td>
<td>4.0490</td>
<td>0.95192</td>
</tr>
<tr>
<td></td>
<td>Denoised</td>
<td>0.0093577</td>
<td>3.8705</td>
<td>2.0177</td>
<td>0.98827</td>
</tr>
</tbody>
</table>

Figure 5: Continued.
The RF algorithm gives better prediction accuracies with less RMSE, MAPE, and MAE values than the models created without this feature selection algorithm. For all the models, the MICA-based model with RF gives a good prediction accuracy with RMSE \( \approx 4.9431 \) and MAE \( \approx 3.3874 \), while the Elman-based model without RF exhibits weak prediction performance with RMSE \( \approx 29.17 \) and MAE \( \approx 22 \). Besides, box plots of the relative errors of the different models with and without the RF algorithm are given in Figure 5(f). Clearly, the models with the RF algorithm show less relative errors and better performance than the models without this feature selection algorithm.

**Table 6: Results of feature selection.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>MAPE</th>
<th>RMSE</th>
<th>MAE</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>With RF</td>
<td>0.035871</td>
<td>11.662</td>
<td>7.6801</td>
<td>0.90811</td>
</tr>
<tr>
<td></td>
<td>Without RF</td>
<td>0.087257</td>
<td>24.312</td>
<td>18.261</td>
<td>0.67848</td>
</tr>
<tr>
<td>Elman</td>
<td>With RF</td>
<td>0.043166</td>
<td>12.187</td>
<td>9.2569</td>
<td>0.89532</td>
</tr>
<tr>
<td></td>
<td>Without RF</td>
<td>0.10501</td>
<td>29.17</td>
<td>22.001</td>
<td>0.5628</td>
</tr>
<tr>
<td>ELM</td>
<td>With RF</td>
<td>0.040099</td>
<td>10.775</td>
<td>8.4514</td>
<td>0.91802</td>
</tr>
<tr>
<td></td>
<td>Without RF</td>
<td>0.075587</td>
<td>24.78</td>
<td>15.914</td>
<td>0.61203</td>
</tr>
<tr>
<td>LSSVM</td>
<td>With RF</td>
<td>0.029561</td>
<td>9.7406</td>
<td>6.4121</td>
<td>0.93326</td>
</tr>
<tr>
<td></td>
<td>Without RF</td>
<td>0.036291</td>
<td>11.169</td>
<td>7.7067</td>
<td>0.91591</td>
</tr>
<tr>
<td>MICA</td>
<td>With RF</td>
<td>0.015686</td>
<td>4.9431</td>
<td>3.3874</td>
<td>0.98286</td>
</tr>
<tr>
<td></td>
<td>Without RF</td>
<td>0.02261</td>
<td>8.1736</td>
<td>4.8672</td>
<td>0.95526</td>
</tr>
</tbody>
</table>

Figure 5: Prediction results of different algorithms: (a) BP prediction results; (b) Elman prediction results; (c) ELM prediction results; (d) LSSVM prediction results; (e) MICA prediction results; (f) boxplot of relative errors of different algorithms.
700MW to 1000MW. The sampling interval of 1 minute and a boiler load varying from collected from the DCS of an operational power plant with a dataset of more than 5000 real-operation data samples was cated to the application of the proposed algorithm in combustion optimization of the boiler to reduce the NOx emission. The modeling accuracy can be improved by using them. Second, combinatorial algorithm can obtain better prediction accuracy than single model. Moreover, the mean absolute percentage errors of the proposed MICA-based NOx emission model are within 1%, and the proposed algorithm can meet industrial demand. Future research will be dedicated to the application of the proposed algorithm in combustion optimization of the boiler to reduce the NOx emission.

Data Availability

A dataset of more than 5000 real-operation data samples was collected from the DCS of an operational power plant with a sampling interval of 1 minute and a boiler load varying from 700 MW to 1000 MW.

4.5. Classification Model Analysis. We show here the results of classification models established based on three datasets: T1, T2, and T3. The results are demonstrated in Table 7, where A, B, C, and D represent the BP, Elman, ELM, and LSSVM learning methods, respectively. For all datasets, the C4.5 algorithm works well and achieves an accuracy rate of up to 80%.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Category</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>20</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>28</td>
<td>20</td>
<td>22</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>26</td>
<td>17</td>
<td>28</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>23</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>2</td>
<td>23</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>29</td>
<td>19</td>
<td>27</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusions

Accurate physical modeling of coal-fired boilers can be quite challenging due to system complexity. In this work, we propose a model intelligent combinatorial algorithm to create a NOx emission model for coal-fired power plants based on real-operation data. The input data samples are firstly preprocessed to remove outliers. Then, wavelet denoising is employed to reduce the noise. The random forest (RF) algorithm is exploited to select the best input variables for improving the model accuracy. On this basis, the C4.5 algorithm is combined with the basic modeling algorithms and used to establish the NOx emission prediction model. After comparison between different models, several conclusions can be drawn. First, data preprocessing and feature selection strategies are effective and promising. The modeling accuracy can be improved by using them. Second, combinatorial algorithm can obtain better prediction accuracy than single model. Moreover, the mean absolute percentage errors of the proposed MICA-based NOx emission model are within 1%, and the proposed algorithm can meet industrial demand. Future research will be dedicated to the application of the proposed algorithm in combustion optimization of the boiler to reduce the NOx emission.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

X. J. Chen designed the research and the article structure and revised the manuscript. H. Y. Zhang carried out the experiments and revised the manuscript. X. X. Xing and H. W. Qin revised the manuscript. All authors read and approved the final paper.

Acknowledgments

The project was supported by the Changchun Science and Technology project (Grant no. 17DY030).

References


