Model NOx Emissions Emitted from Coal-fired Boilers Based on Differential Evolution Fast Learning Network

Zhiwen Shen and Qingwei Li*

College of Energy and Mechanical Engineering, Shanghai University of Electric Power, Shanghai, China

*Corresponding author email: boiler@shiep.edu.cn

Abstract. Low NOx combustion optimization is a simple, efficient and inexpensive NOx emission reduction technology for coal-fired power plants. Establishing NOx prediction model is an important part of this technology. Fast learning network (FLN) is an improved neural network proposed in recent years, which is simple and efficient. However, randomly generated hidden layer thresholds and input weights would affect the performance of FLN. To solve this problem, differential evolution (DE) algorithm is employed to optimize hidden layer thresholds and input weights. The proposed model was applied to some 330MW coal-fired boiler together with FLN. Each model was repeated 51 times to consider the randomness of both models caused by the randomness of DE algorithm and the stochastic initialization of the original FLN. Results showed that the model has better generalization ability, retaining the good approximation ability and stability. Besides, optimization process of the proposed model is very fast and fit for online combustion optimization.

Keywords: FLN; NOx emissions; Differential evolution; Coal-fired boiler; Hidden layer thresholds; Input weights.

Abbreviations

| Abbreviation | Description |
|--------------|-------------|
| AI           | artificial intelligence |
| ELM          | extreme learning machine |
| DE           | differential evolution |
| DE-FLN       | differential evolution fast learning network |
| FLN          | fast learning network |
| LOOCV        | leave-one-out cross validation |
| NOx          | nitrogen oxides |
| RMSE         | root mean square error |
| n            | number of input layer nodes |
| m            | number of hidden layer nodes |
| min          | lower bound of the ith variable |
| max          | the upper bound of the ith variable |
| \( W^{oi}_{lr} \) | connection weight of the lth output layer node and the rth input layer node |
| \( W^{oh}_{lk} \) | connection weight of the lth output layer node and the kth hidden layer node |
| \( W^{in}_{kt} \) | connection weight of the kth hidden layer node and the tth input layer node |
| X            | input matrix |
| Y            | target output matrix |
| G            | hidden layer output matrix |
| W            | output layer weight |
| N            | number of observation samples |
| l            | number of output layer nodes |
| \( X^0_{j,i} \) | the initialized ith variable of the j individual |

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd
1. Introduction

Although renewable energy grows fast in recent years, coal-fired power generation is still an important part of energy utilization in China and even the world[1]. Reduction of nitrogen oxides (NOx) of power plants is the key of the clean coal utilization. Low-nitrogen combustion optimization based on artificial intelligence (AI) is simple, efficient and low cost technology and has been a research hot in the thermal engineering area[2-3]. Establishing an accurate NOx prediction model is an essential part of combustion optimization[2-6]. Neural network has been extensively studied in the area of combustion optimization[6-7]. In 2004, GB Huang et al.[8] proposed a new neural network, namely extreme learning machine (ELM). ELM overcomes the defects of traditional neural networks like time-consuming and being prone to falling into local optima. To achieve better regression accuracy and stability, Guoqiang Li[9] proposed a double parallel forward neural network, namely fast learning network (FLN). Hidden layer thresholds and input weights of FLN are randomly generated and the performance is affected as a result. Quantum adaptation bird swarm algorithm[10] and adaptive quantum grey wolf optimization[11] have been proposed to optimize these parameters to establish NOx emission prediction model. With the increase of the node of hidden layer, parameters needed to be optimized increase greatly, making the parameter optimization problem very complicated. Differential evolutionary (DE) algorithm proposed by Rainer Storn and Kenneth Price has advantages of few control parameters, simple and efficient for optimization problems[12]. So this study attempted to utilize DE to solve the complicated parameter optimization problem of FLN for the first time. The rest of the paper is: FLN and DE algorithms are introduced first. Then the optimization of FLN parameters based on differential evolution is described, followed by the experiment and analysis. At last, this paper is concluded and the future work is given.

2. Fast Learning Network and Differential Evolutionary Algorithm

2.1. Theory of Fast Learning Network

FLN consists of a three layer neural network and a single layer neural network. The two neural networks are parallel. The output layer can receive information from the hidden layer and the input layer simultaneously[9]. The structure of FLN is shown in Figure 1.
Suppose there are \( N \) groups of observation samples \((x_i, y_i)\), where \( x_i=[x_{i1}, x_{i2}, \ldots, x_{in}] \) and \( y_i=[y_{i1}, y_{i2}, \ldots, y_{in}]^T \) are the input and output of the \( i \)th observation sample. Then FLN can be formulated as

\[
\begin{align*}
  y_{ji} &= \sum_{r=1}^{n} W_{jr}^{oi} x_{j} + \sum_{k=1}^{m} W_{jk}^{oh} g \left( b_k + \sum_{r=1}^{n} W_{jr}^{ih} x_{j} \right), \\
  y_{j1} &= \sum_{r=1}^{n} W_{jr}^{oi} x_{j} + \sum_{k=1}^{m} W_{jk}^{oh} g \left( b_k + \sum_{r=1}^{n} W_{jr}^{ih} x_{j} \right), \\
  y_{j2} &= \sum_{r=1}^{n} W_{jr}^{oi} x_{j} + \sum_{k=1}^{m} W_{jk}^{oh} g \left( b_k + \sum_{r=1}^{n} W_{jr}^{ih} x_{j} \right), \\
  & \vdots \\
  y_{jN} &= \sum_{r=1}^{n} W_{jr}^{oi} x_{j} + \sum_{k=1}^{m} W_{jk}^{oh} g \left( b_k + \sum_{r=1}^{n} W_{jr}^{ih} x_{j} \right)
\end{align*}
\]

where \( W_{jr}^{oi}, W_{jk}^{oh} \) and \( W_{jr}^{ih} \) represent the connection weight of the \( l \)th node of the output layer and the \( r \)th node of the input layer, the connection weight of the \( l \)th node in the output layer and the \( k \)th node of the hidden layer and the connection weight of the \( k \)th node of the hidden layer and the \( t \)th input layer node, respectively. \( b_k, m \) and \( g(\cdot) \) represent the threshold of the \( k \)th node of the hidden layer, the node number of hidden layer and the hidden layer activation function, respectively.

Further, equation (1) can be simplified as

\[
Y = W^{oi} X + W^{oh} G = \left[ W^{oi} \begin{bmatrix} X \\ G \end{bmatrix} \right] = \begin{bmatrix} X \\ G \end{bmatrix}
\]

where \( X, Y \) and \( G \) are the input matrix, the target output matrix and the hidden layer output matrix, respectively.

\[
W = \begin{bmatrix} W^{oi} & W^{oh} \end{bmatrix}_{(n+m) \times m}
\]

\[
G = \begin{bmatrix} g(W_{11}^{ih} x_{1} + b_{1}) & \ldots & g(W_{1n}^{ih} x_{n} + b_{1}) \\ \vdots & \ddots & \vdots \\ g(W_{m1}^{ih} x_{1} + b_{m}) & \ldots & g(W_{mn}^{ih} x_{n} + b_{m}) \end{bmatrix}_{m \times N}
\]

At last, the output weight matrix can be obtained by:
\[ W = Y \left[ X^t \right] = Y\beta^t \]  

\[
\begin{align*}
W^{oi} &= W(1 : l, 1 : n) \\
W^{ob} &= W(1 : l, (n + 1) : (n + m))
\end{align*}
\]  

The mentioned calculation procedures of FLN is:
1) Randomly generate hidden layer thresholds and input weights;
2) Calculate the hidden layer output matrix G according to equation (4);
3) Calculate the output layer weight matrix \( \hat{W} \) according to equation (5);
4) Obtain \( W^{oi} \) and \( W^{ob} \) according to equation (6).

2.2. Differential Evolution
Differential evolution evolve individuals by several operations like other heuristic search algorithms. Generally, the operations include initialization operation, mutation operation, crossover operation, and selection operations[13].

2.2.1. Population Initialization. An initialized population is generated stochastically within the feasible solution space of corresponding problem as below:
\[
X^0_{j,i} = x^\text{min}_i + U(0, 1) \times (x^\text{max}_i - x^\text{min}_i) 
\]  

Where \( x^0_{j,i} \), \( x^\text{min}_i \) and \( x^\text{max}_i \) represent the initialized ith variable of the j individual, the lower bound and the upper bound of the ith variable.

2.2.2. Mutation Operation. The mutation operation is carried out among selected individuals to evolve the individuals and is given by
\[
v^G_j = x^G_i + F \times (x^G_{j,i} - x^G_{j,j})
\]  

where \( j_i \neq j_2 \neq j_3 \neq j \). F is the mutation factor.

2.2.3. Crossover Operation. In order to further increase the population diversity, cross operation is carried out as follows:
\[
u^G_{j,i} = \begin{cases} 
  v^G_{j,i}, & \text{if } \text{rand}_i \leq \text{CR} \text{ or } i = i_n \\
  x^G_{j,i}, & \text{Otherwise}
\end{cases}
\]  

where in is random integer. \( \text{rand}_i \) is a uniformly distributed random number in the interval [0, 1].CR is the crossover rate.

2.2.4. Selection Operation. Greedy selection strategy is employed to select better individuals for next generation based on the fitness value of each individual. For the minimization problem, the selection operation can be described as:
\[
x^{G+1}_j = \begin{cases} 
  u^G_j, & \text{If } f(u^G_j) < f(x^G_j) \\
  x^G_j, & \text{Otherwise}
\end{cases}
\]  

3. Optimization of FLN Parameters Based on Differential Evolution
The prediction performance of the FLN is affected by randomly generated input weights and hidden layer thresholds[9]. However, the number of FLN parameters increases with the increase of input nodes and hidden layer nodes dramatically, making the parameter optimization of FLN very complicated. Differential evolution algorithm is famous for few control parameters, simple and efficient. So this study attempts to utilize differential evolution algorithm to solve the complicated
parameter optimization problem of FLN. Leave-one-out cross validation (LOOCV) is used when training the model [14]. The optimization process is shown in Figure 2.

**Figure 2.** Flow chart of DE-FLN model.

4. Experiment and Analysis

4.1. Background Test

The proposed model was applied to some 330 MW power plant boiler. The thermal adjustment test was reported in the literature [15]. During this test, 26 parameters including load, coal feeder speed, primary wind speed, coal quality and so on were selected for the affecting factors, which can be taken as inputs of the proposed DE-FLN. 20 cases were obtained in total. Case 20 was selected as the test sample, and the remaining cases were used as training samples.

4.2. Related Parameter Settings and Evaluation Index

The maximum iteration number and the population size were 200 and 20, respectively. The mutation factor and crossover rate were set as 0.75 and 0.25, respectively. Sig function was selected as the hidden layer activation function. The hidden layer node number was 30. The optimization process was taken 51 times independently.

In order to comprehensively evaluate the proposed model, the root mean square error (RMSE) was introduced as below

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]  

(11)

4.3. Results and Discussion

Predicted results of DE-FLN and FLN are shown in Figure 3. Figure 3(a) shows the original values, predicted values based on DE-FLN and FLN. Figure 3(b) shows the predicted errors based on DE-FLN and FLN. It should be mentioned that the predicted results are the results corresponding to
the run with the medium test RMSE. For the training samples, DE-FLN and FLN both show good approximation ability. However, the DE-FLN model show better generalization ability, which is shown clearly in Figure 3(b).

Figure 3. Predicted results based on DE-FLN and FLN.

DE is a kind of stochastic algorithm. To study the stability of the proposed model, 51 independent runs were carried out. The predicted results of training samples and test samples are shown in Fig. 4. It can be seen that the stability of the two models is both satisfying for training samples. The stability of generalization for the two models is also similar. The stability of generalization of DE-FLN should be improved in the future work.

Figure 4. Predicted error distributions of 51 independent runs.

A typical optimization process of DE-FLN is shown in Fig. 5. The predicted error based on leave-one-out method is decreasing during the optimization process. This explain why the performance of the DE-FLN is better than FLN. The optimization process is very fast and can find satisfactory results within 200 iterations, showing that the proposed model is fit for online combustion optimization.
5. Conclusions
Randomly generated hidden layer thresholds and input weights have effects on the performance of FLN. To solve this problem, DE algorithm is employed to optimize hidden layer thresholds and input weights. The proposed model was applied to some 330MW coal-fired boiler together with FLN. Results showed that generalization ability is enhanced greatly considering the randomness of DE algorithm and the stochastic initialization of the original FLN. Besides, the optimization process is very fast and fit for online combustion optimization. In the future, the stability of generalization should be improved further.

Acknowledgments
The authors declare that they have no conflict of interest.
This work was sponsored by National Key R&D Program of China(2018YFB0604204).

References
[1] BP (2019) BP statistical review of world energy(68 th edition)
[2] Li QW, Yao GH (2017) Improved coal combustion optimization model based on load balance and coal qualities. Energy 132: 204-212
[3] Zhou H, Zheng LG, Cen KF (2010) Computational intelligence approach for NOx emissions minimization in a coal-fired utility boiler. Energy Conversion and Management 51: 580-586
[4] Li QW, Wu J, Wei HQ (2018) Reduction of elemental mercury in coal-fired boiler flue gas with computational intelligence approach. Energy 160: 753-762
[5] Tang ZH, Zhang ZJ (2019) The multi-objective optimization of combustion system operations based on deep data-driven models. Energy 182:37-47
[6] Song JG, Romero CE, Zheng Y, He BS (2016) Improved artificial bee colony-based optimization of boiler combustion considering NOx emissions, heat rate and fly ash recycling for on-line applications. Fuel 172: 20-28
[7] Ilamathi P, Selladurai V, Balamurugan K, Sathyathan VT (2013) ANN-GA approach for predictive modeling and optimization of NOx emission in a tangentially fired boiler. Clean Technologies and Environmental Policy 15:125-131
[8] Huang GB, Zhu QY, Chee KS (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. Neural networks 2: 985-990
[9] Li GQ (2013) Research of a novel artificial intelligent technology and its application to boiler combustion optimization. Yanshan University
[10] Niu PF, Wang QY, Ma YP, Zhao QC, Chen K, Zhao Z (2017) Study on NOx emission from boiler based on quantum adaptation bird swarm algorithm. Acta Metrologica Sinica 38(6): 770-775

Figure 5. Typical optimization process of the parameters of FLN.
[11] Niu PF, Shi CJ, Liu N, Ma YP, Wu ZL, Li JB (2018) Prediction model for boiler NOx emission based on adaptive quantum grey wolf optimization. Journal of Chinese Society of Power Engineering 38(4):278-285

[12] R Storn, K Price (1997) Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization. Journal of Global Optimization 11(4): 341-359

[13] Zou DX, Steven L, Wang GG, Li ZY, Ouyang H (2016) An improved differential evolution algorithm for the economic load dispatch problems with or without valve-point effects. Applied Energy 181:375-390

[14] Wu ZY, Luo WC, Chen Y (2015) Optimal parameter based on pareto law for radial basis function approximation modeling. Systems Engineering & Electronics 37(5):1195-1200

[15] Li GQ, Niu PF, Zhang WP, Liu YC (2013) Model NOx emissions by least squares support vector machine with tuning based on ameliorated teaching-learning-based optimization. Chemometrics and Intelligent Laboratory Systems 126: 11-20