Research and Simulation of Boiler Combustion System Based on Convolution Neural Network

Chunming Ouyang\textsuperscript{1,*}, Li Xiao\textsuperscript{2}, Zhibin Xu\textsuperscript{2} and Weidong Zhang\textsuperscript{3}

\textsuperscript{1}Electric Power Research Institute of Guangdong Power Grid Co., Ltd., Guangzhou, Guangdong, China
\textsuperscript{2}Guangdong electric power academy Energy Technology Co., Ltd, Guangzhou, Guangdong, China
\textsuperscript{3}School of Electronic, Infomation and Electrical, Shanghai Jiao Tong University, Shanghai, China

*Corresponding author e-mail: olivia89@foxmail.com

Abstract. This paper developed a coal-fired boiler modeling method based on CNN (convolutional Neural Network). In recent years, the deep learning using CNN has made remarkable achievements in image research field. This paper combined the image convolution method and the signal system convolution theory in coal-fired boiler modeling, which regards the DCS records as time series signal. This network is consisted of two convolutional layers which map the input to the feature map, followed by two fully connected layers to model the feature to the output. All the parameters in the network, including the convolution kernel parameter are optimized by minimizing the cost function. This model provided a good solution of modeling coal-fired boiler system with characters of time delay, highly nonlinear and multi-variable coupling.

1. Introduction
China is the largest country on coal producing and consuming in the world, and thermal power is 73.8% of the nation’s total generated energy \cite{1}. Therefore thermal power plays an important role in power generation. Coal-fired boiler is a core part of thermal power generation, so for generator unit, precise modeling and simulation are an importance basis for operation survey, fault detection, scheme making, systems maintenance and combustion optimization.

However, combustion mechanism of coal-fired boiler is complex. Types of data which is detected by distributed control system are up to 60 in each time point. Limited by physical environment, control of variables is not in one step. Not only great time delay but also highly nonlinearity exists between the input and the output.

The traditional coal-fired boiler mechanism model is composed of a series of algebraic equations described in physical simulation of the boiler. Complex combustion process accounts for high time cost of traditional modeling method and high requirement of operator’s professional knowledge level. More importantly, as the system operates and the equipment constantly ages, difference between simulation process and real physical process will be bigger and bigger.

In recent years, with the development of machine learning, heuristic modeling algorithm has been widely used in industrial production. Mellit \cite{2} solving some problem in photovoltaic system explained application of artificial intelligence system synoptically. In the reference \cite{3}, the steady-
state operation data collected by the thermal test was applied to model the 600MV tangentially fired boiler. At the same time, genetic algorithm was used to optimize the input of the model. In the reference [4], support vector machine was used to model the emission of NOx in combustion. To reduce NOx emission, the model parameters was optimized by the particle population algorithm. In the reference [5], the minimum variance support vector algorithm was used to build the boiler combustion model, and the way to update the model by online learning was achieved.

In addition, neural network model with strong ability to express nonlinearity and coupling fully demonstrates practical value in industrial applications. In the reference [6], for 800MW boiler combustion model, by online learning, Booth built the neural network model which provides reference and advice about operation parameters adjustment for operators. In the reference [7], modeling of boiler efficiency and NOx emission was established by BP neural network and dynamic fuzzy neural network, the advantage and disadvantage and the suitable environment of two kinds of network analysed. In the reference [8], after difficulties and problems being analysed, 1000MW ultra-supercritical boiler modeling get a great effect with the help of fuzzy neural network model.

This paper developed a coal-fired boiler modeling method based on convolutional neural network and combined the image convolution in coal-fired boiler modeling which regards the DCS records as time series signal. A two-layer convolution neural network and a fully connected neural network are constructed to model the feature map mapped by convolution kernel. Including the convolution kernel parameters, all the parameters in the networks are adjusted and optimized by minimizing the cost function. Consequently, this model can provide a good solution of modeling coal-fire boiler system with character of time delay, high nonlinear and multi-variable coupling.

2. Model foundation of artificial neural network

Artificial neural network is a fitting model which imitates the transmission of human neural signal [9]. Its output signal is expected to come from former true experience which is gathered to be a training set. The process that network learning is realized by iterative optimization of parameters is called training.

A neural network consists of multiple neurons that are partially or fully connected to each other. As shown in the figure 1, there is the working principle of a single neural. X={x_1,x_2,…,x_m} is the output signal of the upper layer, which is the input signal of the current neuron; b is the bias of the input signal; y is the output signal of the current neuron; h(x) is the activation function of the neuron, which maps the input signal to the output signal domain under the action of neurons. So the mathematical description of the working process of a single neuron is

\[
\begin{align*}
  y' &= -w^T x + b \\
  y &= h(y')
\end{align*}
\]

![Figure 1. Working principles of a single neuron.](image)

Different mapping functions can be chosen as activation function according to the requirements of the model. For example
• Sigmoid function, \( h(t) = \frac{1}{1+e^{-t}} \) [10]
• Radial basis function, \( h(t) = \frac{1}{e^{-r^2}} \)
• Softplus
• Relu
• Tanh
• Linear kernel function

The commonly used convolutional neural network [11] is alternately composed of convolutional layers and pooling layers. Each neuron is mapped by activation function, and its output vector is called a feature map. The convolutional layer implements the inner product of the linear convolution kernel and the receptive field corresponding to input sequence.

\[
y(i, j) = x(i, j) \otimes w(t, s)
= \sum_{t=1}^{m} \sum_{s=1}^{n} x(i+t-1, j+s-1) w(t, s)
\]

(2)

The pooling layer aggregates data in the area to obtain the most representative value. The size of the area is determined by the pooling step. The rules of pooling depend on difference requirements, such as maximum pooling, minimum pooling, average pooling and so on. The alternative combination of convolutional layer and pooling layer can lower the dimension and extract abstract feature.

Convolution kernel is a form of vector, and the convolution formula which corresponds to the one-dimensional discrete sequence is

\[
y(t) = x(t) \otimes k(t) = \sum_{\tau=-\infty}^{\infty} x(\tau) k(t-\tau)
\]

(3)

\( k(t-\tau) \) represents the influence of the input signal on the output signal at the time of \( \tau \) from \( t \). For a finite time series \( x \), the value of the output \( y \) is mainly determined by the convolution kernel \( k \) sequence, and the length of the time delay depends primarily on the length of \( k \).

Analog to the one-dimensional discrete sequence, the two-dimensional convolutional neural network expands the convolution kernel appropriately to process the time series.

When the input of the neural network is \( \{x_{t-d}, x_{t-d+1}, \ldots, x_{t}\} \), a multi-time sequence whose delay is \( d \), the two-dimensional convolution kernel is replaced by the one-dimensional whose time dimension is \( d+1 \) because convolution is not required among variables at the same time.

\[
y(t) = \sum_{s=1}^{m} x(t+s-1) w(s)
\]

(4)

According to the calculation formula, the output vector dimension obtained by convolution with time delay of \( d \) is the same as the data dimension calculated by convolution with the time delay of \( n \) times of \( \{d_1, d_2, \ldots, d_n\} \)

\[
d_1 + d_2 + \cdots + d_n = d
\]

(5)

3. Model objects

In this paper, the data used in the model comes from the sampling data of DCS in February of a coal-fired unit in a thermal power plant. The number of monitored objects is up to 60, recorded every 15 minutes, and there are 96 sampling points every day, a total of 2592 sampling records. To avoid introducing extraneous variables which increase the model complexity and to enhance the robustness of the model by reducing the over-fitting, it is necessary to manually select key variables from the system acquisition data of DCS for modeling.
Safety, economy and environmental friendliness should be mainly taken into consideration to measure the boiler combustion system. The corresponding parameter indicators are as follows.

- Main steam pressure $P_{fs}$
- Thermal efficiency $f_t$
- Smoke emission $F_g$
- Boiler efficiency $f_b$

Based on these indicators, the prediction model of the boiler combustion system is established. These indicators are not only directly related to the operating parameters, but also have a mutual coupling relationship with each other.

The change in steam pressure is caused by the load’s change. Therefore, the supply of fuel should be adjusted to maintain the balance between the blown air and the drawn air. Thermal efficiency and boiler efficiency represent the percentage of thermal power which is converted to electric energy. They are global indicators which are closely related to combustion, so these processes should be coordinated to achieve the global optimization. Smoke emission is mainly determined by the characteristic of the fuel and the adequacy of combustion.

11 parameters shown in table 1 are regarded as state variables related to the four outputs.

| Variable name                  | symbol | unit   |
|--------------------------------|--------|--------|
| Load (rate)                    | $L_r$  | %      |
| Fuel supply                    | $Q_f$  | t/h    |
| Total air volume               | $Q_{air}$ | t/h   |
| Inlet air temperature          | $T_{air}$ | °C    |
| Feed water flow                | $Q_w$  | t/h    |
| Feed water pressure            | $P_w$  | Kpa    |
| Feed water temperature         | $T_w$  | °C     |
| Total flow primary and secondary attemperation water | $f_{w1}$ | % |
| excess air coefficient         | $O$    | %      |
| Condenser vacuum               | $V$    | Kpa    |

Due to the lag effect, the value of 4d can be equal to 3, so 2589 input-output samples can be created in 2592 records, the first 80% taken as the training set and the last 20% taken as the test set.

Since the DCS records the absolute information of each variable, the range and the order of magnitude of the different variables also vary greatly. For example, the load rate and boiler efficiency are between 0 and 1, while the feed water pressure is at $10^1$ orders of magnitude, the inlet air temperature is at $10^2$ orders of magnitude, and the total air volume is even at $10^3$ orders of magnitude. Putting data of different orders of magnitude in the same model will greatly increase the training time. Especially in the optimized objective function (loss function), the large scale parameters will occupy the main part of the loss function, which cause the loss function to tend to optimize the large-scale ones. That is the reason why it is difficult to realize optimization, even optimize smaller-scale parameters in the opposite direction. For example, the loss function in this model will give priority to optimize the main steam pressure and smoke emissions, and it is hard to optimize thermal efficiency and boiler efficiency.

In summary, the original data needs normalizing before modeling.

$$S'_{train} = \frac{S_{train} - S_{train\cdot min}}{S_{train\cdot max} - S_{train\cdot min}}$$

$$S'_{test} = \frac{S_{test} - S_{train\cdot min}}{S_{test\cdot max} - S_{train\cdot min}}$$
$S_{train}$ is the input-output sequence of all training samples. $S_{test}$ is the input-output sequence of all test samples. $S'$ is the sequence of samples after normalization.

4. Experiment and Simulation

As mentioned earlier, the specific parameters of the model established in this paper are as follows.

• In the input layer, the value of $d$ representing time lag is equal to 3. As input vector, the size of $\{x_{t-3}, x_{t-2}, x_{t-1}, x_t\}$ is $11 \times 4$.
• In the first layer, there are two convolution kernel whose size is $1 \times 3$ in the convolution layer. The output is 2 feature maps, and the size of each map is $11 \times 2$.
• In the second layer, there are four convolution kernel whose size is $1 \times 2$ in the convolution layer. The output is 4 feature maps, and the size of each map is $11 \times 1$.
• The third layer is a fully connected layer with 20 implicit neurons.
• The fourth layer is a fully connected layer with 10 implicit neurons.
• The fifth layer is the output layer with 4 output neurons.

The activation function of all neurons is Sigmoid function.

In contrast, we have established two models of neural networks additionally.

Model 1: The input is the argument parameter at the current time point, and the output is the dependent variable parameter. It doesn’t need to be processed by the convolution layer, nor does it take the time lag of the boiler into consideration.

$$ f(x_t) = y_t $$

However, lack of time lag sliding window, the dimension of the input parameter is 11, and the hyperparameter becomes 1/4 of the original. In order to avoid the model over-fitting caused by too many parameters, the number of neurons in the hidden layer should be reduced appropriately. Therefore, the model parameters are as follows:

• In the input layer, the value of $d$ representing time lag is equal to 0, $\{x_{t}^{*}\}$ is a vector of size 11.
• The first layer is a fully connected layer with 8 implicit neurons.
• The second layer is a fully connected layer with 6 implicit neurons.
• The third layer is the output layer with 4 output neurons.

Model 2: Time lag sliding window is adopted on the basis of the model 1. The size of window is the same as that of the convolution neural network, and the value of $d$ is equal to 3, but the convolution processing layer was not included. In the convolution neural network, after being processed by two convolution layers, the output of the model is still $44(4 \times 11)$, and the number of parameters that is output to the fully connected layer is not changed, so the number of neuron in fully connected layers in the model 2 equals that of the convolutional neural network.

$$ f(\{x_{t-3}, x_{t-2}, x_{t-1}, x_t\}) = y_t $$

• In the input layer, the time lag $d$ is equal to 3. The size of $\{x_{t-3}, x_{t-2}, x_{t-1}, x_t\}$ is $11 \times 4$.
• The third layer is a fully connected layer with 20 implicit neurons.
• The fourth layer is a fully connected layer with 10 implicit neurons.
• The fifth layer is the input layer with 4 output neurons.

The mean squared error of loss function of three models can be calculated by the following formula.

$$ f_{loss} = \sum_{i} \sum_{t} (y_{i}^{*} - y_{i}) $$

Stochastic gradient descent (SGD) is used as the optimization algorithm. The learning rate is set to 0.06, and the value of the weight decay is equal to 1e-5, and the value of learning momentum is 0.95.

$$ \Delta w_{ij}^{t+1} = 0.95 \cdot \Delta w_{ij}^{t} + 0.05 \cdot 0.01 \cdot \nabla w_{ij} $$
5. Analysis of results
The contrastive analysis of the three models’ convergence speed on the training set and fitting effect on the test set is made by the simulation.

![Figure 2](image)

**Figure 2.** Convergence curves of three models on the training data set.

Respectively iterated the same number of times on the entire training set, the three models get accuracy on the test set shown in Table 2.

| Epoch | Model1 | Model2 | CNN Model |
|-------|--------|--------|-----------|
| 10    | 0.0723 | 0.0151 | 0.0128    |
| 20    | 0.0162 | 0.0130 | 0.0109    |
| 30    | 0.0144 | 0.0125 | 0.0089    |

It can be seen from figure 2 and Table 1 that in the case of fixed parameters, the convolutional neural network model converges faster than the other two models during training, and can converge to lower error values. In contrast to the other two models, with the same times of trainings, fitting degree of Model 3 on the training set is significantly improved.

![Figure 3](image)

**Figure 3.** Comparison of the predicted output of the main steam pressure.
To quantify the difference among the models, we introduce mean absolute error to measure the models’ quality.

$$mae = \frac{1}{n} \sum_{i=1}^{n} |y_i - y'_i|$$  \hspace{1cm} (12)
According to table 3, the prediction effects to outputs by three models are increased successively. Compared with model 1 without time-delay input, the prediction effect of model 2 is improved greatly, which indicate the existence of great time delay in combustion system. Neural network model based on convolution has a better prediction effect on model 2, which represent that the feature of original data can be extracted better after being mapped by the convolution layer. Thus the operation of coal-fired boiler can be reflected more accurately.

6. Conclusion
This paper developed a coal-fired boiler modeling method based on convolutional neural network. Three neural network models are applied to the modeling of coal-fired boilers, and the results of simulation are compared and analyzed in detail. Compared with Model 1 adopting a simple neural network, Model 2 that adopts a neural network with a time-delay window improves the output prediction effect obviously. Thus it can be proved that a large time delay exists in the combustion system. The prediction error of the neural network model mapped by convolution is further reduced than that of model 2, which indicates that the convolutional layer in the time dimension can map the data to a better feature space, which is beneficial to build more accurate simulation of boiler dynamic characteristics.

Acknowledgments
This work was sponsored by China Southern Power Grid Corp science and technology project (Contract No.: GDDW2120150301JJ00035).

References
[1] China Electric Power Enterprise Association, Analysis and Forecast Report of 2014 National Electricity Supply and Demand Situation,2014-2-25, http://www.cec.org.cn/guihuayutongji/gongxufenxi/dianligongxufenxi/2014-02-25/117272.html
[2] Mellit A, Kalogirou S A. Artificial intelligence techniques for photovoltaic applications: A review[J]. Progress in energy and combustion science, 2008, 34(5): 574-632.
[3] Zhou H, Cen K, Fan J. Modeling and optimization of the NOx emission characteristics of a tangentially fired boiler with artificial neural networks[J]. Energy, 2004, 29(1): 167-183.
[4] Zheng L, Zhou H, Wang C, et al. Combining support vector regression and ant colony optimization to reduce NOx emissions in coal-fired utility boilers[J]. Energy & Fuels, 2008, 22(2): 1034-1040.
[5] Gu Y, Zhao W, Wu Z. Online adaptive least squares support vector machine and its application in utility boiler combustion optimization systems[J]. Journal of Process Control, 2011, 21(7): 1040-1048.
[6] Booth R C, Roland W B. Neural network-based combustion optimization reduces NO x emissions while improving performance[C] //Dynamic Modeling Control Applications for Industry Workshop, 1998. IEEE Industry Applications 1998. IEEE, 1998: 1-6.
[7] ZHAO Min. Study of Modeling and Optimization for Combustion Process of the Boiler in Power Plants Based on Neural Network [D]. Hangzhou. Zhejiang university,2010.
[8] Liu X J, Kong X B, Hou G L, et al. Modeling of a 1000 power plant ultra super-critical boiler system using fuzzy-neural network methods[J]. Energy Conversion and Management, 2013, 65: 518-527.
[9] S. Haykin Neural networks, a comprehensive foundation (2nd)Prentice-Hall Inc., New Jersey, USA (1999)

[10] Von Seggern D H. CRC standard curves and surfaces with Mathematica[M]. CRC Press, 2006.

[11] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.