Study of a deep LSTM model for power prediction of micro-combustor

Hongyu Si $^{1,2,*}$, Xiaohui Liang$^{1,2}$, Suxiang Liu$^{1,2}$, Dongliang Hua$^{1,2}$, Dong Hou$^{3,*}$, Meirong Xu$^{1,2}$, Sijie Liu$^{4}$, Ning Mei$^{4}$

$^1$ Energy Institute, Qilu University of Technology (Shandong Academy of Sciences), Shandong Key Laboratory of Biomass Gasification Technology, Jinan 250014, China
$^2$ School of Energy and Power Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China
$^3$ Faculty of Natural Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway
$^4$ College of Engineering, Ocean University of China, Qingdao 266100, China

*Corresponding author e-mail: sihy@sderi.cn

Abstract. A deep “long short-term memory” (LSTM) model, which is a special recurrent neural network (RNN) in deep learning, was established. Data from orthogonal experiments of the micro-combustor was used for training the model. A nonlinear mapping of output power to the geometric parameters of the micro-combustor was built. After changing the geometric parameters of the micro-combustor, the output power was predicted, and the corresponding results were analyzed. The results show that the model from the deep LSTM has a relative error of 1.1% in power prediction, which is within the scope of engineering requirements.

Keywords: LSTM (Long Short-Term Memory), deep learning, nonlinear mapping, micro-combustor.

1. Introduction

As the complexity of industrial production increases, the requirements to establish appropriate monitoring, control, and fault diagnosis systems are getting higher. Commonly used device monitoring and fault diagnosis methods are based on data-driven approaches, model-based approaches and expert knowledge ones. [1-3] The development of sensors, data storage technology, and computer technology provides a foundation for monitoring large number of data and then extracting corresponding mathematical models from the data. Both physical model-based methods and expert knowledge-based approaches require a deep understanding of the physical processes of industrial production systems, resulting to a prolonged modeling cycle. [4] Deep learning is well known after AlexNet won the championship in the ImageNet Large Scale Visual Recognition Challenge in 2012. Deep learning has been widely used in image recognition, speech recognition, and natural language processing. The Recurrent Neural Network (RNN) has a special advantage in processing sequences. The LSTM network, as a variant of RNN, can process the entire sequences in which each group of data is not independent. [5] From a spatial perspective, it builds the correlation between each set of data. The LSTM network
eventually establishes a nonlinear mapping from input data to output data time. For a micro-combustor, changing some of its geometric parameters can affect its output power. However, common simulation methods cannot quickly calculate the change in the final output power when the geometric parameters are altered. In this paper, a deep LSTM algorithm establishes a nonlinear mapping of the geometric parameters of the micro-combustor to its output power.

2. Modeling method

2.1. Single layer LSTM network

LSTM can link the information between each set of data by learning long-term dependency. The outside of the LSTM network is recurrent in the manner of RNN. The internal self-recurrent and information filtering are performed by the LSTM cell. Each cell has the same type of input and output, and inside the unit there are gates which is a way to optionally let information through. The LSTM network structure replaces neuron in the traditional neural network with a cell. The internal structure of the LSTM cell is shown in Figure 1.

Each LSTM cell has two outputs: $C_t$ and $h_t$. $C_t$ stores the state of the network in a long-term, which is the hidden complex relationship between each set of data. $h_t$ reflects the influence of the current input data on the output. The LSTM cell consists mainly of input gate, forget gate, and output gate. The state of the cell, $C_t$, is controlled by the input gate and the forget gate, while the output value $h_t$ of the cell is controlled by the output gate.

![Figure 1. Internal structure of a LSTM Cell](image)

![Figure 2. Single layer LSTM structure](image)

2.1.1. Forward pass algorithm. The algorithm for these three gates is introduced as follows based on the forward pass.

1) Forget gate

The forget gates are mainly used to determine how much of the content at $t-1$ (meaning before the current time) is retained to the current time $t$, and the calculation formula is:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f)$$  \hfill (1)

2) Input gate

There are mainly two functions of the input gate: to input the data at the current time into the cell unit, and to calculate the states of current cell unit after the input.

The formula of the input data at the current time, $i_t$, is:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i)$$  \hfill (2)

The formula of the input unit state at the current time, $\tilde{c}_t$, is:

$$\tilde{c}_t = \sigma(w_c \cdot [h_{t-1}, x_t] + b_c)$$  \hfill (3)

After the current input, the LSTM algorithm needs to combine the previous state $c_{t-1}$, forget gate output $f_t$, input data at the current time $i_t$, and input unit state at the current time $\tilde{c}_t$, to calculate the current state of the cell $c_t$, the expression is:
\[ c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \]  \quad (4)

The symbol \( \circ \) means that the elements at the corresponding positions of the two matrices are multiplied. Due to the control of the forget gate, the LSTM can retain long-term information and also can store relevant information in the structure. Meanwhile, irrelevant information does not enter long-term memory due to the control of the input gate.

3) output gate

The output gate \( o_t \) controls the effect of long-term memory on the current output. The expression of \( o_t \) is:
\[ o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \]  \quad (5)

The final output of the LSTM network is determined by the output gate and the state of the cell unit:
\[ h_t = o_t \circ \tanh(c_t) \]  \quad (6)

The expression of the activation function sigmoid function used in the forward pass and its derivative are:
\[ \sigma(y) = \frac{1}{1+e^{-y}} \]  \quad (7)
\[ \sigma'(y) = \sigma(y)(1 - \sigma(y)) \]  \quad (8)

The expression of the tanh function and its derivative are:
\[ \tanh(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}} \]  \quad (9)
\[ \tanh'(y) = 1 - \tanh^2(y) \]  \quad (10)

After the forward pass calculation is performed, the LSTM algorithm needs to perform error back-propagation calculation to correct the weight \( W \) and the offset \( b \).

2.1.2. Error back-propagation calculation. The single-layer LSTM structure is shown in Figure 2. Each cell unit has two input values, one is the previous cell output \( h_{t-1} \). The parameters corresponding to this input include: forget gate weight \( W_{fh} \), offset \( b_f \), input gate weight \( W_{ih} \), offset \( b_i \), state weight \( W_{ch} \), offset \( b_c \), output gate weight \( W_{oh} \), offset \( b_o \). The other input is the data input at current time \( W_{xh}, W_{cx}, \) and the corresponding weights include: forget gate weight \( W_{fx}, \) input gate weight \( W_{ix}, W_{cx}, \) and output gate weight \( W_{ox}. \) Because of these two different input values, the error back-propagation calculation is performed by BPTT (Back Propagation Through Time) algorithm to calculate the propagation of the error along the direction of time and along the direction of output to the current input. The error back-propagation algorithm is based on the gradient descent method for iterative optimization, the weights and offsets are continuously corrected in order to obtain the minimum error between the forward pass result and the target value.

At time \( t \), the output value of the LSTM network is \( h_t \), and the expression defining the error \( \delta_t \) of the final output of the network is:
\[ \delta_t = \frac{\partial E}{\partial h_t} \]  \quad (11)

The error term mainly consists of two parts: the error \( \delta^T_{t-1} \) caused by the propagation in the time direction and the error \( \delta^T_{t-1} \) along input to output direction.

1) Algorithm for weight gradient along time direction

At time \( t-1 \), the error \( \delta^T_{t-1} \) caused by propagation along the time direction can be obtained by the chain rule, and its expression is:
\[ \delta^T_{t-1} = \delta^T_t \frac{\partial E}{\partial h_t} \]  \quad (12)

Since \( f_t, i_t, c_t, o_t \) are all functions of \( h_{t-1} \), the errors \( \delta^T_{f_t}, \delta^T_{i_t}, \delta^T_{c_t}, \delta^T_{o_t} \), which propagate to \( f_t, i_t, c_t, o_t \) at \( t-1 \), can be obtained by functions \( (4), (5), (6), (7), (8), (9), \) and \( (10) \). The expressions are:
\[ \delta^T_{f_t} = \delta^T_t \circ o_t \circ (1 - \tanh^2(c_t)) \circ c_{t-1} \circ f_t \circ (1 - f_t) \]  \quad (13)
\[ \delta^T_{t+1} = \delta^T_t \circ o_t \circ (1 - \tanh^2(c_t)) \circ \tilde{c}_t \circ i_t \circ (1 - i_t) \] (14)
\[ \delta^T_{c,t} = \delta^T_t \circ a_t \circ \tanh(c_t) \circ i_t \circ (1 - \tilde{c}_t)^2 \] (15)
\[ \delta^T_{o,t} = \delta^T_t \circ \tanh(c_t) \circ a_t \circ (1 - o_t) \] (16)

So the expression of the error back propagating to \( t-1 \) is:
\[ \delta_{t-1} = \delta^T_t \circ \frac{\partial E}{\partial h_{t-1}} = \delta^T_{f,t}W_{fh} + \delta^T_{i,t}W_{ih} + \delta^T_{c,t}W_{ch} + \delta^T_{o,t}W_{oh} \] (17)

The expression of the error propagating to any time \( k \) is:
\[ \delta^T_k = \bigoplus_{j=k}^{t-1} \delta^T_{f,t}W_{fh} + \delta^T_{i,t}W_{ih} + \delta^T_{c,t}W_{ch} + \delta^T_{o,t}W_{oh} \] (18)

2) The weight gradient propagating from output to input.

If the current layer is \( l \), then the error term of the \( l-1 \) layer is calculated from the weighted input of the \( l-1 \) layer using the error function, and its expression is:
\[ \delta_{t-1}^{l-1} = (\delta^T_{f,t}W_{fh} + \delta^T_{i,t}W_{ih} + \delta^T_{c,t}W_{ch} + \delta^T_{o,t}W_{oh})f^{l-1}(W_x x_t + b_x) \] (19)

3) The gradient algorithm to offset term.

The gradient of offset term is the offset gradient at each moment, and the expression of the output gate is:
\[ \frac{\partial E}{\partial b_{o,t}} = \frac{\partial E}{\partial (W_x x_t + b_x)} \frac{\partial (W_x x_t + b_x)}{\partial b} \delta^T_{o,t} \] (20)

The output offset gradients of other gates are calculated using the same formula. The offset gradient of the forgetting gate is \( \partial_{f,t}^T \), and the offset gradient of the input gate is \( \partial_{i,t}^T \).

![Figure 3. The structure of deep LSTM](image)

Through the equations (1) to (20), each weight and offset term required by forward pass and back gradient propagation of single-layer LSTM network are completed. Then an iterative update of each term to correct the initial weight and offset, until the network accuracy requirement is met, or the number of iterations is reached. [7]

2.2. Deep LSTM network.

The deep LSTM network is a variant of the recurrent neural network that enhances the expressive ability of the model. The schematic diagram of the deep LSTM network structure is shown in Figure 3. There is only one fully connected layer between \( x_t \) and output \( y_t \) at each moment for the single-layer LSTM network. There is a single-layer neural network between \( x_t \) and \( y_t \), while a neural network with a larger number of layers in the time direction.
The deep LSTM network has multiple hidden layers from $x_t$ to $y_t$. Therefore, higher-level information can be extracted from the input, which enhances the complexity and robustness of the model, and enables the modeling construction of complex physical processes.

3. Description of combustor model and data analysis

3.1. Description of combustor’s physical model
The combustor referred here is a device for transferring heat from combustion of a fuel to a heat exchange medium in a thermal cycle system by heat exchange principle. The device having a rated power of 35 kW and a rated fuel consumption of 3.8 kg/h. The main part of the body is made of Q235 steel, which is welded and threaded.

The structure of the combustor is shown in Figure 4. The main working body is composed of three parts.

- Fan: The fan is connected to the combustion chamber via a flange (17), and the required air volume of the combustion chamber is provided by the analysis section.
- Combustion chamber: the combustion chamber comprises a main combustion chamber (8), a combustion chamber constriction (7) and a secondary combustion chamber (6). A baffle (11) is provided in the main combustion chamber, the main function of which is to adjust the flow field in the combustion chamber. The ignition plug (10) is inserted into the main combustion chamber.
- Water jacket body: The water jacket body is mainly composed of a water jacket shell (1), a water jacket inner liner (3) and a hot fin (5). The water jacket shell and the inner liner form a sandwich layer, and internal radial hot fins are distributed to enhance heat transfer. The water jacket is the component used for heat exchange. [6]

![Figure 4. The structure of the combustor](image)

![Figure 5. The comparison of the predictions from LSTM and the actual data.](image)
The working principle of the device is as follows: the rotation of the motor drives the combustion air blower (12) and the atomizer (20) to rotate, and the fuel took in by the fuel pump (16) is sent to the atomizer by the oil pipe (19) to atomize. The air sent to the combustion chamber by the combustion air blower is mixed with atomized fuel in the main combustion chamber (8) and ignited by the ignition plug. The mixed gas is fully burned through the combustion chamber constriction (7) and the secondary combustion chamber (6), and then folded back. The heat was transferred to heat exchange medium inside the water jacket when passing through the inner wall of the water jacket and the hot fin (5). The heat exchange medium circulates through the entire pipeline by the water pump, and transfers the heat to the appropriate position. The wasted gas is discharged from mouthpiece (21).

3.2. Data test and analysis
There are 12 geometric parameters, as shown in Table 1. These 12 parameters may have a certain impact on the performance of the combustor. Since the influence of single factor on the output power of the combustor cannot be identified, the orthogonal experiment method was carried out.

| geometric parameter name | A | B | C | D |
|--------------------------|---|---|---|---|
| bending radius of impeller blade | E | F | G | H |
| opening angle of lower compartment | I | J | K | L |

Table 1. The geometric parameters
IOP Conf. Series: Earth and Environmental Science 514 (2020) 042073  doi:10.1088/1755-1315/514/4/042073

Table 2. Original data

| Index | A | B | C    | D    | E | F    | G    | H    | I    | J    | K    | L    | P/kW |
|-------|---|---|------|------|---|------|------|------|------|------|------|------|------|
| 1     | 60| 30| 1.7  | 90   | 16| 1    | 15.8 | 15   | 24   | 0    | 0    | 90   | 29.13|
| 2     | 60| 30| 1.7  | 90   | 14| 0    | 15.3 | 13   | 22   | upward shift 0.5 | left 0.5 | 85   | 28.85|
| 3     | 60| 30| 1.7  | 90   | 18| 3    | 16.3 | 17   | 20   | downward shift 0.5 | right 0.5 | 95   | 27.89|
| 4     | 60| 32| 1.5  | 85   | 16| 1    | 15.8 | 13   | 22   | upward shift 0.5 | right 0.5 | 95   | 28.31|
| 5     | 60| 32| 1.5  | 85   | 14| 0    | 15.3 | 17   | 20   | downward shift 0.5 | 0    | 90   | 28.29|
| 6     | 60| 32| 1.5  | 85   | 18| 3    | 16.3 | 15   | 24   | 0    | left 0.5 | 85   | 29.6 |
| 7     | 60| 34| 2    | 95   | 16| 1    | 15.8 | 17   | 20   | downward shift 0.5 | left 0.5 | 85   | 28.1 |
| 8     | 60| 34| 2    | 95   | 14| 0    | 15.3 | 15   | 24   | 0    | right 0.5 | 95   | 28.33|
| 9     | 60| 34| 2    | 95   | 18| 3    | 16.3 | 13   | 22   | upward shift 0.5 | 0    | 90   | 29.27|
| 10    | 62| 30| 1.5  | 95   | 16| 0    | 16.3 | 15   | 22   | downward shift 0.5 | 0    | 85   | 28.29|
| 11    | 62| 30| 1.5  | 95   | 14| 3    | 15.8 | 13   | 20   | 0    | left 0.5 | 95   | 28.7 |
| 12    | 62| 30| 1.5  | 95   | 18| 1    | 15.3 | 17   | 24   | upward shift 0.5 | right 0.5 | 90   | 28.4 |
| 13    | 62| 32| 2    | 90   | 16| 0    | 16.3 | 13   | 20   | 0    | right 0.5 | 90   | 29.07|
| 14    | 62| 32| 2    | 90   | 14| 3    | 15.8 | 17   | 24   | upward shift 0.5 | 0    | 85   | 28.23|
| 15    | 62| 32| 2    | 90   | 18| 1    | 15.3 | 15   | 22   | downward shift 0.5 | left 0.5 | 95   | 28.4 |
| 16    | 62| 34| 1.7  | 85   | 16| 0    | 16.3 | 17   | 24   | upward shift 0.5 | left 0.5 | 95   | 29.71|
| 17    | 62| 34| 1.7  | 85   | 14| 3    | 15.8 | 15   | 22   | downward shift 0.5 | right 0.5 | 90   | 28.21|
| 18    | 62| 34| 1.7  | 85   | 18| 1    | 15.3 | 13   | 20   | 0    | 0    | 85   | 29.95|
| 19    | 64| 30| 2    | 85   | 16| 3    | 15.3 | 15   | 20   | upward shift 0.5 | 0    | 95   | 30.13|
| 20    | 64| 30| 2    | 85   | 14| 1    | 16.3 | 13   | 24   | downward shift 0.5 | left 0.5 | 90   | 29.43|
| 21    | 64| 30| 2    | 85   | 18| 0    | 15.8 | 17   | 22   | 0    | right 0.5 | 85   | 27.8 |
| 22    | 64| 32| 1.7  | 95   | 16| 3    | 15.3 | 13   | 24   | downward shift 0.5 | right 0.5 | 85   | 28.19|
| 23    | 64| 32| 1.7  | 95   | 14| 1    | 16.3 | 17   | 22   | 0    | 0    | 95   | 29.08|
| 24    | 64| 32| 1.7  | 95   | 18| 0    | 15.8 | 15   | 20   | upward shift 0.5 | left 0.5 | 90   | 28.32|
| 25    | 64| 34| 1.5  | 90   | 16| 3    | 15.3 | 17   | 22   | 0    | left 0.5 | 90   | 28.93|
| 26    | 64| 34| 1.5  | 90   | 14| 1    | 16.3 | 15   | 20   | upward shift 0.5 | right 0.5 | 85   | 28.22|
| 27    | 64| 34| 1.5  | 90   | 18| 0    | 15.8 | 13   | 24   | downward shift 0.5 | 0    | 95   | 28.74|

The orthogonal table is L_{27}(3^{13}), in the parameter setting of the LSTM network, the timestep is set to 3, and the amount of training data is 21, and the whole group data is used for training. The batch-size is 21, and the network input is 12-dimensional, which includes 12 geometric parameters that affect the power of the combustor. The output of the network is 1 dimensional, and the output value is the combustor power. The number of hidden layer nodes is 100, the learning rate is 0.01, the number of iteration steps is 5000, and the loss function uses the mean square error function.

The predicted results are shown in Fig. 5. In the above combustor power prediction, the maximum relative error is 1.1%. The mathematical model established by the deep LSTM network can be used to predict the power of the micro-combustor.

Table 3. The relative error of the predictions and actual data

| Actual value/kW | Predicted value/kW | Relative error |
|-----------------|--------------------|---------------|
| 28.19           | 28.14              | 0.2%          |
| 29.09           | 28.75              | 1.1%          |
| 28.32           | 28.27              | 0.2%          |
| 28.93           | 28.85              | 0.2%          |
| 28.22           | 28.49              | 0.96%         |
| 28.74           | 28.64              | 0.3%          |

4. Conclusion
Through the deep LSTM network, a nonlinear mapping between the individual geometric parameters of the micro-combustor to the output power can be established. When several geometric parameters of the
combustor change, the output power can be predicted by a trained deep LSTM network, and the maximum relative error of the predicted result is 1.1%.

Acknowledgments
This work was financially supported by Key R and D project of Shandong Province (2019LYXZ027).

References
[1] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri. A review of process fault detection and diagnosis Part I: Quantitative model-based methods[J]. Computers and Chemical Engineering, 2003(27):293-311
[2] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri. A review of process fault detection and diagnosis Part II: Qualitative models and search strategies[J]. Computers and Chemical Engineering, 2003(27):313-326
[3] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, Kewen Yin. A review of process fault detection and diagnosis Part III: Process history based methods [J]. Computers and Chemical Engineering, 2003(27):327-346
[4] Khaoula Tidriri, Nizar Chatti, Sylvain Verron, Teodor Tiplica. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges[J]. Annual Reviews in Control, 2016(42):63-81
[5] ZHANG Jun Yang, WANG Hui Li, GUO Yang, HU Xiao. A review of deep learning [J/OL]. 2018, 35(7). 2017-08-17 [2017-08-17] http://www/arocmag/com/artcicle/02-2018-07-067.html
[6] SI Hongyu, The numerical simulation and experimental study on the flow and heat transfer characteristics of a micro-combustor, Qing Dao: The ocean university of China, 2006
[7] [Yoshua Bengio, Ian Goodfellow, Aaron Courville, Deep learning [M]: http://www.deeplearningbook.org/, 2015 284-290]