Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Electrode regulating system modeling in electrical smelting furnace using recurrent neural network with attention mechanism

Shenyi Ding[a], Zhijie Wang[a,*], Weijian Kong[a,*], Honghai Yang[b], Guangxiao Song[a]

[a] College of Information Science and Technology, Donghua University, Shanghai 201620, China
[b] College of Environmental Science and Engineering, Donghua University, Shanghai 201620, China

ARTICLE INFO

Article history:
Received 13 March 2019
Revised 10 May 2019
Accepted 20 May 2019
Available online xxx

Communicated by Dr. Nianyin Zeng

Keywords:
Dynamic system modeling
Recurrent neural network
Attention mechanism
Electrical smelting furnace
Data driven

ABSTRACT

Electrical smelting furnace (ESF) is the primary equipment to produce steer, nonferrous metals and other materials. In smelting process, it is beneficial to keep the smelting current stable within a reasonable range to improve product quality and reduce energy consumption. ESF is characterized by nonlinearity, strong coupling and time variation, which makes it difficult to establish accurate mathematical models. Therefore, we propose a data-driven recurrent neural network (RNN) using gated recurrent unit (GRU) with attention mechanism not only to establish the relationship between the electrode position and the smelting current but also to reveal the internal dynamic variation between three-phase currents for the electrode regulating system. Our proposed model is tested on the actual production data collected from a fused magnesium furnace (one kind of ESF) in Liaoning Province of China. Numerical results show RNN excels at processing sequential data and describing inner changes of the dynamic system; GRU alleviates long-term dependency problems in industrial big data; attention mechanism can identify and put emphasis on the crucial points containing key information in long sequential data. The results indicate that the proposed model is effective and feasible for the modeling of electrode regulating system.

1. Introduction

Electrical smelting furnace (ESF) is a device that smelts ores or metals through high temperature generated by electrode arcs. ESF has great flexibility in technology and can effectively remove impurities like sulfur and phosphorus, so it is widely used for producing important industrial materials such as steel, magnesia, and alumina. ESF is available in several styles and this paper mainly discusses three-phase alternating current (AC) ESF.

Maintaining the steadiness of the smelting current in ESF is conducive to reduce energy consumption (mainly electric energy). ESF is a kind of power-consuming furnace. During the smelting process, ESF may exert detrimental influence on the grid such as waveform distortion, voltage fluctuation, power factor reduction, and three-phase imbalance. For purpose of lightening these disadvantages, State Grid Corporation of China (SGCC) requires a contract with those enterprises using ESF to restrict their maximum electricity demand. Once the actual demand surpasses the specified demand, enterprises have to pay an extra fine. Since the voltage is basically stable in the smelting process, the difficulty of the demand (the specified demand) forecasting arises from the dynamical fluctuation of the smelting current. Scholars have already carried out some researches on demand forecasting issue. Yang and Chai [1] presented a radical basis function neural network (RBFNN) based demand forecasting method using partial autocorrelation function (PACF) to decide the input numbers of the neural network. Another demand forecasting model raised by Yang et al. [2] adopted bat algorithm to train the parameters of RBFNN and the forecasting results were optimized using a asymmetrical penalty function.

Maintaining the stability of the smelting current also contributes to improve product quality. There usually exists a rational current setpoint according to human experience, and the actual current can fluctuate within a reasonable range around this setpoint. However, people have to adjust the position of three-phase electrodes to change the length of the arcs, that is, to change the value of the arc resistance, to offset the variation of resistance of molten pool caused by granularity and impurity composition of raw materials as well as working stages alternation (e.g., heating and melting, feeding, and exhausting stages) so that the actual current can track this setpoint. Wu et al. [3] proposed an intelligent control method to classify smelting conditions and applied neural

* Corresponding authors.
E-mail addresses: wangzj@dhu.edu.cn (Z. Wang), kongweijian@dhu.edu.cn (W. Kong).

https://doi.org/10.1016/j.neucom.2019.05.060
0925-2312/© 2019 Published by Elsevier B.V.

network controller to handle normal and abnormal conditions based on simultaneous perturbation stochastic approximation (SPSA) and rule-based reasoning respectively. Wu et al. [4] put forward a hybrid intelligent optimal control technique consisting of PI control based feed-forward compensator, iterative learning control based feedback compensator, case based-reasoning based current presetting model, and rule based-reasoning based current controller. Ma and Zhu [5] designed a BP neural network as a controller to adjust the current value back to normal. Wu et al. [6] proposed a control-designed dynamic model of ESF and verified a series of control rules under disturbance-free circumstance, and then proved the convergence and robustness of close-loop control system under disturbance case. Other papers focus on current setpoints optimization. In [7], a rule based-reasoning based model of abnormal condition identification was established considering the current tracking errors, the change rate of the currents, and the duration for currents fluctuation. Accordingly, the setpoint was regulated to a proper value through case based-reasoning method.

Establishing an electrode regulating system model to obtain the relationship between the electrode position and the smelting current is the basis of the above researches. As a matter of fact, ESF features strong nonlinearity, strong coupling, hysteresis and is sensitive to boundary conditions. Wang et al. [8] calculated the mathematical relationship between electrode position and arc length, and then deduced a mechanism model for the smelting current. This is a beneficial attempt to help us better understand ESF in principle, but it is difficult to apply in practice, since there are too many interference factors in the actual smelting process. Thereby, the black box model plays an irreplaceable role. The neural network model, as a special black box model avoiding the cognitive obstacles in intermediate process, has been explored to apply in electrode regulating system modeling. Wu et al. [9] discussed a novel hybrid model on the basis of mathematical model and compensation Elman neural network model, and proved that neural network is helpful to improve the accuracy of the hybrid model.

Although neural networks improve the effect of electrode regulating system modeling, there are two remaining problems restricting their applications. On one hand, traditional neural networks can only handle a limited amount of data and are vulnerable to be trapped in local minimum when dealing with industrial big data. On the other hand, there are some crucial points in the data sequence of the smelting process, which current models have not made full use of. These crucial points contains the key information of the smelting process that is either explicit (such as working stages alternation) or implicit (such as granularity or impurity composition variation of raw materials). To overcome the aforementioned two problems, we employ gate recurrent unit (GRU) on the basis of vanilla recurrent neural network (RNN) and integrate attention mechanism. Recurrent neural network (RNN) is widely used for sequence processing, for instance, music classification [10,11] and machine translation [12,13]. With a well-designed connection between hidden layers, RNN is able to memorize information from all the past components of the sequential inputs implicitly [14]. However, RNN may encounter long-term dependency problems caused by gradient exploding or vanishing [15], especially in industrial process where large number of sequential data need to be dealt with. Gate recurrent unit (GRU) is proposed in [16] to solve long-term dependency problems. One of its vital improvements is to introduce gated units so that the connection weights between hidden layers can adjust dynamically, enabling networks to collect useful information from previous inputs even if these inputs have pasted a long time. Attention mechanism was firstly presented in image processing field to imitate human's behavior on visual images. When people look at an image, they don't actually see every pixel of the whole image at one time. Instead, they focus their attention on specific parts of the image as required. Later, attention mechanism is widely used in nature language processing [17,18], audio recognition [19,20], and so on to assist primary models in concentrating on key information. Inspired by these researches, this paper combines the attention mechanism with GRU-RNN to capture the hidden critical points in the long sequential data of the smelting process. Experiments in a typical kind of ESF, fused magnesium furnace (FMF), show that our proposed model has higher prediction accuracy than current models and can be applied into practice.

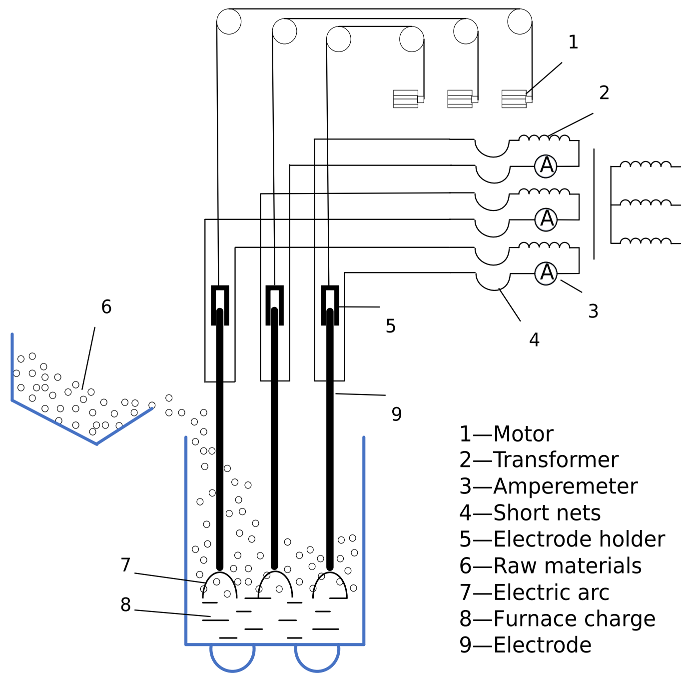


Fig. 1. Schematic diagram of equipment for electrical smelting furnace.

The rest of this paper is organized as follows: Section 2 describes the smelting process in ESF. Section 3 presents the electrode regulating model based on GRU-RNN with attention mechanism. Section 4 explains the selection of model structure parameter and Section 5 gives the analysis of the results. Section 6 gives conclusions and discussions for this paper.

2. Description of an electrical smelting furnace

A brief description of the production equipments in ESF is illustrated in Fig. 1. The whole production process can be divided into three stages: starting, smelting and ending. At the starting stage, operators spread a layer of raw materials at the bottom of the furnace and insert three-phase electrodes into a suitable position, and then some broken electrode blocks are added between three electrodes to induce the formation of arcs. After these preparations, AC power is exerted to electrodes and the smelting process starts. During the smelting stage, raw materials are fed into the furnace discontinuously. Some of them are smelted by the heat of arcs and others are smelted in the molten pool, both lead to the rise of molten surface level and the shortening of arcs. Operators need to adjust the position of electrodes in light of their experience to make sure that the smelting currents are stable around the setpoints. There are two special stages in the smelting stage, which are feeding stage and exhausting stage. The newly added solid raw materials' resistivity is higher than the resistivity in the molten pool, thus the current will decrease during the feeding phase, and the operators are called for adjusting the position of the electrodes carefully to prevent the unheated raw material from falling directly into the molten pool, resulting in a decrease in product quality. The smelting process also produces a portion of gas that affects the product quality, so during the exhausting phase, the operators move the electrodes to allow the gases to exit out of the furnace. These two special stages occur repeatedly at roughly fixed intervals. When the molten pool level approaches the upper surface of the furnace, the smelting process ends and the power is extinguished.

With the aforesaid analysis, keeping the current within a reasonable range is beneficial for ensuring product quality, reducing energy consumption, and lessening the impact on the grid. The challenge is that the electrode regulation systems of most enterprises are accomplished by human operators relying heavily on their experience, and they can only roughly know the trend of the current instead of accurately understanding how much the electrode position change will affect the current. In this paper, we choose the neural network method to build a model for the three-phase electrode currents. We select the historical value of the three-phase currents, the control signals of the three-phase electrodes, and the working stages as the input of the model, and the three-phase current values for the next period of time are chosen as the output of the model.

3. Electrode regulating model based on GRU-RNN with attention mechanism

In the field of system modeling, neural networks have been widely used due to their ability to learn arbitrary nonlinear mapping. Traditional static feedforward neural networks such as BP neural networks whose outputs are only affected by the current inputs, and does not reflect inner state changes of the system, is inappropriate for our problem. RNN is adept in dealing with sequential data. The hidden layers of RNN are connected so that the system state of last moment can be memorized, which will affect the output together with the input at the current moment. However, if the input sequence is too long (namely the data is too big), RNN without any improvement, which is called vanilla RNN, is incapable of remembering those information far from the current step, resulting in long-term dependency problem. An efficient solution is to replace the units of the hidden layer with gated units such as long short-term memory (LSTM) [21] and GRU, using the gate structure to control the flow of information. GRU is a simplified version of LSTM, which turns out to reduce network parameters while ensuring performance in many researches [22,23], so we choose GRU in our model. Another issue we are concerned with is that boundary conditions change frequently during the smelting process. These factors have a large impact on the current. We hope that the model can identify the information in these key time steps (the crucial points) and put emphasis on them. Inspired by the successful application of attention mechanism in many machine learning fields, we introduce this mechanism to our model.

3.1. Vanilla RNN

A vanilla RNN is described by equations below:

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{1}$$

$$y_t = \sigma(W_{hy}h_t + b_y) \tag{2}$$

where x_t, h_t, y_t denote the input vectors, hidden states and output vectors respectively, and W denotes weight matrixes between layers. For example, W_{xh} means the weight from input vector x to hidden state vector h. Subscript t of each vector represents the time step t. Function σ is the sigmoid function, which is computed by $\sigma(x) = \frac{1}{1+\exp(-x)}$. b is the bias vector. Though RNN is quite suitable for time-series processing, training it through back propagation through time (BPTT) algorithm turns out to be troublesome since the backpropagated gradients will increase or decrease every time step. As the training time grows to some extent, the backpropagated gradients either become infinite (gradients exploding problem) or zero (gradients vanishing problem). A solution to gradient exploding is using gradient clipping [24] while a solution to gradient vanishing is using gated recurrent unit.

3.2. Gated recurrent unit

Gated recurrent unit is calculated by equations below:

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{3}$$

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \tag{4}$$

$$\tilde{h_t} = \tanh(W_{xh}x_t + r_t * W_{hh}h_{t-1} + b_h) \tag{5}$$

$$h_t = z_t * h_{t-1} + (1 - z_t) * \tilde{h_t} \tag{6}$$

$$y_t = \sigma(W_{hy}h_t + b_y) \tag{7}$$

where the meaning of the subscript is consistent with vanilla RNN. z_t and r_t are update gate and reset gate, respectively. $\tilde{h_t}$ is the candidate memory and "*" is the elementwise multiplication. In each time step t, we first calculate the values of r_t and z_t according to the input at the current time step x_t as well as the hidden layer value at the previous time step h_{t-1}. r_t is used to control the degree to which the information of the previous moment is brought into the current state. z_t is used to control the degree of ignoring the memory of the previous moment and the degree of memorizing new information $\tilde{h_t}$. Then x_t, r_t, and h_{t-1} jointly determine the candidate memory $\tilde{h_t}$. Next, the state of the hidden layer is updated by forgetting some previous information and remembering new information. It is worth mentioning that the amount of previous memories to be deleted are equal to that of the new memories to be added, which guarantees that the network's memory is kept within a rational range. In contrast to vanilla RNN, the update scale of GRU is controlled by z_t rather than fixed as in Eq. (1). Finally, the new state of hidden layer is passed through the sigmoid function as the output.

3.3. Attention mechanism

Attention mechanism works through below equations.

$$u_t = \tanh(W_w x_t + b_w) \tag{8}$$

$$\alpha_t = \frac{\exp(u_t v_w)}{\sum_{i=1}^{t} \exp(u_t v_w)} \tag{9}$$

$$s = \sum_{i=1}^{t} (\alpha_t u_t) \tag{10}$$

The input of attention mechanism x_t here is the hidden state of GRU h_t in Eq. (6). Firstly, x_t is converted to a hidden representation, u_t, by a one-layer multilayer perceptron. Then, by comparing the similarity of u_t with a corresponding vector v_w, we measure the significance of a input and calculate a normalized significance

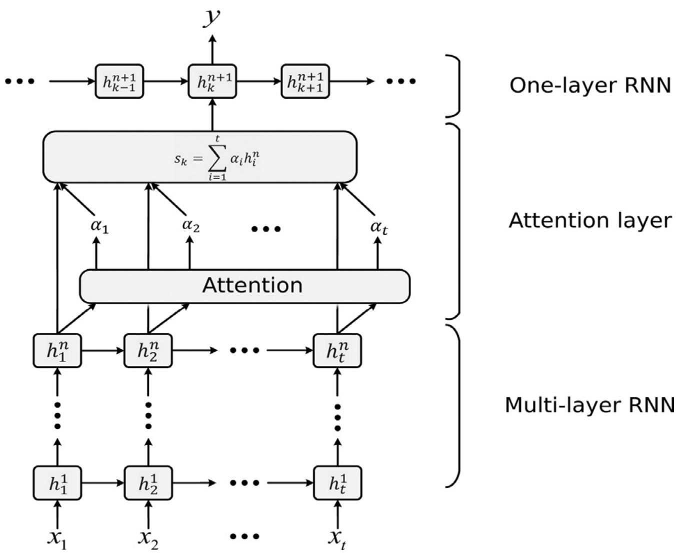


Fig. 2. GRU-RNN model with attention mechanism.

weight α_t through a softmax function. This corresponding vector v_w can be regarded as a high level representation of prominent information, and is randomly initialized and jointly learned during the training process. At last, the weighted sum of the inputs is calculated and could be used as the input of next RNN layer (see the attention layer in Fig. 2).

3.4. Structure of our proposed model

The whole structure of our proposed model is depicted in Fig. 2. We next elaborate on the calculation process of the entire model. The sequential data of each input variable is divided into batches whose length (the number of time steps) is T. Each batch corresponds to one input training sample. The dimension of one batch is [T, I], where I is the number of input variables (8 in this paper). The dimension of the output of muitl-layer RNN for one batch is [T, H], where H is the number of hidden units. Afterwards, attention mechanism calculates the importance of each vector in dimension of T and gives a weighted sum, whose dimension is [1, H]. After that, the one-layer RNN extracts information and gives the required output with dimension of [T, O], where O is the number of output variables (3 in this paper).

4. Simulation preparation

4.1. Data pre-processing

Our proposed model is verified on real manufacturing data acquired from a FMF in Liaoning Province of China. As a typical kind of ESF, FMF is widely used to produce a prominent refractory, fused magnesia. 30,000 sets of consecutive data are produced during a smelting process of FMF and we use them to build our model, 80% as the training samples and 20% as the test samples. The original data consist of eight variables: three-phase currents, stepping motor pulse signals for three-phase electrodes, exhausting stage indicator, and feeding stage indicator. Stepping motors control the lifting of electrodes directly, so the amount of pulses and the position of the electrodes can be approximated as linear relations. When the FMF is in exhausting stage or feeding stage, corresponding variable is set to 1, otherwise it is set to 0. Since each variable differs greatly in their values, we normalize them separately through equation below:

$$x = \frac{x' - x'_{min}}{x'_{max} - x'_{min}} \tag{11}$$

x is the normalized input data. x' is the original input data. x'_{min} and x'_{max} are the minimum and maximum values of each input data. Similarly, the output data is normalized by the same equation.

4.2. Model establishment

In line with the depiction in Fig. 2, we build a electrode regulating system model where the first RNN is a multi-layer recurrent neural network and the second RNN is a single-layer recurrent neural network. We next explain the network parameter selection. It is generally believed that stacking several layers of recurrent units contributes to extracting deeper features of the inputs; nevertheless, excessive number of layers will slow down the efficiency of training process and will not bring an increase in accuracy. Through our experiment, setting the number of hidden layers in first RNN to 3 is the most appropriate. However, we find out that a single-layer recurrent neural network is sufficient for the second RNN because the first RNN along with the attention mechanism have accomplished information compression and extraction outstandingly.

As is mentioned above, BPTT algorithm is applied for tuning the network parameters. We take the parameters W_{hy} and b_y in Eq. (7) as an example to illustrate how the parameters are tuned. The cross-entropy loss function L applied in this paper is:

$$L = -\frac{1}{n}\sum_x [\hat{y}\ln(y) + (1-\hat{y})\ln(1-y)] \tag{12}$$

where $\hat{y}$ is the real value and y is the predicted value. Compared with the mean square error loss function, the cross-entropy loss function can help neural networks learn faster from their errors. We can deduce the partial derivative of W_{hy} and b_y:

$$\begin{aligned}\frac{\partial L}{\partial W_{hy}} &= \frac{y_t - \hat{y}_t}{y_t(1-y_t)}\sigma'(y_t)h_t = (y_t - \hat{y}_t)h_t \\ \frac{\partial L}{\partial b_y} &= \frac{y_t - \hat{y}_t}{y_t(1-y_t)}\sigma'(y_t) = y_t - \hat{y}_t\end{aligned} \tag{13}$$

In Eq. (13), the backpropagated error is related to the difference between the predicted value and the actual value. However, besides this difference, the backpropagated error under mean square error loss function is also related to the derivative of the activation (sigmoid in this paper) function [25]. When the sigmoid function is close to 0 or 1, the derivative is close to 0 accordingly, which will cause the propagation error to become very small. Cross-entropy loss function avoids this propagation-error-slow-down problem. Therefore, we choose cross-entropy loss function in this paper. Subsequently, W_{hy} and b_y are tuned through below equation:

$$\begin{aligned}W_{hy} &= W_{hy} - \eta\frac{\partial L}{\partial W_{hy}} \\ b_y &= b_y - \eta\frac{\partial L}{\partial b_y}\end{aligned} \tag{14}$$

Other network parameters are tuned in the same way according to the chain rule.

Learning rate η is a crucial parameter to regulate weights in training process. When this parameter is too large, the loss function will fluctuate; when the value is too small, the convergence process will become slow. To avoid this point, adam optimizer [26] is adopted since it can automatically adjust learning rate in the right way given an initial value. Meanwhile, we introduce dropout network [27] so as to alleviate overfitting problem, and

Please cite this article as: S. Ding, Z. Wang and W. Kong et al., Electrode regulating system modeling in electrical smelting furnace using recurrent neural network with attention mechanism, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.05.060

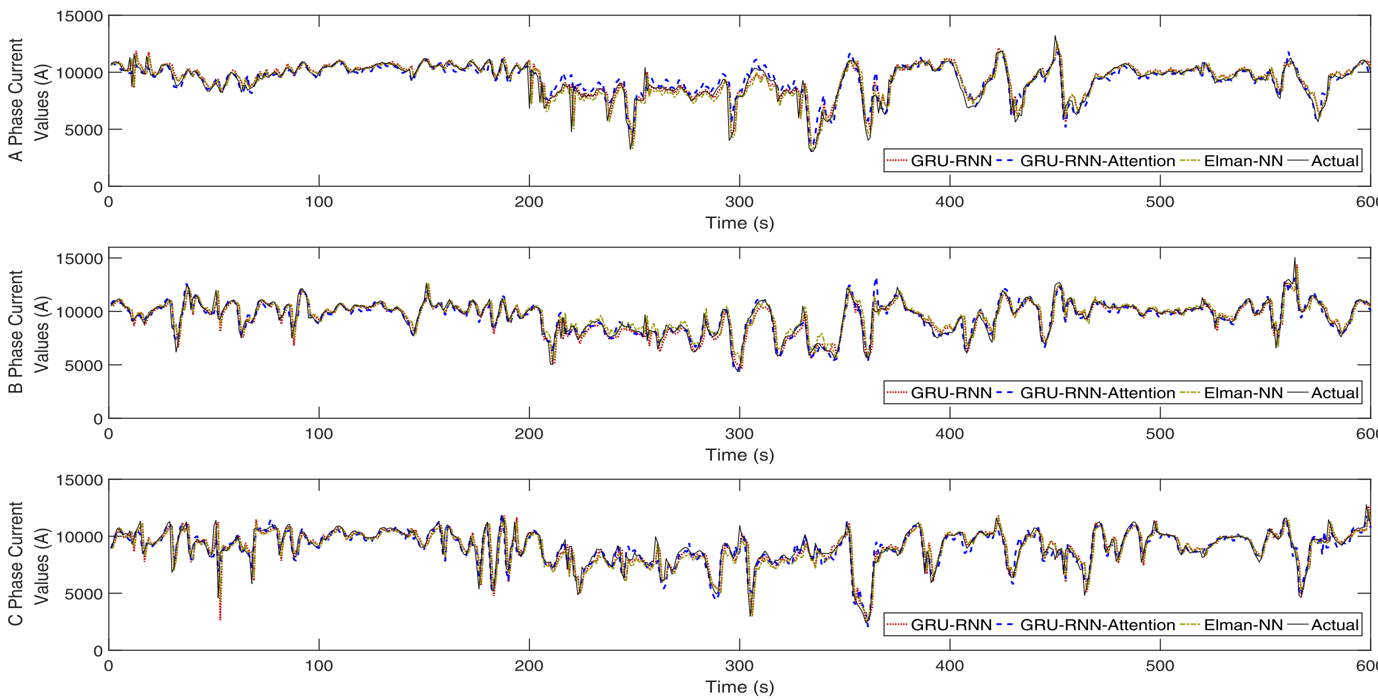


Fig. 3. Prediction results of three-phase currents.

Table 1
Network parameters.

Network parameters	Sampling interval 1 s	5 s	10 s
Epoches	10	10	10
Initial learning rate	0.001	0.001	0.0008
Hidden layer units	23	25	25
Keep probability	0.8	0.8	0.8
Time steps	10	10	10

the keep probability is set to 0.8. The training process terminates when the loss function no longer drops or the maximum of iterations is reached. The selection of other network structure parameters listed in Table 1 are similar to [10].

5. Experimental results analysis

We establish an electrode regulating system model using a 3-hidden-layer of GRU-RNN in Section 5.1 and discover that GRU-RNN works well in short-term prediction of the smelting current (this paper refers to the prediction 10s ahead of the current time step as short-term prediction). However, such short-term prediction is senseless in the relatively long smelting process. In order to increase the prediction time (the prediction time is $\tau \times T$, where τ is the sampling interval and T is the number of time steps that each output sample covers) to meet the demand of efficient prediction (larger prediction time implies more efficient the smelting current prediction), we can increase the sampling interval τ while keep T unchanged. Yet increasing the sampling interval causes the corresponding data points in the sequence sampled from the real crucial points to become sparse. Accordingly, these sparse data points bring difficulties for GRU-RNN model to fully identify the real crucial points, resulting in a sharp decrease in prediction accuracy (see Table 3 for details, which will be discussed in Section 5.2). Attention mechanism is thus introduced in Section 5.2 to help GRU-RNN model to catch the crucial points when the sampling interval increases (meaning that the prediction time increases). With the aid of attention mechanism, the MAPE (Mean Absolute Percentage Error) with a relative long prediction time (50 s) of our proposed model is 6.53% which meets the industrial requirements. In Section 5.3, we further increase the prediction time (100 s) and notice that the MAPE of our proposed model maintains around 7.75%, satisfying the demand of industrial application as well.

5.1. Performance of GRU-RNN in dealing with industrial big data

Fig. 3 shows the outputs of GRU-RNN, GRU-RNN-Attention and Elman neural network [9] under the test samples along with the actual current values. The test samples contains 6000 sets of consecutive data with a sampling interval of 1 s and is divided into 600 batches (one batch contains 10 sets of data). Each time one batch is fed into the model and the model outputs a batch of ten sets of data, that is, predicts the three-phase current values of next 10 s. 600 batches of data are sequentially input into the model to obtain 600 batches of outputs. Since the whole output sequence is too long to plot, we only draw the first 60 batches of outputs (equivalent to 600 s) in Fig. 3. Intuitively, all three models have good prediction results. The error probability distributions (EPD) of the three models are shown in Fig. 4. It is widely believed that the EPD of a good prediction model should be close to the normal distribution with small variation. We can obviously observe that the distribution of prediction error of Elman neural network is a little more scattered while the EPD of GRU-RNN and GRU-RNN-Attention model are clustered more tightly. Additionally, we calculate RMSE (Root Mean Square Error) and MAPE as two error indicators to further assess the reliability of our proposed model, where RMSE approximately reflects the absolute error and MAPE reflects the relative error. From Table 2 we can see more instinctively that both GRU-RNN and GRU-RNN-Attention outperform Elman neural network since the absolute errors and the relative errors of both GRU-RNN and GRU-RNN-Attention models are lower. During the smelting process in FMF, if the difference between the predicted value and the actual value is within a certain range ($\pm 10\%$), the predicted value is considered to be acceptable. Hence, we define

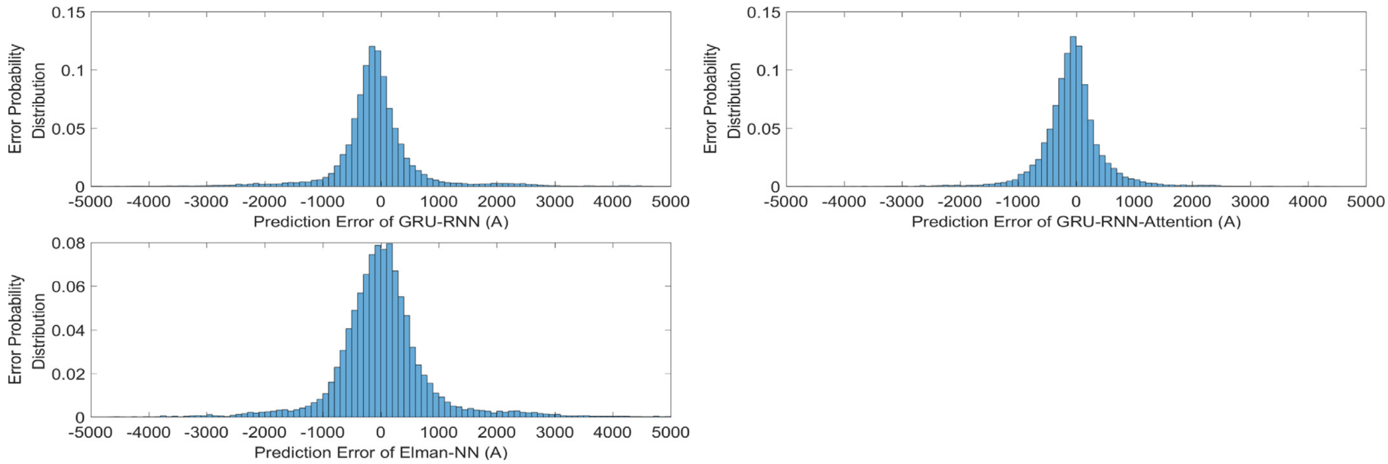


Fig. 4. Error probability distribution of GRU-RNN, GRU-RNN-Attention, and Elman-NN.

Table 2
Error indicators.

	RMSE	MAPE	Accuracy
Elman-NN	838.7	0.1337	0.8298
GRU-RNN	753.3	0.0687	0.8687
GRU-RNN-Attention	532.4	0.0662	0.9079

another indicator, the accuracy of a model, in equation below:

$$A_i = \begin{cases} 1, & if \quad \frac{|X_{pred(i)} - X_{real(i)}|}{X_{real(i)}} \le 0.1 \\ 0, & otherwise \end{cases}$$

$$accuracy = \frac{\sum_{i=1}^{n} A_i}{n} \times 100\% \tag{15}$$

where n is the length of output sequence. The amount of those prediction results within the range of error is counted to calculate accuracy. Table 2 shows the accuracy of GRU-RNN and GRU-RNN-Attention model are higher than Elman neural network. In conclusion, GRU-RNN and GRU-RNN-Attention perform better than Elman neural network, benefiting from the capability of GRU in dealing with sequential data.

It is worth noting that the sampling interval here is 1 s and the number of time steps T is 10, meaning the prediction time is 10 s, and such short-term prediction makes little sense in real practise. When we increase the prediction time by increasing τ to 5 s to get an efficient prediction, GRU-RNN acts poor in forecasting the crucial points (see the data point in Fig. 5 at the 1100th time step), causing a low prediction accuracy failing to meet industrial requirements.

5.2. Performance of attention mechanism in capturing crucial points

When we increase the sampling interval to 5 s (meaning the prediction time increases to 50 s and the number of dataset drops to 6000), the performance of the two models diverges. GRU-RNN model failed to make accurate predictions in some moments and tend to output average values instead of detailed information (see the B phase prediction results in Fig. 5). The outputs of GRU-RNN-Attention model remain consistent with the actual value in general. We can perceive that around the 1100th time step in Fig. 5, where the current value drops drastically. According to the dataset, the FMF is under feeding stage during that period which exerts a great impact on the current value. GRU-RNN-Attention model catches this feeding stage apparently better than GRU-RNN model. Fig. 6 indicates that the EPD of GRU-RNN model is more dispersed and irregular than that of GRU-RNN-Attention model. In Table 3,

Table 3
Error indicators.

	RMSE	MAPE	Accuracy
GRU-RNN	1319.8	0.2447	0.7167
GRU-RNN-Attention	690.3	0.0653	0.8817

Table 4
Error indicators.

	RMSE	MAPE	Accuracy
GRU-RNN	1238.0	0.2128	0.7122
GRU-RNN-Attention	704.5	0.0775	0.8889

the RMSE of GRU-RNN model is nearly twice as much as that of GRU-RNN-Attention model and the MAPE of GRU-RNN model is nearly four times as much as that of GRU-RNN-Attention model. Furthermore, the accuracy of GRU-RNN model drops to 71.67%, where the accuracy of GRU-RNN-Attention model remains steadily at 88.17%. These results indicate that attention mechanism is vital to improve the accuracy of our proposed model in efficient prediction.

5.3. Verification of the capability of attention mechanism in capturing crucial points

To further validate the capability of attention mechanism in capturing crucial points, we do the same experiments as in Section 5.2 except that the sampling interval is expanded to 10 s (the prediction time here is thereby 100 s). Accordingly, the crucial points representing key information (such as working stages alternation, granularity or impurity composition variation of raw materials) becomes more sparse, which presents less available information for the model to utilize. In Fig. 7, the outputs of GRU-RNN model deviate the actual ones in many time periods, implying a poor prediction of the model. For example, at the moment when the current value drops greatly, the predicted value fails to keep up with this change (see the 550th time step Fig. 7). On the contrast, the prediction results of GRU-RNN-Attention model accord with the actual value well (see Fig. 7). Furthermore, the EPD of GRU-RNN-Attention model is still closed to a normal distribution with small variation, but that of GRU-RNN model is not (see Fig. 8). As for the error indicators, Table 4 shows that the relative error and absolute error of the GRU-RNN model are higher than the GRU-RNN-Attention model, and the accuracy is lower than the

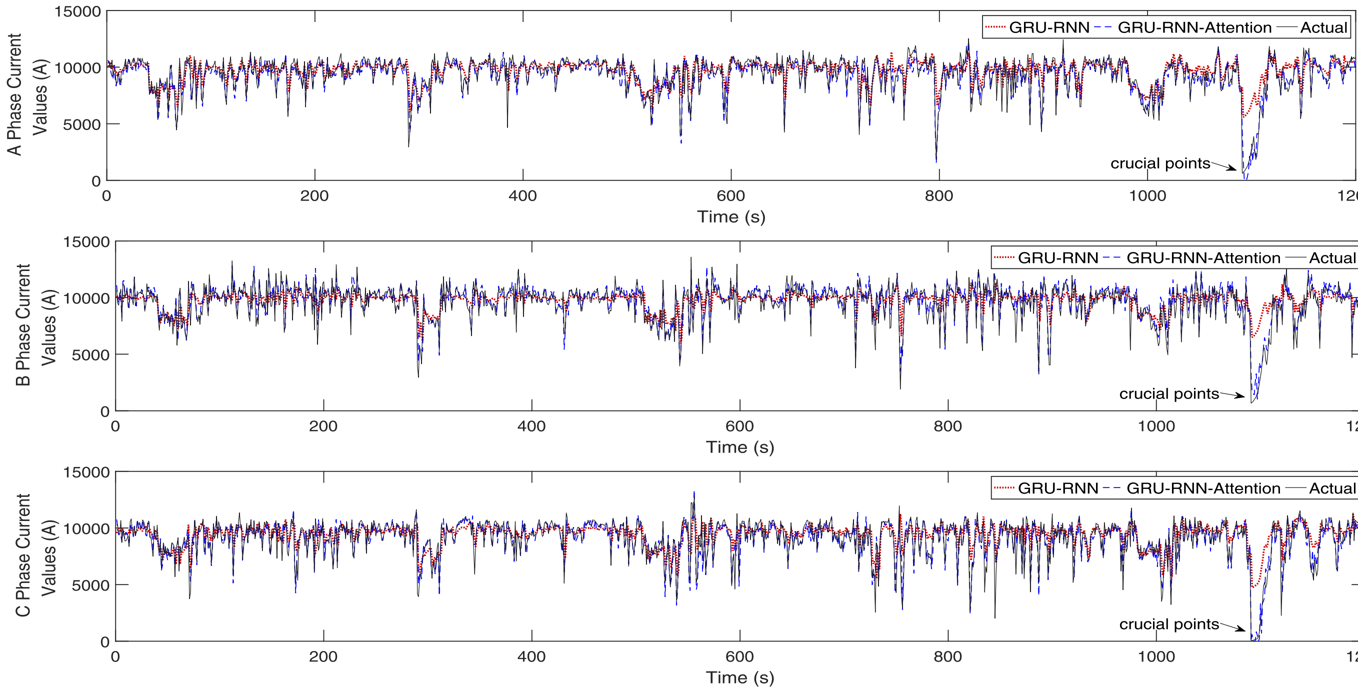


Fig. 5. Prediction results of three-phase currents.

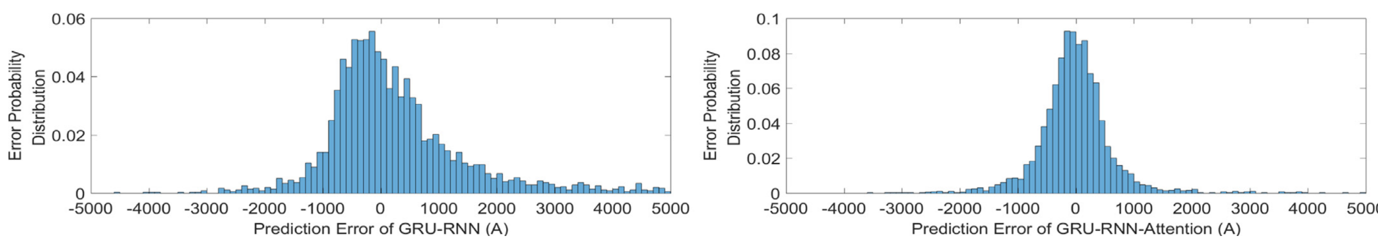


Fig. 6. Error probability distribution of GRU-RNN and GRU-RNN-Attention.

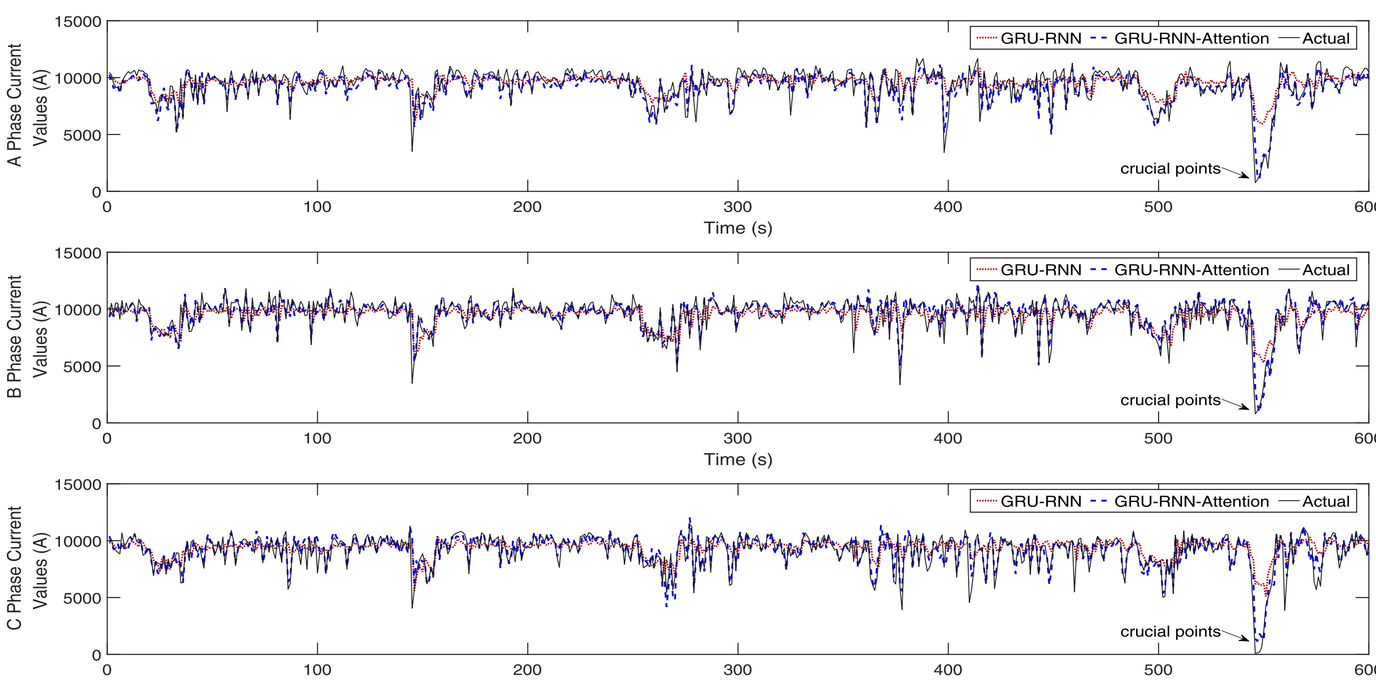


Fig. 7. Prediction results of three-phase currents.

Please cite this article as: S. Ding, Z. Wang and W. Kong et al., Electrode regulating system modeling in electrical smelting furnace using recurrent neural network with attention mechanism, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.05.060

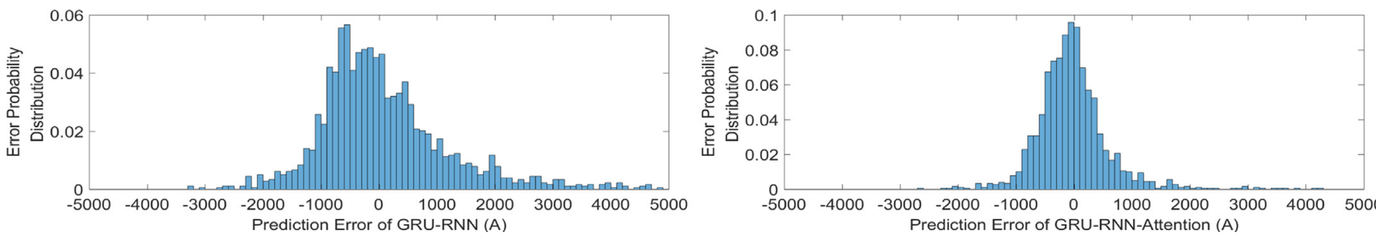


Fig. 8. Error probability distribution of GRU-RNN and GRU-RNN-Attention.

latter. These results indicate that GRU-RNN-Attention model performs better than GRU-RNN model for efficient prediction.

6. Conclusion

In this paper, we establish a recurrent neural network using gated recurrent unit with attention mechanism for the modeling of electrode regulation system in fused magnesium furnace. The main contributions of this paper can be summarized as following three points. Firstly, in order to fully reflect the dynamic characteristics of the smelting process, RNN is chosen; for the purpose of alleviating the long-term dependency problem in big data, GRU is adopted. Secondly, the attention mechanism is incorporated into the RNN framework, which captures the crucial points during the smelting process efficiently. Specifically, the experimental results show that when the sampling interval is short, i.e., the prediction time is short (the prediction time is $\tau \times T$, where τ is the sampling interval and T is the number of time steps that each output sample covers), both GRU-RNN and GRU-RNN-Attention are able to acquire enough vital information to make good predictions, benefiting from the capability of GRU in dealing with sequential data; when the sampling interval becomes longer, GRU-RNN-Attention can catch the crucial points more efficiently than GRU-RNN model, resulting in better prediction of the three-phase currents. Thirdly, the proposed GRU-RNN-Attention model reaches an accuracy of about 89% with a prediction time of 100 s, which meets the industrial requirements. In the future, data preprocessing procedure [28,29] could be introduced to decompose the input data or to change its segmentation way so that RNN can extract features more effectively. At the same time, we will delve more deeply into the attention mechanism to enhance its ability to identify key information. Furthermore, some other state-of-the-art deep learning methods [30–34] are worthy of trying.

Conflict of interest

None.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants no. 61603088) and National Key R&D Program of China (no. 2016YFB0601404).

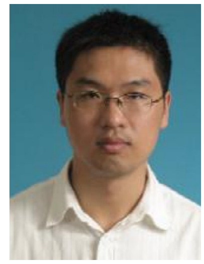
References

[1] J. Yang, T. Chai, Data-driven demand forecasting method for fused magnesium furnaces, in: Proceedings of Conference on Intelligent Control and Automation, 2016, pp. 2015–2022.
[2] J. Yang, L.Y. Wang, X. Wang, Z.W. Wu, Very short-term demand forecasting for fused magnesium furnaces based on specific penalty function, in: Proceedings of 2014 11th World Congress on Intelligent Control and Automation, WCICA, IEEE, 2014, pp. 788–793.
[3] Z.W. Wu, Y.J. Wu, T.Y. Chai, Intelligent control of fused magnesium furnaces based on SPSA, J. Shanghai Jiaotong Univ. 45 (8) (2011) 1095–1100.
[4] Z.W. Wu, T.Y. Chai, J. Fu, J. Sun, Hybrid intelligent optimal control of fused magnesium furnaces, in: Proceedings of Conference on Decision and Control, 2010, pp. 3313–3318.
[5] W.L. Ma, S.X. Zhu, Intelligent Control Algorithm of Electric-Fused Magnesia Furnace Based on Neural Network, Springer, New York, 2014.
[6] Z.W. Wu, T.F. Liu, Z.P. Jiang, T.Y. Chai, L.N. Zhang, Nonlinear control tools for fused magnesium furnaces: design and implementation, IEEE Trans. Ind. Electron. 65 (9) (2018) 7248–7257.
[7] Z.W. Wu, Y.J. Wu, T.Y. Chai, J. Sun, Data-driven abnormal condition identification and self-healing control system for fused magnesium furnace, IEEE Trans. Ind. Electron. 62 (3) (2015) 1703–1715.
[8] Y. Wang, Z.Z. Mao, H.X. Tian, Y. Li, P. Yuan, Modeling of electrode system for three-phase electric arc furnace, J. Central South Univ. Technol. 17 (3) (2010) 560–565.
[9] Y.J. Wu, J. Zhang, L. Zhang, H. Yue, T.Y. Chai, Hybrid modeling for electrode regulating system of electrical smelting furnace for magnesia, Chin. J. Sci. Instrum. 30 (6) (2009) 13–19.
[10] G. Song, Z. Wang, F. Han, S. Ding, M.A. Iqbal, Music auto-tagging using deep recurrent neural networks, Neurocomputing 292 (2018) 104–110.
[11] S. Nakayama, S. Arai, Dnn-lstm-crf model for automatic audio chord recognition, in: Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence, ACM, 2018, pp. 82–88.
[12] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the Properties of Neural Machine Translation: Encoder-decoder Approaches, Empirical methods in natural language processing, 2014, pp. 103–111.
[13] Z. Tan, J. Su, B. Wang, Y. Chen, X. Shi, Lattice-to-sequence attentional neural machine translation models, Neurocomputing 284 (2018) 138–147.
[14] Z. Wang, J. Wang, Y. Wu, State estimation for recurrent neural networks with unknown delays: a robust analysis approach, Neurocomputing 227 (2016).
[15] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Netw. 5 (2) (1994) 157–166.
[16] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning Phrase Representations using RNN Encoder-decoder for Statistical Machine Translation, Empirical methods in natural language processing, 2014, pp. 1724–1734.
[17] D. Bahdanau, K. Cho, Y. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, International conference on learning representations, 2015.
[18] M.T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, Empirical methods in natural language processing (2015) 1412–1421.
[19] L. Chao, J. Tao, M. Yang, Y. Li, Z. Wen, Audio Visual Emotion Recognition with Temporal Alignment and Perception Attention, 2016, arXiv:1603.08321.
[20] C. Yu, K.S. Barsim, Q. Kong, B. Yang, in: Multi-level Attention Model for Weakly Supervised Audio Classification, 2018.
[21] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[22] B. Athiwaratkun, J.W. Stokes, Malware classification with LSTM and GRU language models and a character-level CNN, in: Proceedings of IEEE International Conference on Acoustics, 2017.
[23] F. Rui, Z. Zuo, L. Li, Using LSTM and GRU neural network methods for traffic flow prediction, in: Proceedings of Youth Academic Conference of Chinese Association of Automation, 2017.
[24] R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, in: Proceedings of International Conference on Machine Learning, 2013, pp. 1310–1318.
[25] M. A.Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.
[26] D. Kingma, J. Ba, Adam: a method for stochastic optimization, International conference on learning representations (2015).
[27] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving Neural Networks by Preventing Co-adaptation of Feature Detectors, 2012, arXiv:1207.0580.
[28] N. Ur Rehman, D.P. Mandic, Filter bank property of multivariate empirical mode decomposition, IEEE Trans. Signal Process. 59 (5) (2011) 2421–2426.
[29] B. Esmael, A. Arnaout, R.K. Fruhwirth, G. Thonhauser, Multivariate time series classification by combining trend-based and value-based approximations, in: Proceedings of International Conference on Computational Science and its Applications, Springer, 2012, pp. 392–403.
[30] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643–649.

[31] S. Bai, J.Z. Kolter, V. Koltun, An Empirical Evaluation of Generic convolutional and Recurrent Networks for Sequence Modeling, 2018, arXiv:1803.01271.
[32] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, T. Young, Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach, IEEE Trans. Med. Imaging 33 (5) (2014) 1129–1136.
[33] Z. Wu, C. Shen, A. Van Den Hengel, Wider or deeper: revisiting the resnet model for visual recognition, Pattern Recognit. 90 (2019) 119–133.
[34] N. Zeng, Z. Wang, Y. Li, M. Du, X. Liu, A hybrid EKF and switching PSO algorithm for joint state and parameter estimation of lateral flow immunoassay models, IEEE/ACM Trans. Comput. Biol. Bioinform. 9 (2) (2012) 321–329.

Shenyi Ding received the B.S. degree from College of Information Science and Technology, Donghua University in 2014. Since September 2015, he takes a successive postgraduate and doctoral program in information science and technology at Donghua University, China. His research interests include deep learning and dynamic system modeling and control.

Zhijie Wang received his B.E., M.E. and Ph.D. degrees from College of Information Science and Technology, Donghua University, Shanghai, China, in 1991, 1994 and 1997, respectively. He did his Post Doc from the University of Tokyo, Tokyo, Japan in 2002. He is currently a Professor with the College of Information Science and Technology, Donghua University, Shanghai, China. His main research interests are computational neuroscience, neural network, and deep learning.



Weijian Kong received his M.E. and Ph.D. degrees from College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2008 and 2014, respectively. He is currently a lecturer with the College of Information Science and Technology, Donghua University, Shanghai, China. His current research interest focuses on plant-wide optimization and control for the complex industrial systems.

Honghai Yang received her B.E. and M.E degrees from College of Energy and Power Engineering, Xi'an Jiaotong University, Shanxi Province, China, in 1989, 1992, respectively, received Ph.D. degree from College of Environmental Science and Engineering, Donghua University, Shanghai, China, in 2006. She studied at University of Stuttgart, Germany since 2003.6 to 2005.8. She is currently an associate professor with the College of Environmental Science and Engineering, Donghua University, Shanghai, China. Her main research interests are energy conversion and management, heat and mass transfer, energy saving and new technology in HVAC system.

Guangxiao Song received the M.S. degree in Computer Science and Technology from Yangtze University, China, in 2016. Since September 2016, he has been a Ph.D. candidate in information science and technology at Donghua University, China. His research interests include deep learning and music information retrieval.