Research Article

The Bidirectional Information Fusion Using an Improved LSTM Model

Tianwei Zheng , Mei Wang , Yuan Guo, and Zheng Wang

Xian University of Science and Technology, Xian 710054, China

Correspondence should be addressed to Mei Wang; wangm@xust.edu.cn

Received 15 January 2021; Revised 21 February 2021; Accepted 12 April 2021; Published 21 April 2021

Academic Editor: Hsu-Yang Kung

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

The information fusion technology is of great significance in intelligent systems. At present, the modern coal-fired power plant has the fully functional sensor network. However, many data that are important for the operation of a power plant, such as the coal quality, cannot be directly obtained. Therefore, the information fusion technology needs to be introduced to obtain the implied information of the power plant. As a practical application, the soft measurement of coal quality is taken as the research object. This paper proposes an improved LSTM model combined with the bidirectional deep fusion, alertness mechanism, and parameter self-learning (DFAS-LSTM) to realize online soft computing for the coal quality analyses of industries and elements. First, a latent structure model is established to preprocess the noisy and redundant sensor network data. Second, an alertness mechanism is proposed and the self-learning method of the activation function parameters is used for the data feature extraction. Third, a deeply bidirectional fusion layer is added to the long short-term memory neural network model to solve the problem of the insufficient accuracy and the weak generalization. Using the historical data of the sensor network, the DFAS-LSTM model is established. Then, the online data of the sensor network is input to the DFAS-LSTM model to implement the online coal quality analyses. Experiment shows that the accuracy of the coal quality analyses is increased by 1%–2.42% compared to the traditionally bidirectional LSTM.

1. Introduction

Long short-term memory (LSTM) neural network is a variant model of the recurrent neural network [1, 2]. In recent years, there have been many studies on the LSTM. Reference [3] proposed a novel DC-Bi-LSTM model. Reference [4] developed a selective multimode long short-term memory network. Reference [5] studied a novel layered network to solve the problem of the pedestrian trajectory prediction. Besides, aiming at the problem of the 3D motion recognition, reference [6] introduced a new gating mechanism into LSTM to increase the reliability of the sequential input data and to adjust the effect on updating the long-term context information stored in the memory cell.

For the past few years, neural network models based on the bidirectional long short-term memory (Bi-LSTM) networks have achieved excellent application performance in their respective fields and have shown a vitality of the

Bi-LSTM in the field of the sequential data processing [7, 8]. Therefore, the basic research of the bidirectional long short-term memory network is of a great significance for the performance upgrading of the neural network model based on it as well as the development of a new neural network model in specific application field.

In recent years, neural network technology has shown excellent performance in pattern recognition, automatic control, signal processing, auxiliary decision-making, and other fields [9, 10]. In particular, the coal is an important energy in the world, and the way of the coal energy utilization still has a great improvement space. Meanwhile, neural network technology has the practical value for adjusting the energy utilization way. According to the data provided by BP p.l.c. in 2019, world's total primary energy consumption is 583.88 EJ, of which coal consumption is 157.86 EJ. The coal plays an important role in the consumption of the primary energy [11]. The main utilization

mode of the coal is combustion in coal-fired power plants [12]. The types of the coal used in coal-fired power plants are complex and changeable. In coal-fired power plants, the boiler combustion system is the heart. The change of the coal quality directly affects the combustion state of the boiler, as well as the stability, safety, and economy of the boiler combustion system. The timely acquisition of the coal quality information is of a great significance for ensuring the smooth operation and the adequacy of the fuel combustion in the boiler.

At the same time, it can also provide power for improving the economic benefits of the coal-fired power plant and the energy utilization rate of the coal-fired industry. So far, the coal quality measurement technologies include neutron activation analysis, laser-induced breakdown spectroscopy, and near-infrared spectroscopy [13, 14]. The above technologies realize the real-time measurement of the coal quality in coal-fired power plants, but all rely on expensive hardware facilities.

For the application, an improved LSTM model with the bidirectional fusion, alertness mechanism, and parameter self-learning (DFAS-LSTM) is used for coal quality computing in coal-fired power plants. The real-time coal quality computing method is not yet found in publications, and the soft computing of the coal quality in coal-fired power plant is a major challenge for many years. However, coal-fired power plants generally have a certain degree of intelligence. They have a strong sensor network composed of edge devices and central processing systems. In the sensor network, there are a large number of sensors with complete categories, which provide a wealth of the edge data for the system. These edge data contain rich information about coal-fired power plants systems. Some of the information, such as coal quality information, is hidden in the edge data in the form of the high-dimensional data features. The way to realize element analyses of the coal quality through edge data mining can avoid the interference to the system operation as well as the cost increase caused by the addition of the hardware facilities. It is cheap and convenient as well. To achieve this goal, this paper proposes a DFAS-LSTM model, which is used to mine coal quality-related information in the edge data of the coal-fired power plants. At the same time, this paper uses the alertness mechanism for the alerting abnormal data information, the improved activation function, and the deeply bidirectional fusion structure of the LSTM for improving the performance of the model. The logic diagram of the soft computing system is shown in Figure 1.

The following parts of this paper include related work, data set and its latent structure, establishment of the alertness mechanism, deeply bidirectional fusion LSTM modeling, experiment and result analyses, and conclusions.

2. Related Work

This section discusses the development trend of the efficient use of energy in the world and the related work of scholars on the optimization of the LSTM structure and the coal utilization from various countries.

At present, the world is in a trend of economic transformation and efficient use of energy. The sustainable development trend of the economy puts forward higher requirements for the rational and efficient use of the primary energy. According to the statistics of the International Energy Agency (IEA), the energy intensities of some countries are shown in Figure 2.

Among primary energy sources, coal is mainly used for power generation in coal-fired power plants. The coal-fired power plants are facing increasingly severe challenges. The introduction of the intelligent technologies and methods is necessary for them to adapt to the historical trend of the efficient energy utilization.

In recent years, the generation capacity of wind and solar power has continued to grow. However, due to the high volatility of the renewable energy from wind and solar power generation, coal-fired power plants must be used to cover the gap between renewable energy generation and load to maintain the stability of the frequency and the power supply stability [15]. Therefore, the method of the coal-fired power generation is still irreplaceable.

According to the survey data from BP p.l.c., driven by the wind and solar energy, the growth of the renewable energy has reached a record level, accounting for more than 40% of the primary energy growth in 2019 [16]. This also means that the ever-changing energy structure places higher demands on coal-fired power generation technology.

2.1. Optimization of the LSTM Model. Aiming at increasing the depth in the time dimension, reference [17] extended a highway LSTM (HW-LSTM) model by adding highway networks inside an LSTM and used it for language modeling. As an application of the recursive neural network, reference [18] studied the capture of the behavior trajectory through a large context window and achieved the purpose of solving the data sparseness and improving the robustness. Combined with the attention mechanism and the character-level convolutional neural network, reference [19] proposed several new classification architectures based on the long short-term memory (LSTM) language model and the gated recurrent unit (GRU) language model. Reference [20] formulated precipitation nowcasting as a spatiotemporal sequence forecasting problem. By extending the fully connected LSTM (FC-LSTM) to have convolutional structures in both the input-to-state and state-to-state transitions, they proposed the convolutional LSTM (ConvLSTM) and used it for the precipitation nowcasting problem. Reference [21] presented a novel unified framework (LSTM-E) for exploring the learning of the LSTM and visual-semantic embedding. In order to solve the problem of the energy load forecasting, reference [22] presented a novel forecasting model based on long short-term memory algorithms. Reference [23] designed an architecture with the purpose of serving as a model which can generate sequence samples, while simultaneously classifying a given sequence.

2.2. Optimization of the Coal Energy Utilization. At present, the process of the power plant intelligence is constantly

FIGURE 1: The logic diagram of the soft computing system for the industrial and the elemental coal quality analyses.

FIGURE 2: Energy intensities of some countries.

advancing. Reference [24] developed a dynamic model of the drum-boiler using NARX neural networks, which can forecast the actual pressure and water level of the drum-boiler. Reference [25] proposed a pilot program aiming at developing a comprehensive knowledge base for power plants by using principal component analysis and artificial neural networks. The program used the principal component analysis method to filter noise in the prediagnosis stage and evaluated the neural network model based on representative data of the power plant. Reference [26] developed a diagnosis system of the power plant gas turbine to detect the deterioration of the turbine. By using artificial neural network, the system can be used for predicting the deterioration of the main component. Reference [27] proposed a new method for predicting the output of a power plant by using a feedforward neural network. The network used ambient temperature, atmospheric pressure, relative humidity, and vacuum as the input parameters to predict the average hourly output of the power plant.

At present, coal quality information of a coal-fired power plant cannot be obtained in real time. As a result, untimely coal quality information lacks guiding value for the optimization of coal combustion, which leads to the underutilization of the coal in coal-fired power plants. As a method to improve the utilization of the coal, an online soft measurement model

for coal quality is proposed in this paper. The system framework is shown in Figure 3.

3. Data Set and Its Latent Structure

This section discusses the data acquisition and preprocessing including the selection of conventional measurement points, principal component analysis, and independent component analysis.

3.1. Selection of Conventional Measurement Points. In the sensor network of a coal-fired power plant, there are many kinds of conventional measurement points, and the redundancy among the measured data is serious. Most of the conventional measurement points have no obvious correlation with coal quality. These points lack the guidance for the coal quality soft computing. Therefore, it is necessary to screen the conventional measurement points of the coal-fired power plant and obtain the effective data from the monitoring points.

The operation process of a coal-fired power plant includes complex physical and chemical reaction processes. The existing operation mechanism research of the coal-fired power plant provides a basis for the selection of conventional measurement points [28]. According to the laws of the energy conservation, material conservation, and the actual process flow in the operation of a coal-fired power plant, 190 conventional measurement points related to the coal quality are determined from the sensor network of the coal-fired power plant. Some relevant measurement points used in soft computing of the coal quality are shown in Table 1.

In coal-fired power plants, the sensors used in conventional measurement points have various working principles and complex working environment; these result in the serious data redundancy and noise. At the same time, the range and accuracy of the data from different measuring points are significantly different, as well as the correlation with coal quality. In order to ensure the performance of the model, it is necessary to preprocess the data of conventional measurement points in coal-fired power plants.

For the monitoring data coming from conventional measurement points in coal-fired power plants, the bad points in the data are eliminated first. The data preprocessing model realizes this function mainly through the following steps:

FIGURE 3: The system framework.

TABLE 1: Some relevant measurement points used in soft computing of the coal quality.

No.	Name/unit
1	Active power of the generator (MW)
2	Total power of the generator (MW)
3	Total coal supply (T/h)
4	Current of the coal feeder (A)
5	Current of the coal mill (A)
6	Primary air volume at mill inlet (T/h)
7	Primary air temperature at mill inlet (°C)
8	Steam temperature at the outlet of the final reheater (left side) (°C)
9	Steam temperature at the outlet of the final superheater (left side) (°C)
10	Flow of the feed water (T/h)
11	Temperature of the feed water (°C)
12	Pressure of the feed water (MPa)
13	Drum pressure (MPa)
14	Drum water level (mm)
15	Secondary air flow rate (T/h)
16	Desuperheating water temperature of the superheater (°C)
17	Outer wall temperature of the high-pressure cylinder (°C)
18	Inner wall temperature of the high-pressure cylinder (°C)
19	Inlet air temperature of the secondary heater (°C)
20	Primary air pressure at outlet of the air preheater (MPa)
21	Secondary air pressure at the inlet of the air preheater (MPa)
22	Outlet air pressure of the forced draft fan (MPa)
23	Current of the supply fan (A)
24	Current of the primary fan (A)
25	Outlet air pressure of the primary fan (MPa)
26	Differential pressure between the secondary air box and the furnace (MPa)
27	Inlet air temperature of the primary fan (°C)
28	Outlet air temperature of the primary air heater (°C)
29	Inlet air temperature of the primary air heater (°C)
30	Negative pressure at the outlet of the extension end low-temperature superheater (MPa)
31	Negative pressure at the outlet flue of the fixed end superheater (MPa)

TABLE 1: Continued.

No.	Name/unit
32	Temperature of the primary air (°C)
33	Temperature of the hot secondary air (°C)
34	Primary air duct pressure of the furnace (MPa)
35	Main pipe pressure of the hot primary air (MPa)
36	Outlet pipe temperature of the reheater (left side) (°C)
37	Outlet pressure of the reheater (left side) (MPa)
38	Working environment temperature of the coal mill (°C)
39	Inlet temperature of the reheater desuperheater (left side) (°C)
40	Desuperheating water flow of the reheater (left side) (T/h)
41	Primary air pressure at mill inlet (MPa)
42	Pressure of the reheater inlet (MPa)
43	Inlet steam temperature of the low-pressure cylinder (°C)
44	Exhaust temperature of the high-pressure cylinder (°C)
45	Exhaust pressure of the high-pressure cylinder (MPa)
46	Outlet air temperature of the secondary heater (°C)
47	Steam pipe pressure of the high and intermediate pressure cylinder (MPa)
48	Steam pipe temperature of the high and medium pressure cylinder (°C)
49	Differential pressure between the secondary air box and the furnace (MPa)
50	Oxygen concentration at chimney inlet (%)
51	Outlet flue gas temperature of the air preheater (°C)
52	Outlet flue gas pressure of the air preheater (kPa)
53	Flue gas temperature at the inlet of the air preheater (°C)
54	Inlet flue gas pressure of the air preheater (kPa)
55	Main motor current of the air preheater (A)
56	The current of the induced draft fan (A)
57	Flue gas pressure at the inlet of the induced draft fan (kPa)
58	Low-pressure cylinder exhaust temperature (°C)
59	Inlet flue pressure of the fixed end economizer (MPa)
60	Feed water temperature at economizer inlet (°C)
61	Pressure of the main water supply pipe at economizer inlet (MPa)
62	Inlet flue pressure of the economizer at expansion end (MPa)

- (1) According to the measurement range information of each measurement point, remove the measurement point data which obviously deviates from the measurement range
- (2) Combined with the actual operation experience of the power plant, remove the data which obviously deviates from the experience value under the current working condition

At the same time, in order to shorten the time of the data processing and remove redundant information from the monitoring data, this paper proposes a feature extraction method based on the latent structure model for the monitoring data. Two common methods, principal component analysis (PCA) [29, 30] and independent component analysis (ICA) [31, 32], are used in this latent structure model.

3.2. Principal Component Analysis. In coal-fired power plants, there are many conventional measurement points. The soft computing method, by which we obtain the coal quality-related information through edge data mining, requires the processing of the high-dimensional data. The principal component analysis method is a commonly used data dimensionality reduction method, which aims at maximizing the variance of the data after dimensionality reduction. The PCA algorithm is used to map the raw data to

the low-dimensional feature space with most of the information saving. At the same time, the algorithm realizes the data compression and avoids too many parameters of the following neural network.

In this paper, the data of the conventional measurement points and the coal quality test are used as the test data of the model. Set the data set of the conventional measurement points of the coal-fired power plants after data preprocessing as U .

$$U = \begin{bmatrix} X_{0,t_0} & X_{0,t_1} & \cdots & X_{0,t_{N-1}} \\ X_{1,t_0} & X_{1,t_1} & \cdots & X_{1,t_{N-1}} \\ \vdots & \vdots & \vdots & \vdots \\ X_{M-1,t_0} & X_{M-1,t_1} & \cdots & X_{M-1,t_{N-1}} \end{bmatrix}, \quad (1)$$

where $\{X_{M-1,t_0}, X_{M-1,t_1}, \dots, X_{M-1,t_{N-1}}\}$ is the data of the M th coal-fired power plants' conventional measuring points at the sampling time of the $t_0, t_1, \dots, t_{N-1}$, respectively. N is the number of samples, and M is the total number of the conventional measuring points. The correlation matrix R of the U^T is calculated by

$$R = E[U^T U]. \quad (2)$$

Find the eigenvalue λ_{ij} of the R , queue the eigenvalues from the big to the small, select the first d bigger eigenvalues,

and calculate the eigenvectors corresponding to the d eigenvalues. After normalization, record them as $U_j, j = 1, 2, \dots, d$. The transformation matrix A is composed of the U_j .

$$A = (U_1, U_2, \dots, U_d). \quad (3)$$

K - L transformation is applied to the sample set U^T . If the transformed matrix is I , then

$$I = A^T U^T, \quad (4)$$

where vector I is the low-dimensional data obtained after principal component analysis, and the data dimension is reduced to d dimension. After the principal component analysis, the correlation of the data is removed by using the two-order statistical information, the processed data may still have higher order redundant information, and the components of the data may not have mutual independence. Therefore, the independent component analysis is used to obtain the independent component of the data after the principal component analysis.

3.3. Independent Component Analysis. Independent component analysis (ICA) is a data analysis method that aims at the independence of the processed data components. Taking Fast-ICA algorithm as an example, this paper describes the implementation of the ICA algorithm. Fast-ICA algorithm, also known as fixed point algorithm, is a kind of the fast optimization iterative algorithm, which has forms based on kurtosis, likelihood, and negative entropy. In this section, taking Fast-ICA algorithm based on negative entropy as an example, suppose the data to be processed is X ; then the Fast-ICA algorithm determines the separation matrix W by observing the data X . The data result Y of the independent component analysis is shown in

$$Y = WX. \quad (5)$$

According to the information theory, among random variables with the same variance, Gaussian random variables have the largest differential entropy. According to the central limit theorem, the stronger the non-Gaussianity of the Y , the greater its negative entropy and the stronger its independence. The decision basis for the Fast-ICA algorithm based on negative entropy is the maximization of the negative entropy. The definition of the negative entropy is given in

$$J_g(Y) = H(Y_{\text{Gauss}}) - H(Y), \quad (6)$$

$$H(Y) = - \int p(Y) \lg p(Y) dY, \quad (7)$$

where $J_g(Y)$ is the negative entropy, Y_{Gauss} is a Gaussian random variable with the same variance as Y , and $H(Y_{\text{Gauss}})$ is the entropy of the random variable. In practical applications, in order to avoid using the unknown probability density distribution function of the variable Y , the approximate formula of the negative entropy is shown in

$$J(Y) \approx [E\{G(Y)\} - E\{G(Y_{\text{Gauss}})\}]^2, \quad (8)$$

where E is the mean calculation and G is a nonlinear function.

Because the data is generally standardized before being analyzed by Fast-ICA algorithm, the constraint is given by in

$$\|WX\| = \|W\| = 1. \quad (9)$$

From equation (8), combined with the method of the Lagrange multipliers, equation (10) can be obtained as follows:

$$E\{XG(WX)\} + \beta W = 0, \quad (10)$$

where β is the Lagrangian multiplier.

In practical applications, the iterative algorithm and equation (10) can be combined to realize the processing of the independent component analysis on the data.

4. Establishment of the Alertness Mechanism

The basic theory of the alertness mechanism originates from philosophy, cognitive psychology, social science, and linguistics. It is a new mechanism to strengthen the key information of the data based on prior knowledge [33]. The data used for online measurement of the coal quality is the actual operation data of a coal-fired power plant. In the actual operation of a coal-fired power plant, due to the segmented control behavior of the control system or the subjective operation of the operators, the regularity of the raw data in the time series will be destroyed in some places. Because this kind of the damage will reduce the accuracy of the model, and according to the prior knowledge, it has a certain alertness feasibility. In this paper, the alertness mechanism is introduced to data processing for the purpose of optimizing the model. The specific implementation process is as follows:

- (1) According to the prior knowledge of the operation in a coal-fired power plant, determine the data location of conventional measurement points that need the active alert, such as the changing points of the total coal input of the coal mill, plant load, and boiler water level.
- (2) The active alert matrix C is defined. The alert matrix is a sparse matrix. The value of each element in the matrix is the alert weight. The data position weight that needs alert is not 0, and the rest position elements are 0.

$$C \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & W_{j,t_1} & \cdots & 0 \\ W_{i,t_0} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & W_{h,t_{N-1}} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (11)$$

where $W_{i,t_0}, W_{j,t_1}, \dots, W_{h,t_{N-1}}$ are alert weights and subscripts $i, j, h, t_0, t_1, t_{N-1}$ indicate the location and time of the alert data.

- (3) Modify the data; introduce the alertness mechanism to the feature vector extracted by the latent structure model. Let the input of the alertness mechanism be matrix I ,

$$I = \{I_{t_0}, I_{t_1}, \dots, I_{t_{N-1}}\}, \quad (12)$$

where $I_{t_0}, I_{t_1}, \dots, I_{t_{N-1}}$ are d -dimensional eigenvectors. The data after the introduction of the alertness mechanism are shown in equation (13), where N is the number of samples, $I_{t_0}, I_{t_1}, \dots, I_{t_{N-1}}$ represent the sampling time series, I is the original data, and W is the alert weight.

$$I_A \triangleq I + C \times I = \begin{bmatrix} I_{0,t_0} & I_{0,t_1} & \cdots & I_{0,t_{N-1}} \\ I_{1,t_0} & \vdots & \cdots & I_{1,t_{N-1}} \\ \vdots & (1 + W_{j,t_1}) \times I_{j,t_1} & \cdots & I_{2,t_{N-1}} \\ (1 + W_{i,t_0}) \times I_{i,t_0} & I_{j+1,t_1} & \cdots & \vdots \\ I_{i+1,t_0} & I_{j+2,t_1} & \cdots & (1 + W_{h,t_{N-1}}) \times I_{h,t_{N-1}} \\ \vdots & \vdots & \vdots & \vdots \\ I_{d-1,t_0} & I_{d-1,t_1} & \cdots & I_{d-1,t_{N-1}} \end{bmatrix}. \quad (13)$$

- (3) The object of the alertness mechanism is some specific discrete data points, while attention mechanism introduces attention mechanism to all or part of the data.
- (4) The purpose of introducing alertness mechanism to data in the model is to reduce the damage caused by the subjective operation behavior of operators or the segmented control behavior of the control system in coal-fired power plants, while the general purpose of the attention mechanism is to pay attention to the key information in data and enhance the ability of the neural network to data mining.

The introduction of the alertness mechanism enhances the stability of the training model and suppresses the influence of the abnormal data fluctuation on the accuracy of the model output. To a certain degree, it reduces the damage of the unpredictable factors to the data regularity and optimizes the performance of the DFAS-LSTM model proposed in this paper.

5. Deeply Bidirectional Fusion LSTM Modeling

This section discusses the establishment of the deeply bidirectional fusion LSTM including improved activation function of parameters self-learning, structure of the deeply bidirectional fusion LSTM, and Encoder-Decoder framework with attention mechanism.

- (4) The output of the alertness mechanism is the input of the long short-term memory neural network in the model. With the training process, the weight of the alertness mechanism and the neural network parameters are modified by the gradient descent method.

The alertness mechanism proposed in this paper refers to the mechanism of the attention mechanism in neural network, and it is obviously different from attention mechanism [34, 35]. The main differences between alertness mechanism and attention mechanism are as follows:

- (1) The initial setting of the alertness mechanism's weight is based on the prior knowledge of practical problems, while attention mechanism does not need the support of the prior knowledge.

5.1. Improved Activation Function of the Parameter Self-Learning. Long short-term memory (LSTM) is a variant of the recurrent neural network. Different from the traditional recurrent neural network, LSTM neural network uses three gate controllers: input gate, output gate, and forgetting gate. On the basis of the original short-term memory, memory units are added to maintain long-term memory. Reference [36] studied the recent LSTM variants, summarized the results of 5400 experiments, and found that the forgetting gate and output activation function are the most critical components. Compared with the traditional recurrent neural networks model, the long short-term memory neural network uses gate structure, which enhances the selective memory ability of the neural network and overcomes the problem that the traditional recurrent neural networks are prone to gradient explosion and gradient dispersion in dealing with long-term sequential problems. Therefore, it has a unique advantage in dealing with long-term sequential problems. Activation function is an indispensable part of the long short-term memory neural network, and the activation function of its input gate plays an important role in the mapping process from input to neuron state [37]. Tanh function is a commonly used activation function of the bidirectional long short-term memory neural network. The expression is given in equation (14).

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (14)$$

The tanh activation function and its derivative curve are shown in Figure 4.

As shown in Figure 4, tanh function has a wide saturation region, in which the derivative of the tanh function is almost zero. Because the neural network uses the gradient descent method to modify the network weight, when the activation function enters the saturation region, the modification of the weight will be very slow. When tanh is selected as the activation function, if there is a large number of the data input, the weight parameters may be congested due to the slow correction of the weight, resulting in a longer training time or even inability to train. To solve the above problems of the tanh activation function, this paper proposes an improved multiparameter hyperbolic tangent activation function $f(x)$, whose expression is given by

$$f(x) = \lambda \tanh(\gamma h) + \eta x, \quad (15)$$

where λ regulates the output amplitude of the activation function, γ regulates the scale of the independent variable, and η regulates the gradient of the activation function, reflecting the gradient limit of the activation function. The function curve of the improved activation function $f(x)$ is shown in Figure 5.

λ , γ , and η are adjustable parameters in neural network, and their values affect the performance of the neural network model. In this paper, λ , γ , and η are set as optimization variables. After the initial value is set, with the training process, the gradient descent method is used to optimize the values of λ , γ , and η . After the training, the appropriate values of λ , γ , and η are determined and solidified into the model. In the model training process, the value optimization process of λ , γ , and η is shown in Figure 6.

It can be seen from Figure 6 that the parameters λ , γ , and η in the improved activation function of the multiparameters self-learning realize self-learning with the training process. The parameter correction process is relatively slow, and the parameter correction of λ , γ , and η in this model tends to be stable after about 50000 steps. After training convergence, the activation function parameters λ , γ , and η in this model are about: $\lambda = 0.2442114$, $\gamma = 2.82857346$, and $\eta = 0.01878586$. The improved hyperbolic tangent activation function avoids the difficulty due to gradient saturation during training. At the same time, this method realizes the process of the independent optimization in the model. The influence of the subjective parameters on the performance of the neural network is weakened. The parameter self-learning method proposed in this paper is also suitable for some common superparameters and provides a new idea for the optimization of the superparameters in neural network.

5.2. Structure of the Bidirectional Deep Fusion LSTM. When using deep learning methods to deal with sequence problems, RNN is a common and effective method. During the operation of a coal-fired power plant, it takes a long time for the material and energy of the coal to be completely converted. Data on coal quality is scattered over a long time series. As a kind of the RNN, LSTM has better performance in dealing with long sequence problems than traditional RNN. Bidirectional long short-term memory (Bi-LSTM) networks,

FIGURE 4: The tanh activation function and its derivative curve.

FIGURE 5: Curve of the improved activation function $f(x)$.

FIGURE 6: Curve of the activation function parameter as training converges.

compared with unidirectional long short-term memory networks, consider the relevance of the data in the reverse direction, which helps to fully mine the relevance of the data

in the forward and reverse direction [38]. On the basis of the Bi-LSTM, this paper proposes a structure of the deeply bi-directional fusion LSTM model. The structure of the deeply bi-directional fusion LSTM model is shown in Figure 7.

The structure of the deeply bi-directional fusion LSTM is an important part of the DFAS-LSTM proposed in this paper. It uses the fusion layer to realize the fusion of the forward and the reverse data in the hidden layer of the model. Compared with the deeply bi-directional fusion LSTM, the traditional bi-directional fusion LSTM is essentially two independent unidirectional LSTM networks, and there is no bi-directional data fusion in the hidden layer of the network. The structure design of the forward and reverse data splitting hinders the ability of the hidden layer in the neural network to extract bi-directional data features. In this paper, the deeply bi-directional fusion LSTM structure constructed by DFAS-LSTM model overcomes the defect of the traditional bi-directional LSTM by fusion structure, which makes the neural network structure have stronger ability of the data feature representation. Deeply bi-directional fusion LSTM structure is the core structure of the DFAS-LSTM model, which consists of the input layer, forward LSTM layer, reverse LSTM layer, bi-directional data fusion layer, and output layer. Among them, the fusion layer is the key structure to realize bi-directional data fusion in DFAS-LSTM model and also the key to distinguish the deeply bi-directional fusion LSTM from the traditional deeply bi-directional LSTM.

In the fusion layer, bi-directional fusion weight, sigmoid function, and Encoder-Decoder unit are set. In this structure, the input of this structure is the output of the upper bi-directional LSTM neuron, and the output is the input of the lower LSTM neuron. After the data enters the fusion layer, the forward and reverse neuron output data are given fusion weights. After bi-directional data superposition, sigmoid function is used to output the fused vector. This vector outputs the vector data of the specified dimension through the Encoder-Decoder framework that introduces the attention mechanism. This output is connected to the bi-directional LSTM neuron node of the lower layer. The structure of the bi-directional data fusion layer is shown in Figure 8.

In order to visually characterize the working process of the fusion layer, mathematical expressions are given. Suppose the input of the fusion layer, that is, the output of the previous layer, is $\mathbf{Y}_f$ and $\mathbf{Y}_b$, respectively. Then, the mathematical operation process of the fusion layer is shown in equations

$$\mathbf{Y} = \text{sigmoid}(\mathbf{Y}_f \times \mathbf{W}_f + \mathbf{Y}_b \times \mathbf{W}_b), \quad (16)$$

$$\mathbf{Y}_f = [\gamma_{f1}, \gamma_{f2}, \gamma_{f3}, \dots, \gamma_{fn}], \quad (17)$$

$$\mathbf{W}_f = [w_{f1}, w_{f2}, w_{f3}, \dots, w_{fn}], \quad (18)$$

$$\mathbf{Y}_b = [\gamma_{b1}, \gamma_{b2}, \gamma_{b3}, \dots, \gamma_{bn}], \quad (19)$$

$$\mathbf{W}_b = [w_{b1}, w_{b2}, w_{b3}, \dots, w_{bn}], \quad (20)$$

where $\mathbf{Y}$ is the input of the Encoder-Decoder framework, $\mathbf{Y}_f$ is the forward output matrix, $\mathbf{Y}_b$ is the reverse output matrix,

FIGURE 7: Structure of the deeply bi-directional fusion LSTM model.

FIGURE 8: Structure of the bi-directional data fusion layer.

$\mathbf{W}_f$ is the forward fusion weight, and $\mathbf{W}_b$ is the reverse fusion weight. Sigmoid function enhances the nonlinear fitting ability of the fusion layer, adjusts the fusion result range, and enhances the representation ability of the model.

5.3. Encoder-Decoder Framework with Attention Mechanism.

As shown in Figure 8, the fusion layer uses the Encoder-Decoder framework to adjust the data after bi-directional data fusion. The working mechanism of the Encoder-Decoder framework is as follows: use encoder to map the input to the specified dimension space, get the fixed dimension decoding vector $\mathbf{C}$, and then use decoder structure to decode the decoding vector $\mathbf{C}$. This structure realizes the function of the data feature acquisition and data structure adjustment.

In this paper, the fusion layer of the DFAS-LSTM model is constructed based on Encoder-Decoder framework, and attention mechanism is added. The attention mechanism in deep learning is essentially similar to the selective visual attention mechanism of the human beings. Its core goal is to select the primary and secondary information of the current task, so as to achieve the purpose of paying attention to the primary information and ignoring the secondary information. Compared with the traditional Encoder-Decoder framework which introduces attention mechanism encodes the input into a vector sequence. At the same time, each vector in the vector sequence obtains different attention weights according to its importance to the target output. Then, the decoder structure decodes the vector sequence with attention weight. In the fusion layer, the Encoder-Decoder framework of this structure can effectively obtain the key information in the bidirectional fusion data and improve the data feature extraction ability of the fusion layer. The Encoder-Decoder framework with attention mechanism is shown in Figure 9.

6. Experiment and Result Analyses

This section discusses the performance of the DFAS-LSTM in coal quality computing. Content includes presetting and training, metrics soft computing of the industrial coal quality analyses, and soft computing of the elemental coal quality analyses.

6.1. Presetting and Training. Our algorithm is implemented in Tensorflow-1.12.0 with the Python wrapper and using eight cores of a 3.6 GHz Intel Core i7-7700 CPU and two NVIDIA GeForce GTX 1080 Ti GPUs. The data for model training and testing comes from actual operating data of a coal-fired power plant. Among them, the coal quality data comes from the laboratory data of a coal-fired power plant. Coal quality data includes industrial analyses data and elemental analysis data. The soft computing method proposed in this paper can realize industrial analyses and element analyses of the coal quality based on these data. Among them, the industrial analyses of the coal quality include low calorific value, total moisture, ash content, and volatile; the elemental analysis of the coal quality includes carbon content, hydrogen content, oxygen content, nitrogen content, and sulfur content.

The DFAS-LSTM model is used to solve the problem of the coal quality soft computing. The framework of the DFAS-LSTM model for coal quality soft computing is shown in Algorithm 1.

6.2. Metrics. In order to intuitively represent the ability of the model, this paper uses the fitting index and root mean square error as the evaluation index of the model. Equation (21) shows the expression of the fitting index.

FIGURE 9: Encoder-Decoder with attention mechanism.

$$R_f = 1 - \sqrt{\frac{\sum_i^n (y_i - y_{ip})^2}{\sum_i^n y_i^2}}, \quad (21)$$

where R_f is the fitness index, y_i is the real value of the sample, and y_{ip} is the output value of the model.

The model proposed in this paper belongs to regression model, using root mean square error as the loss function of the model; the loss function can evaluate the deviation degree between the model output and the real value, and the smaller the value, the better the robustness of the model. The expression of the RMSE loss function is given in

$$\text{RMSE} = \sqrt{\frac{1}{n} \times \sum_i^n (y_i - y_{ip})^2}, \quad (22)$$

where RMSE is the root mean square error, n is the number of samples, y_i is the true value of samples, and y_{ip} is the output value of the model.

6.3. Soft Computing of Industrial Coal Quality Analyses. Industrial analyses of the coal, also called technical analysis or practical analysis of the coal, is the basis for evaluating coal quality and an important indicator for understanding coal quality. In this section, soft computing of the industrial analyses is carried out, which includes low calorific value, total moisture, ash content, and volatile. The specific meaning of each industrial analysis involved is as follows:

- (1) Low calorific value: the low calorific value of the coal refers to the heat produced by the combustion of the coal under atmospheric pressure, after deducting the vaporization heat of the moisture in the coal, the remaining heat that can actually be used
- (2) Total moisture: it is the moisture that the coal sample loses when it is in the air and reaches equilibrium with air humidity
- (3) Ash content: the ash content of the coal refers to the residue left after the coal is completely burned
- (4) Volatile: the volatile of the coal is the content heated by insulation at a certain temperature, and

Input: Data of the historical conventional measurement points of the coal-fired power plants X_h ; Real-time conventional measurement data of the coal-fired power plants X_r ; Coal quality test data of the coal-fired power plants y_r ;

Output: Real-time coal quality data in the furnace of the coal-fired power plants y ;

- (1) Remove noise from historical data X_h and filter it;
- (2) Standardize the data, process the standard data by PCA and ICA algorithm to obtain data X_i ;
- (3) Initialize weight parameters, batch the data to get X_{ib} , and input X_{ib} into DFAS-LSTM;
- (4) Use alertness mechanism to process data X_{ib} and obtain data X_a ;
- (5) Data X_a input to LSTM, which is based on improved activation function and fusion structure;
- (6) Use the coal quality test data y_r to compare with the output of the neural network to obtain the cost function C ;
- (7) Use the optimizer to optimize the cost function C by updating the weight parameters of the neural network;
- (8) After the model is stable, the optimization ends and the model parameters are solidified;
- (9) Take X_r as input, output coal quality information y in real time.

ALGORITHM 1: The soft computing of the coal quality analyses by using DFAS-LSTM model.

FIGURE 10: Statistical results of the industrial analyses accuracy.

the moisture is subtracted from the escaped material

In order to verify that the DFAS-LSTM model proposed in this paper has advantages, in this section, based on the data, soft computing of the industrial analyses is realized by using DFAS-LSTM, conventional Bi-LSTM model, Bi-LSTM model with improved activation function, and Bi-LSTM with alertness mechanism. Each model runs 20 times, the accuracy obtained is averaged, and the statistics are shown in Figure 10.

Using the DFAS-LSTM model proposed in this paper, based on the conventional measurement points data of a coal-fired power plant, the soft computing of the above industrial analyses' information is realized. In chronological order, 20 data points were selected at random time to show the actual value of industrial analyses and the soft

computing result for comparison. The result is shown in Figure 11.

6.4. Soft Computing of Elemental Coal Quality Analyses.

In this section, DFAS-LSTM model is used to achieve soft computing for elemental analysis of the coal. Elemental analysis of the coal is to detect and analyze the element content in coal. The element content in coal is an important indicator of the coal quality. The element analyses data used in this paper is based on the received basis. In this section, elemental analysis specifically includes carbon content, hydrogen content, oxygen content, nitrogen content, and sulfur content. Similarly, elemental analysis soft computing is realized by using DFAS-LSTM, conventional Bi-LSTM model, Bi-LSTM model with improved activation function, and Bi-LSTM with alertness mechanism. Each model runs 20 times, the accuracy obtained is averaged, and the statistics are shown in Figure 12.

FIGURE 11: The results of industrial analyses by using DFAS-LSTM model: (a) the actual and test value of the low calorific value; (b) the actual and test value of the total moisture; (c) the actual and test value of the ash content; (d) the actual and test value of the volatile.

FIGURE 12: Statistical results of the elemental analyses accuracy.

For elemental analyses of the coal, the same work has been done as industrial analyses. Using the DFAS-LSTM model proposed in this paper, the soft computing of the above elemental analyses is realized. In chronological

order, 20 data points were selected at random time to show the actual value of the elemental analyses and the soft computing result for comparison. The result is shown in Figure 13.

FIGURE 13: The results of elemental analyses by using DFAS-LSTM model: (a) the actual and test value of the carbon content; (b) the actual and test value of the hydrogen content; (c) the actual and test value of the oxygen content; (d) the actual and test value of the nitrogen content; (e) the actual and test value of the sulfur content.

7. Conclusions

In this paper, the information fusion technology applied in the coal-fired power plant is discussed. As a practical application, the soft measurement of the coal quality in the power plant is achieved by the information fusion method. Combining the sensor network of a coal-fired power plant, an improved LSTM model with the bidirectional fusion, alertness mechanism, and parameter self-learning (DFAS-LSTM) is proposed to realize the soft computing of the coal quality. The use of the alertness mechanism can suppress the interference information; the use of the improved activation function of parameters self-learning can improve the accuracy of the model; the use of the bidirectional fusion structure can improve the accuracy and the

generalization ability of the model. In order to verify the superiority of the DFAS-LSTM model proposed in this paper, the model is compared with conventional Bi-LSTM model, Bi-LSTM model with improved activation function, and Bi-LSTM with alertness mechanism. For the test of the model, the data of the coal-fired power plant is used to achieve the industrial and elemental analyses of the coal quality. To be specific, the industrial analyses include low calorific value, total moisture, ash content, and volatile. The accuracies of industrial analyses using DFAS-LSTM model are 85.32%, 82.21%, 83.79%, and 83.01%. The elemental analyses include carbon content, hydrogen content, oxygen content, nitrogen content, and sulfur content. The accuracies of the elemental analyses using DFAS-LSTM model are 84.65%, 80.29%, 81.26%, 81.82%, and 81.20%. The

verification shows that the DFAS-LSTM model proposed in this paper basically completes the functions of industrial and elemental analyses, which provides support for online analyses of the coal quality in coal-fired power plants. Traditional measurement methods rely on expensive equipment and cannot achieve coal quality analyses in real time. The soft measurement method proposed in this paper avoids the weakness of the traditional measurement methods and saves the cost. In the future, the correlation between conventional measurement points and coal quality will be further analyzed. By removing the measurement points with low correlation, the data dimension will be further reduced. The model parameters and the time consumed for analyses will reduce at the same time.

Data Availability

The data used to support the findings of this study are available from the corresponding author Mei Wang, whose e-mail is wangm@xust.edu.cn.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the paper.

Acknowledgments

This work was supported by the Shaanxi Province Science and Technology Project (2016GY-040) and the National Natural Science Foundation of China (51804249).

References

- [1] S. D. Kumar and D. Subha, "Prediction of depression from EEG signal using long short-term memory (LSTM)," in *Proceedings of the 3rd international conference on trends in electronics and informatics (ICOEI)*, pp. 1248–1253, Tirunelveli, India, April 2019.
- [2] O. Barut, L. Zhou, and Y. Luo, "Multitask LSTM model for human activity recognition and intensity estimation using wearable sensor data," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8760–8768, 2020.
- [3] Z. Ding, R. Xia, J. Yu, X. Li, and J. Yang, "Densely connected bidirectional LSTM with applications to sentence classification," *Natural Language Processing and Chinese Computing*, vol. 1, pp. 278–287, 2018.
- [4] Y. Huang, W. Wang, and L. Wang, "Instance-aware image and sentence matching with selective multimodal LSTM," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2310–2318, 2017.
- [5] H. Xue, D. Q. Huynh, and M. Reynolds, "A Hierarchical LSTM model for pedestrian trajectory prediction," in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1186–1194, Lake Tahoe, NV, USA, March 2018.
- [6] J. Liu, A. Shahroudy, D. Xu, and G. Wang, "Spatio-temporal LSTM with trust gates for 3D human action recognition," *Computer Vision - ECCV 2016*, vol. 40, pp. 816–833, 2016.
- [7] S. O. Sahin and S. S. Kozat, "Nonuniformly sampled data processing using LSTM networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1452–1461, 2019.
- [8] B. Liu, N. Liu, G. Chen, X. Dai, and M. Liu, "A low-cost vehicle anti-theft system using obsolete smartphone," *Mobile Information Systems*, vol. 2018, Article ID 6569826, 16 pages, 2018.
- [9] H. Lu, M. Zhang, X. Xu, Y. Li, and H. T. Shen, "Deep fuzzy hashing network for efficient image retrieval," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 166–176, 2020.
- [10] B. Siegel, "Industrial anomaly detection: a comparison of unsupervised neural network architectures," *IEEE Sensors Letters*, vol. 4, no. 8, pp. 1–4, 2020.
- [11] C. Zou, Q. Zhao, G. Zhang, and B. Xiong, "Energy revolution: from a fossil energy era to a new energy era," *Natural Gas Industry B*, vol. 3, no. 1, pp. 1–11, 2016.
- [12] Y. Zhu, R. Zhai, H. Peng, and Y. Yang, "Exergy destruction analysis of solar tower aided coal-fired power generation system using exergy and advanced exergetic methods," *Applied Thermal Engineering*, vol. 108, pp. 339–346, 2016.
- [13] J. Singh and S. Thakur, *Laser-induced Breakdown Spectroscopy*, Elsevier, Amsterdam, Netherlands, 2020.
- [14] Z. Zang, X. Qiu, Y. Guan, E. Zhang et al., "Determining moisture content of traditional Chinese medicines using a near-infrared LED-based moisture content sensor with spectrum analysis," *Optical and Quantum Electronics*, vol. 51, no. 5, pp. 51–133, 2019.
- [15] H. Schwarz and X. Cai, "Special issue: development of renewable energy and smart grid," *Frontiers in Energy*, vol. 11, no. 2, pp. 105–106, 2017.
- [16] B. Looney, *BP Statistical Review of World Energy: BP Statistical Review*, Baker Library, Hanover, NH, USA, 2020.
- [17] G. Kurata, B. Ramabhadran, G. Saon Sathy, and A. Sathy, "Language modeling with highway LSTM," in *Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 244–251, Okinawa, Japan, December 2017.
- [18] S. Y. Tseng, S. N. Chakravarthula, B. R. Baucom, and P. Georgiou, "Couples behavior modeling and annotation using low resource LSTM language models," *Interspeech*, vol. 12, no. 1, pp. 898–902, 2016.
- [19] B. Athiwaratkun and J. W. Stokes, "Malware classification with LSTM and GRU language models and a character-level CNN," in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2482–2486, New Orleans, LA, USA, March 2017.
- [20] X. Shi, Z. Chen, H. Wang et al., "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 2015, pp. 802–810, 2015.
- [21] Y. Pan, T. Mei, T. Yao et al., "Jointly modeling embedding and translation to bridge video and language," 2016, <https://arxiv.org/abs/1505.01861>.
- [22] K. Amarasinghe, D. L. Marino, and M. Manic, "Deep neural networks for energy load forecasting," in *Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pp. 1483–1488, Scotland, UK, June 2017.
- [23] D. L. Marino, K. Amarasinghe, and M. Manic, "Simultaneous generation-classification using LSTM," in *Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, Athens, Greece, December 2016.
- [24] E. Oko, M. Wang, and J. Zhang, "Neural network approach for predicting drum pressure and level in coal-fired subcritical power plant," *Fuel*, vol. 151, pp. 139–145, 2015.
- [25] A. Ayodeji, Y.-k. Liu, and H. Xia, "Knowledge base operator support system for nuclear power plant fault diagnosis," *Progress in Nuclear Energy*, vol. 105, pp. 42–50, 2018.

- [26] M. Talaat, M. H. Gobran, and M. Wasfi, "A hybrid model of an artificial neural network with thermodynamic model for system diagnosis of electrical power plant gas turbine," *Engineering Applications of Artificial Intelligence*, vol. 68, pp. 222–235, 2018.
- [27] M. Rashid, K. Kamal, T. Zafar et al., "Mathavan. Energy prediction of a combined cycle power plant using a particle swarm optimization trained feedforward neural network," in *Proceedings of the 2015 International Conference on Mechanical Engineering, Automation and Control Systems (MEACS)*, pp. 1–5, Novosibirsk, Russia, December 2015.
- [28] J. Li, Z. Wu, K. Zeng, G. Flamant, A. Ding, and J. Wang, "Safety and efficiency assessment of a solar-aided coal-fired power plant," *Energy Conversion and Management*, vol. 150, pp. 714–724, 2017.
- [29] H. Chen, W. Rong, X. Ma et al., "An extended technology acceptance model for mobile social gaming service popularity analysis," *Mobile Information Systems*, vol. 2017, Article ID 3906953, 12 pages, 2017.
- [30] S. Liu, L. Feng, J. Wu, G. Hou, and G. Han, "Concept drift detection for data stream learning based on angle optimized global embedding and principal component analysis in sensor networks," *Computers & Electrical Engineering*, vol. 58, pp. 327–336, 2017.
- [31] Z. Uddin, A. Ahmad, M. Iqbal et al., "Adaptive step size gradient ascent ICA algorithm for wireless MIMO systems," *Mobile Information Systems*, vol. 2018, Article ID 7038531, 9 pages, 2018.
- [32] T. Venkatakrishnamoorthy and G. U. Reddy, "Cloud enhancement of NOAA multispectral images by using independent component analysis and principal component analysis for sustainable systems," *Computers & Electrical Engineering*, vol. 74, pp. 35–46, 2019.
- [33] J. Bowler and P. Bourke, "Facebook use and sleep quality: light interacts with socially induced alertness," *British Journal of Psychology*, vol. 110, no. 3, pp. 519–529, 2019.
- [34] D. K. Jain, R. Jain, Y. Upadhyay, A. Kathuria, and X. Lan, "Deep Refinement: capsule network with attention mechanism-based system for text classification," *Neural Computing and Applications*, vol. 32, no. 7, pp. 1839–1856, 2020.
- [35] Y. Wang, C. Hu, K. Chen et al., "Self-attention guided model for defect detection of aluminium alloy casting on X-ray image," *Computers & Electrical Engineering*, vol. 88, pp. 1–8, 2020.
- [36] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.
- [37] M. Chammas, A. Makhoul, and J. Demerjian, "An efficient data model for energy prediction using wireless sensors," *Computers & Electrical Engineering*, vol. 76, pp. 249–257, 2019.
- [38] Y. Yao and Z. Huang, "Bi-directional LSTM recurrent neural network for Chinese word segmentation," *Neural Information Processing*, vol. 40, pp. 345–353, 2016.