

Retraction

Retracted: Application Study of Sigmoid Regularization Method in Coke Quality Prediction

Complexity

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Complexity. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] S. Yan, H. Zhao, L. Liu, Q. Sang, P. Chen, and J. Li, "Application Study of Sigmoid Regularization Method in Coke Quality Prediction," *Complexity*, vol. 2020, Article ID 8785047, 10 pages, 2020.

Research Article

Application Study of Sigmoid Regularization Method in Coke Quality Prediction

Shaohong Yan,^{1,2} Hailong Zhao,^{2,3} Liangxu Liu,³ Qiaozhi Sang,² Peng Chen,² and Jie Li³

¹College of Sciences, North China University of Science and Technology, Tangshan 063200, China

²College of Metallurgy and Energy, North China University of Science and Technology, Tangshan 063200, China

³North China University of Science and Technology Innovation of Mathematical Modeling Laboratory, Tangshan 063200, China

Correspondence should be addressed to Jie Li; lijie@ncst.edu.cn

Received 7 April 2020; Accepted 22 June 2020; Published 20 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Shaohong Yan 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.

Coke is an indispensable and vital flue for blast furnace smelting, during which it plays a key role as a reducing agent, heat source, and support skeleton. Models of prediction of coke quality based on ANN are established to map the functional relationship between quality parameters M_b , A_d , V_{daf} , $S_{t,d}$ and caking property (X , Y , and G) of mixed coal and quality parameters A_d , $S_{t,d}$, coke reactivity index (CRI), and coke strength after reaction (CSR) of coke. A regularized network training method based on Sigmoid function is designed considering that redundancy of network structure may lead to the learning of undesired noise, in which weights having little impact on performance and leading to overfitting are removed in terms of computational complexity and training errors. The cascade forward neural network with validation is found to be the most suitable one for coke quality prediction, with errors around 5%, followed by feedforward neural network structure and radial basis neural networks. The cascade forward neural network may play a guiding role during the coke production.

1. Introduction

With the growing trend of large-scale blast furnace, smelting effect of blast furnace and its economic and technical indicators are more deeply influenced by the quality and performance of coke [1]. Unfortunately, various types and proportions of mixed coal can lead to different coke quality [2]. In that case, it is of significance to study the relationship between physical and chemical properties of mixed coal and that of coke in controlling the coke quality.

Due to multidimension, being dynamic, incompleteness, and uncertainty, the data of coking test is difficult to collect and handle, and this makes the actual production process still in an unscheduled and irregular state and causes the occurrence of the “rich data and poor knowledge” phenomenon [3]. Rule data mining technology is capable of finding the potential link between historical data, promoting information transmission, digging knowledge from data, and providing decision basis [4]. Numerous researches and experiments have been carried out in response to coke

quality prediction. Alvarez, R et al. predicted the CSR property of coke using the law of addition [5]. Roest et al. used different statistical analysis tools (MLR and PCR) and the ANN technique to solve the same problem [6]. Using the thought of coal blend with petrography, Tao Peisheng guided the coke quality prediction and determination of the coal blending ratio based on coal-rock phase composition and reflectance pattern of single coal while taking into account coal maceral and coke strength, etc. [7]. Although the above studies enrich the theory of coke quality prediction models, rare deep data rule mining is adopted for relevant data [8]. A large number of models predicting coke quality have been proposed up to now, a majority of which are merely based on coal characteristics and limited to the same coal geography origin, but no general applicable prediction formula has been developed up to now [9].

The statistical law of abundant coking testing data indicates the extremely strong nonlinear relationship between the physical and chemical properties of coke and that of mixed coal [10]. A variety of existing artificial coal blending

plans are inefficient and the application of corresponding results is quite poor, which fails to meet the production need of modern enterprises [11]. Fortunately, the adoption of neural network technology can decrease time consumption and reduce economic costs through rule mining of the experimental data and rationally predicting the physical and chemical properties of coke from mixed coal.

The application of data rule mining techniques is here and there right now. It becomes an important mission of interdisciplinary researchers in coking industry and mathematical application fields to guide the data mining process and draw meaningful results using the professional knowledge in coking industry.

2. Factors Affecting Coke Quality

In steel industry, coke as a fuel has been used to provide heat for the melting of slag and metal [12]. Coke is also used as a reduction agent to reduce iron ore to elemental iron. Hence, the parameters influencing the coke property are of great importance [5]. Coke production is shown in Figure 1. First of all, the raw materials are collected from the raw material yard to the mixing warehouse, crushed by a crusher, then put into a coke oven for calcination, and finally put into a blast furnace for smelting.

2.1. Mixed Coal Properties. The properties of mixed coal play key roles in determining the coke quality. Moisture, ash, volatile, sulfur, and caking property affect it in different degrees. Table 1 shows the range of content of mixed coal corresponding indicators and the impact on coke production.

2.1.1. Full Moisture (M_t). It is the sum of intrinsic and external moisture in coal. Too much moisture content is bad for processing and transportation; it also deteriorates thermal stability and thermal conductivity when burning, reduces the coke yield, and extends the coking cycle when coking. Hence, it is specified to be below 8% [13].

2.1.2. Ash (A_d). Ash in coal remains in coke after coking. Too much ash will lead to sudden drop of the cold state strength, the increase of residue in a blast furnace, and finally production reduction. Ash in coke can be guaranteed within the required range only if the ash in mixed coal is controlled, and the ash in coke is generally 1.3-1.4 times that in mixed coal. So the ash of mixed coal in general is controlled between 9% and 10% [1].

2.1.3. Volatiles (V_{daf}). The metamorphism of coal can be reflected through volatiles. The volatiles can be calculated roughly on a weighted average basis according to the volatile of single coal. The product yield of coking gas and chemical products can be improved by appropriately adding high volatile coal, with the best content being controlled in the range of 24%–30% [14].

2.1.4. Sulfur Content ($S_{t,d}$). Sulfur content is the harmful component in coke, most of which is brought into the blast furnace by coke. The sulfur content of pig iron is directly affected by it, and the quality of pig iron is declined. Therefore, the control of sulfur content is indispensable. Related study [15] shows that 60%-70% sulfur content from coal is transferred into coke, the inorganic sulfur in coal is transferred into coke as sulfocompounds, and the remaining part stays in ash in the form of sulfate and sulfide. The coke rate of mixed coal is 70%-80% usually, while the sulfur content in coke is 80%-90% of that in coal [16]. Thus, the sulfur content of the mixed coal should be constrained in the range of 0.6%-0.7%.

2.1.5. Caking Property. Caking property refers to the ability of coal forming plastic substance during coking, which serves as a necessary condition in coking process and affects the coking property of mixed coal [17]. It commonly can be reflected by three indexes, X, Y, and G in glial layer. Among them, shrinkage X can be used to estimate whether the final shrinkage of coal can cause difficulty of coke pushing, the maximum thickness of the glial layer Y indicates the amount of liquid produced with the mixed coal, and the bond index G, to a certain extent, reflects the content of the gum. So, the caking property of mixed coal can be measured by the mentioned indicators.

2.2. Index Assessing the Coke Quality. Coke provides heat and reducing gas, occupying an important position in the furnace. The hot strength (coke reactivity index CRI and coke strength after reaction CSR) is a main index in quality judgment of coke, and so are ash, sulfur, and volatile [15, 17].

2.2.1. Ash (A_d). Ash in coke is mainly composed of SiO_2 , Al_2O_3 , and other acidic oxide composition with high melting point. In the blast furnace refining slag discharge requires a lot of solvent to reduce the melting point of the compound [18]. The ash content of the coke all comes from mixed coal. The higher the ash content, the less the carbon content of the coke, the more solvent required, the more slag, and the lower yield [19]. The ash in coke is an inert component, and it will reduce the caking property of mixed coal, increase the coke crack, and reduce the coke mechanical strength. At the same time, the alkali metal oxide in the ash component also catalyzes the CO_2 reaction of the coke and increases the CO_2 reaction rate.

2.2.2. Sulfur Content ($S_{t,d}$). 80%–85% of sulfur content in coke, as one of indicators affecting coke quality, comes from the mixed coal. It influences the sulfur content in steel from the blast furnace smelting and the design of the blast furnace and destroys the environment severely [20].

2.2.3. Hot Strength of Coke. The coke supplies heat and carbon in the blast furnace, serving as a reducing agent for the furnace reaction; its hot strength is a mechanical strength

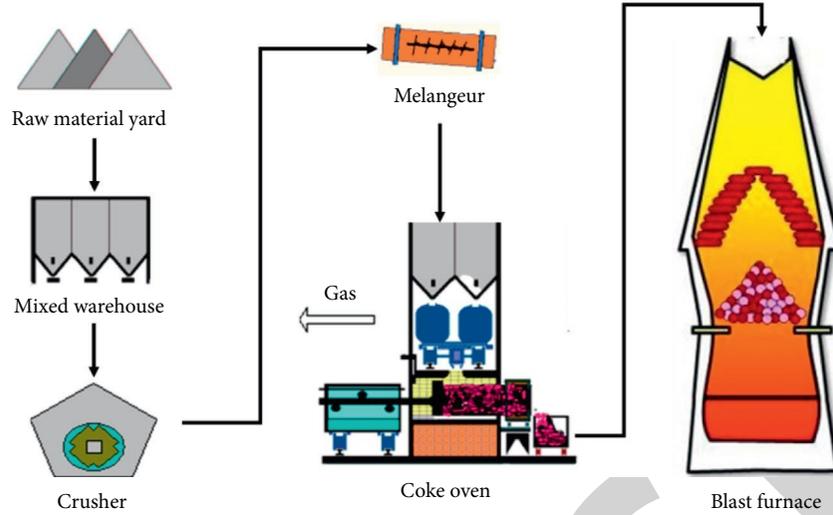


FIGURE 1: Coke production.

TABLE 1: The range of content of mixed coal corresponding indicators and the impact on coke production.

| Index of mixed coal | Range (%) | Effect |
|---------------------|--------------|---|
| M_t | $0 << 8$ | Bad for processing and transportation, deteriorates thermal stability and thermal conductivity when burning, and reduces the coke yield and extends the coking cycle when coking. |
| A_d | $9 << 10$ | Leads to sudden drop of the cold state strength, the increase of residue in a blast furnace, and finally production reduction. |
| V_{daf} | $24 << 30$ | The product yield of coking gas and chemical products can be improved by appropriately adding high volatile coal. |
| $S_{t,d}$ | $0.6 << 0.7$ | The sulfur content of pig iron is directly affected by it, and the quality of pig iron is declined. |

index reflecting the thermal performance [21]. This indicator characterizes the ability to resist crushing and abrasion subjected to thermal stress and mechanical force at specific temperature and atmosphere. The thermal performance is commonly denoted by coke reactivity index (CRI) and coke strength after reaction (CSR). The lower the CSI and the higher the CSR, the better the thermal performance of coke. At the same time, they are the most critical parameters used to evaluate the high temperature performance of coke [22].

3. Neural Networks Based on Domain Knowledge

Neural networks are often merely trained and generalized in data-driven mechanisms. However, it helps to speed up the search of approximation and improve the prediction quality if adding priori knowledge (including knowledge such as symmetry, invariance common sense, and so on) of the desired approximation function to networks. Domain knowledge focuses on the investigation of important issues or concepts in the located fields, as well as the interrelationships of them, which can make up for the shortcomings of unclear directivity and unexplainable results.

Based on the current situation that high-quality coking coal resources are relatively scarce, a multicoal source coking method is proposed. Considering that the nonlinear

relationship between parameters of mixed coal and coke determined by the inherent properties of coking environment, process and chemical reaction, cannot be described accurately by functions simply, neural networks are very suitable for coke quality prediction in coking systems, which avoids the function description between system characteristic indexes and dependent variables, gets relationship between input and output according to memory and feature extraction, and has features of generalization, distributed knowledge storage, associative memory, and parallel processing.

3.1. Multilayer Feedforward Backpropagation (FB) Network.

The pretreated samples are randomly presented to the neural network in traditional algorithm, and the output signals of neurons are revealed through forward recursion next [23]:

$$y_q = f(v_q) = f(u_q - \theta_q) = \left(\sum_{j=1}^n w_{qj} x_j - \theta_q \right). \quad (1)$$

The parameters are corrected in the negative gradient direction in terms of the selected error energy function $\zeta(n) = f(e_j(n))$ after obtaining the error signal $e_j(n) = d_j(n) - y_j(n)$ from comparing the expected value and the actual output [23].

$$w_{ji}^{(s)}(k+1) = w_{ji}^{(s)}(k) + \eta^{(s)} \delta_j^{(s)} y_j^{(s-1)},$$

$$\delta_j^{(s)} = \begin{cases} e_j(n) \phi_j'(v_j(n)), & j \text{ is the putput node,} \\ \phi_j'(v_j(n)) \sum_k \delta_k(n) w_{kj}(n), & j \text{ is the hidden node.} \end{cases} \quad (2)$$

Another algorithm is investigated owing to the slow convergence of the original one, which can be roughly divided into the following:

- (1) Heuristic improved algorithms, including back-propagation with momentum updates, search convergence method, and bulk update of variable learning rates
- (2) Numerical optimization techniques, including conjugate gradient backpropagation, recursive back-propagation based on least squares, and backpropagation with adaptive activation function

The concepts of sensitive area and its width of Sigmoid function $f(\cdot)$ are discussed in literature [24]; each implicit neuron has a corresponding respective sensitive area; directions of neurons depend on that of the hidden node. The hidden nodes interact with each other in the feature space formed by themselves, which affects the training performance of neural network. The input interval of hidden node whose output is within $(0, a)$ is defined as the a -level sensitive area A_a of the node [25]:

$$A_a = \{x \in R^{n \times 1} : 0 < f(x^T w + \theta) < a\}, \quad (3)$$

where x is the vector inputted into synapse, and its weights are $x_j, j = 1, 2, \dots, n, x = [x_1, x_2, \dots, x_n]^T$. x_j is connected to the neuron q through the weight of synapses w_{qj} ; the nonlinear activation function $f(\cdot)$ is able to regulate the amplitude of outputs and enhance abilities of classification, function approximation, anti-noise jamming, and so on; θ is the threshold to reduce the cumulative input of the activation function.

The width of A_a is defined as [25]

$$G_a = P_a - P_0 = \frac{\ln(a/(1-a))}{\|w\|}, \quad (4)$$

where P_a and P_0 represent the equivalent hyper-surface when outputs $f(x^T w + \theta)$ of hidden nodes are a and 0 , respectively, and $\|w\|$ is the L_2 norm of matrix w (also known as Euclidean norm).

3.2. Radial Basis Function (RBF) Network. The function approximation of the multilayer perceptron is realized by nesting the weighted sums, and the RBF is itself a general approximator following the interpolation theory, and it is delicate and tight in the mathematical point of view. Different from the random approximation of multilayer perceptions, RBF based on radial basis functions includes two stages in solving nonlinear mapping problems: high-dimensional transformation and least squares estimation of input samples. The idea is to map the samples non-linearly to

high-dimensional space where the single weighted sum is performed and eventually outputs results. Its rationality is explained by the configurable Cover theorem. The classical training method is to train the hidden layer by using the K-means clustering algorithm in unsupervised mode and then to calculate the weight vector of the output layer using the recursive least squares method. The method has the characteristics of simple calculation and accelerated convergence. The following calculation is carried out once the samples are given:

- (1) It is assumed that the training set $\{\varphi(i), d(i)\}_{i=1}^N$ and the number of iterations $n = 1, 2, \dots, N$ are known already. Initialize $\hat{w}(0) = 0, p(0) = \lambda^{-1}I$, where λ is a small positive number.
- (2) For a given encoder C , minimize the total clustering variance $\sum_{j=1}^K \sum_{C(i)=j} \|x_i - \hat{\mu}_j\|^2$ according to the clustering mean $\{\hat{\mu}_j\}_{j=1}^K$, resulting in the clustering mean $\{\hat{\mu}_j\}_{j=1}^K$, and minimize the encoder $C(i) = \arg \min_{1 \leq j \leq K} \|x(i) - \hat{\mu}_j\|^2$. Repeat the step until the clustering result remains unchanged.
- (3) Calculate $\mathbf{p}(n) = \mathbf{p}(n-1) - (\mathbf{p}(n-1)\Phi(n)\Phi^T(n) \mathbf{p}(n-1) / 1 + \Phi^T(n)\mathbf{p}(n-1)\Phi(n))$, where $p(n) = R^{-1}(n)$, and $R(n)$ is the $K \times K$ order related function of the hidden units: $\mathbf{R}(n) = \sum_{i=1}^n \phi(x_i)\phi^T(x_i)$, $\phi(x_i) = [\varphi(x_i, x_1), \varphi(x_i, x_2), \dots, \varphi(x_i, x_K)]^T$; among them, $\varphi(x_i, x_j) = \exp(-(1/2\sigma_j^2)\|x_i - x_j\|^2)$ $j = 1, 2, \dots, K$.
- (4) $g(n) = p(n) \cdot \Phi(n)$.
- (5) $a(n) = d(n) - \hat{w}^T(n-1)\Phi(n)$.
- (6) $\hat{w}(n) = \hat{w}(n-1) + \mathbf{g}(n)a(n)$: the weight vector of the output layer is obtained.

3.3. Algorithm to Better the Neural Network Generalization. The training process of the network can be regarded as the process of constructing the fitting curve. The network is of pretty good generalization when the mapping calculated by it is correct although inputs are beyond the training samples. There appears overfitting phenomenon with redundant network structure in which the extra synapses memorize unexpected characteristics of information. It can be readily seen that the wider the sensitive area defined by (4) is, the stronger the node generalization ability is. In addition, the ideal function is the smoothest function of approximation and mapping functions for a given error, taking up fewer computing resources. And the process of seeking for it is called network pruning. In this section, a regularized network training algorithm based on the sensitive area of Sigmoid function is developed, and the specific process is shown in Figure 2.

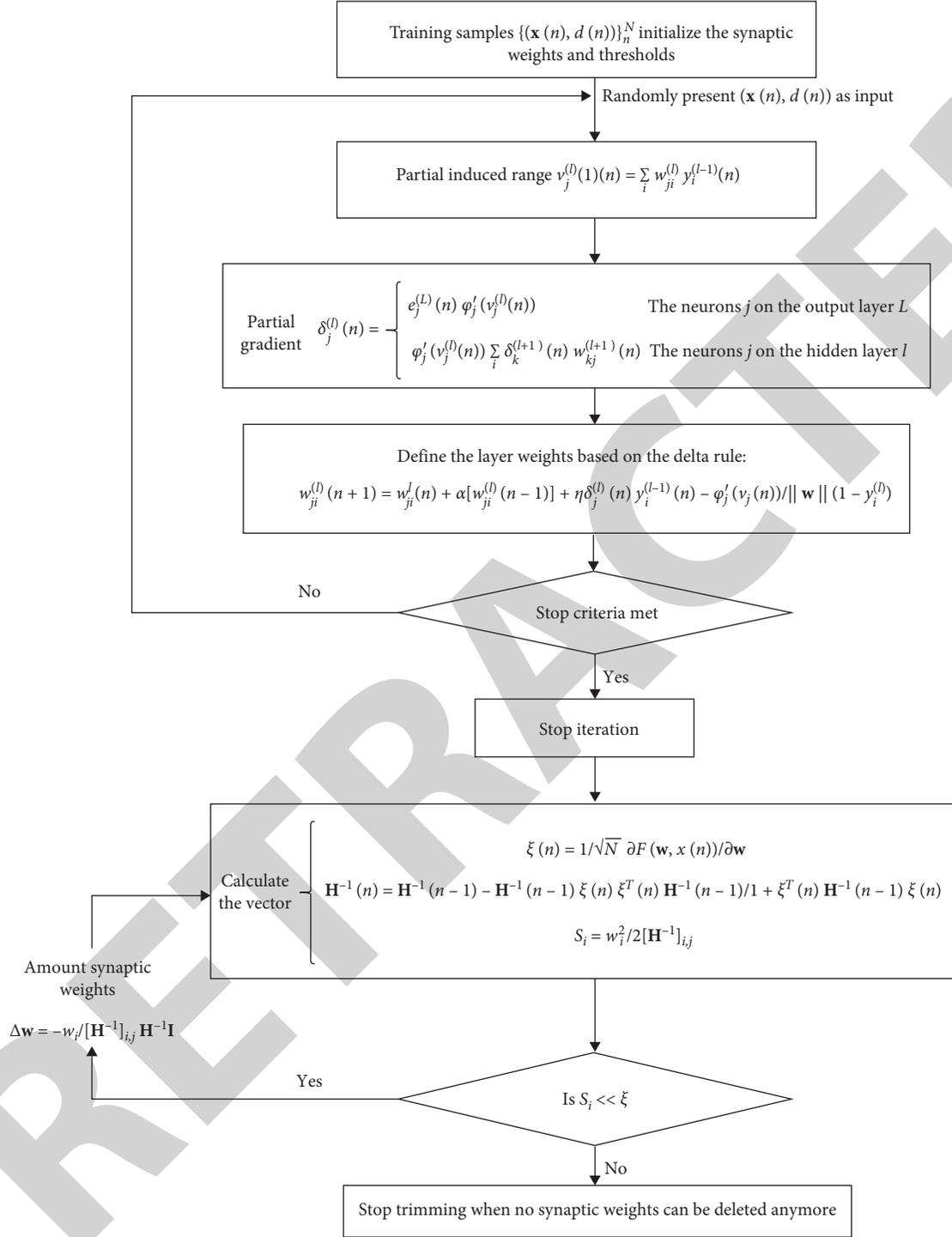


FIGURE 2: The regularized network training algorithm based on the sensitive area of Sigmoid function.

The sensitive area of the activation function proposed in equation (4) is introduced to the mean square error performance function:

$$\varsigma(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) - G_a(n), \quad (5)$$

where the neuron k is an output node.

The wider the sensitive area is, the stronger the generalization ability of the node is. The parameters in the parallel network are modified considering the distribution of sensitive areas and the anti-jamming ability of the network. Then excess weights, that is, the weights having little impact on performance and leading to overtraining, are removed in the pruning process using the second-order information of the error function and considering the complexity and the

training error performance. The final parameters then make it possible to minimize the growth of performance functions when removing them away, which are ideal compromises considering complexity and error performance, and they can further enhance generalization capabilities.

The training is actually a fitting process of nonlinear inputs and outputs, and the generalization of network can be seen as the nonlinear interpolation of the verification data. It is said that the network loses the ability to generalize in other samples when overfitting occurs. Numerous hidden units in a cascade network are likely to store too much noise effect. In that case, it is of necessity to use cross-validation.

Check the verification error in the corresponding parameters state under the set viewing cycle, and enter the next cycle if passing the validation. The training error converges along with the increase of training times, while the validation error monotonically decreases and then rises. The network starts to capture the noise information after crossing the minimum point. Hence, the state is regarded as the stopping criterion to reduce the occurrence of overfitting.

4. Coke Quality Prediction Model

4.1. Data Preprocessing. In general, the direct input of untreated data is not optimal. For example, the finite limit of the logistic activation function is (0, 1), but the sample value is tremendous compared with its finite limit, which results in the function being almost saturated and the training being stagnated. In addition, the backpropagation algorithm is similar to the LMS algorithm, of which the calculation time is heavily dependent on the condition number $\lambda_{\max}/\lambda_{\min}$. The $\lambda_{\max}/\lambda_{\min}$ of nonzero mean input is bigger than that of the zero-mean input.

Hence, it is required that mean value of the data on the entire training set is close to 0. The conventional processing is the mean centering and variance regulation. Let the input and output modes $A \in R^{n \times m}$, $C \in R^{p \times m}$ be arranged by columns, calculate the mean value of the n -th row of A and the p -th row of C , and subtract this mean value in each row. Calculate the variance of the n -th row in A and the p -th row in C , and divide the corresponding values in each row. The processing of the input matrix A and the output pattern matrix C should be synchronous.

4.2. Multilayer Feedforward Backpropagation Network. The parameters of mixed coal are selected as inputs and the coke quality as outputs in the FB model shown in Figure 3; cascade forward network can be used to solve more complex problems and improve the training precision, in which each subsequent layer is connected to the input layer and the adjacent one and the output is directly influenced by the input layer.

The excitation function is used to process the result of the summer, and the nonlinear function is generally used to maximize the efficiency of the network. In this paper, the logsig and tansig functions are used to perform the combination trial and error.

In the cascade forward backpropagation (CF) structure, it can be seen that the accuracy and fit effect that the training method can obtain are better than that of other training functions; hence, it is fixed as the training function in this paper.

The number of hidden nodes depends on the number of training samples, the size of the noise, and rules hidden in data. Commonly, it is best for the number of hidden nodes to be twice the number of input layers. It is illustrated by lots of tests carried out for 12–16 nodes in a single hidden layer that 16 nodes produce the best prediction result.

4.3. RBF. N samples are given in advance and the parameter r specifies the division between the training set ℓ' and the verification one ℓ'' . Then, there are $(1-r)N$ samples in ℓ' and rN samples in ℓ'' . It is a good choice that the parameter r remains at a fixed value of 0.2 according to the simulation experiment by Keams; that is, the training set is assigned 80% of the samples and the remaining samples are distributed in the validation set ℓ'' .

The number of input nodes depends on the dimension of the input vector x , written as m_0 .

The hidden layer size m_1 is determined by the number of planned clusters, controlling the network performance and the computational complexity.

The clustering mean $\hat{\mu}_j$ is obtained via the clustering algorithm and serves as the center of the basis function $\varphi(x_j)$ which is known as x_j .

$\sigma = d_{\max}/2K$ is the extended parameter, where d_{\max} denotes the maximum distance between the centers ensuring that the hidden layer unit is neither too sharp nor too flat.

5. Simulation Results and Analysis

800 sets of mixed coal quality parameters and their corresponding coke quality indexes are randomly selected from the 1000 sets of preprocessed data to train FB, CF, and RBF networks, respectively.

5.1. Results and Analysis of Training Results. RBF can meet the conditions required by the Cover theorem. The adjustment of weight and threshold is considered by functions newrbe and newrb in constructing the network, so there are no specific training and learning functions for the network. In that case, only the training results of FB and CF networks are compared through error curves in Figures 4 and 5:

As can be seen from the figure, the results of training, verification, and test of the two networks basically remain consistent. Under almost identical iteration number, the performance of cascade forward network reaches 0.01, showing a slightly better behavior than that of the forward network (0.05) with the training time being close, while the prediction errors of both networks are comparatively small on the whole.

5.2. Results and Analysis of Prediction Results. The remaining 200 sets of mixed coal quality parameters are

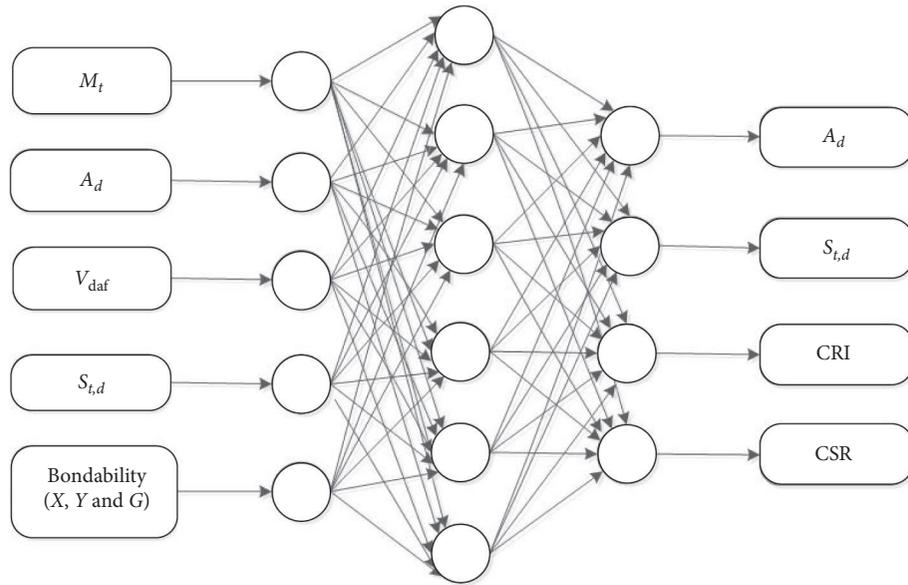


FIGURE 3: The structure of FB network.

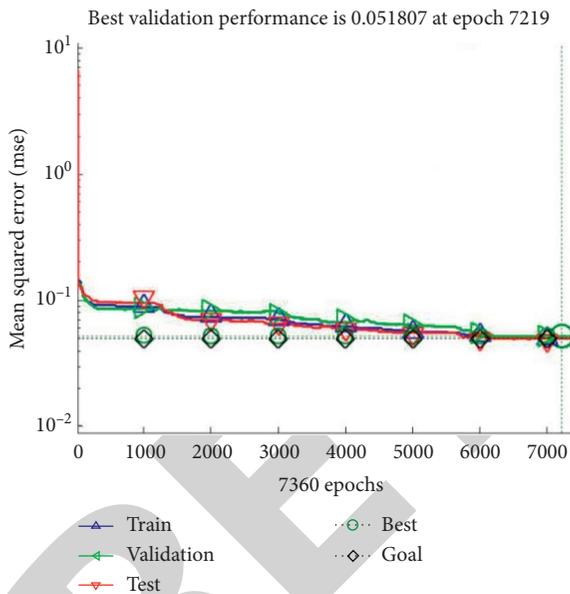


FIGURE 4: The training error curve of FB network.

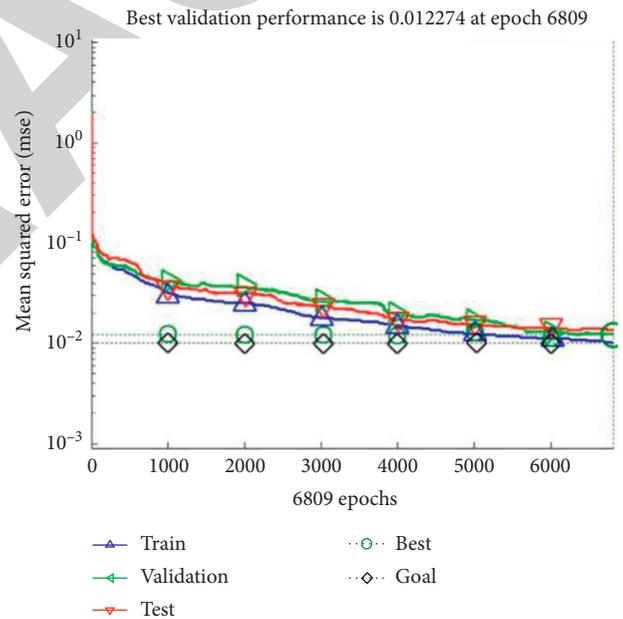


FIGURE 5: The training error curve of CF network.

utilized to predict the coke quality by the above trained networks, and predictions of ash, sulfur, CRI, and CSR are shown in Figures 5–8. Prediction errors of 5% and 10% represent different levels of accuracy required by the business.

The line with the symbol of “◁” in Figures 6–9 denotes the error margin of 10%, The line with the symbol of “▷” in Figures 6–9 denotes the error margin of –10%, the line with the symbol of “△” represents that of 5%, the line with the symbol of “▽” represents that of -5% and the line with the symbol of “○” represents that of 0%. The symbols of “x”, “-,” and “|” show the prediction results of the FB, CF, and RBF networks, respectively.

Only three sets of ash prediction error are more than 10% in the FB, CF, and RBF networks, while the error between 5% and 10% exists in 51, 29, and 48 sets, respectively. There are 13 and 12 sets in the FB and the RBF networks, respectively, while there is one set in the CF network with the prediction error being more than 8%. It can also be illustrated from Figure 6 that the predictive values of CF network are closer to the 0% error margin line as a whole compared to the results of FB and RBF networks. Hence, the conclusion obviously can be drawn that the CF network performs much better than the others in terms of ash prediction.

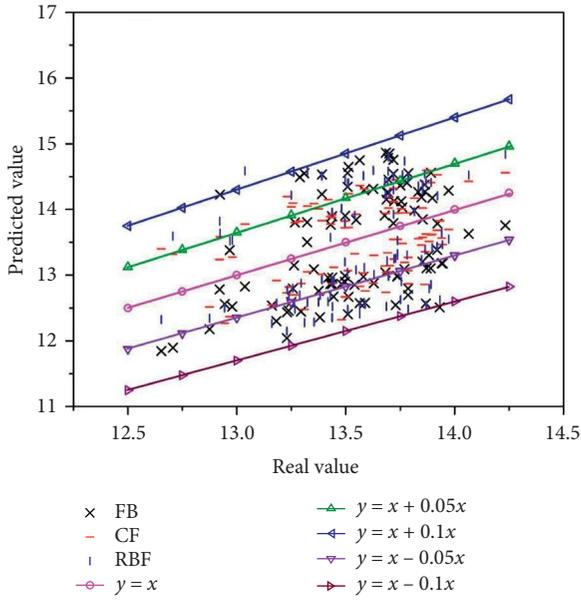
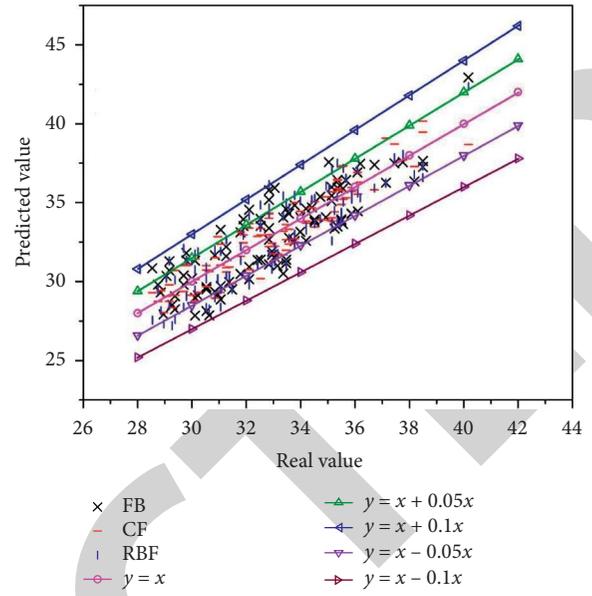
FIGURE 6: Comparison of prediction errors of A_d .

FIGURE 8: Comparison of prediction errors of CRI.

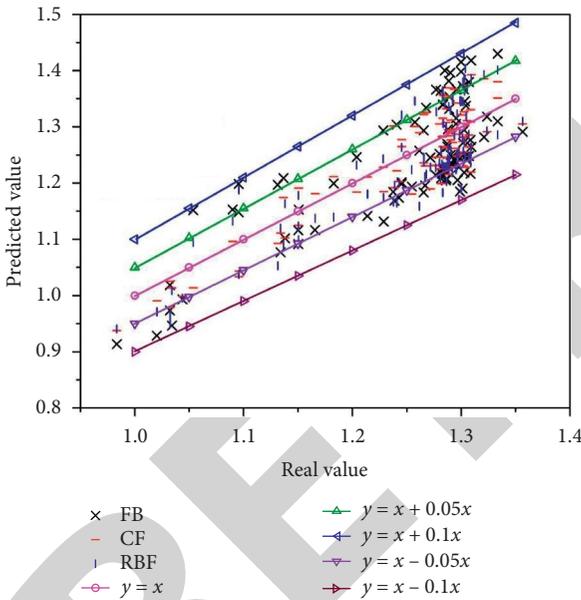
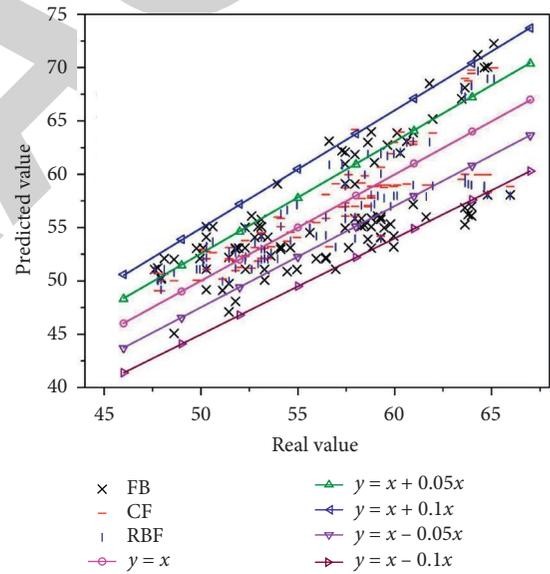
FIGURE 7: Comparison of prediction errors of $S_{r,d}$.

FIGURE 9: Comparison of prediction errors of CSR.

The prediction errors of sulfur content of three networks are all within 10%, and errors of 54, 17, and 27 sets are beyond 5%, respectively. The forecast of sulfur content is more excellent than that of ash and the RBF is better than BP for prediction. It can also be seen from Figure 7 that the number of the predictive values of CF network outside the 5% error margin line is significantly less than that of FB and RBF networks. The CF network behaves excellently in the completion of sulfur content prediction.

The errors are also controlled within 10% in the three networks as for the CRI index, 29, 10, and 32 sets of samples among them having the errors over 5%. The prediction effect of CRI is much better than that of the former indicators. It is shown that the number of the predictive values of FB and

RBF networks outside the 5% error margin line is significantly greater than that of CF network and a majority of errors of CF are within the 5% error margin line.

The prediction of FB network is not so good for CSR, with 10 sets having errors between 10% and 15%, while there are only 2 sets in two other networks with the same errors. With 62, 19, and 24 sets of errors over 5%, respectively, in the three networks, the overall prediction effect of CRI is worst among the predictions of the four properties. It is shown that most of the predictive values of FB network are distributed between the 5% and 10% error margin line, and the effect is comparatively poor; values of CF network near the yellow line are slightly more than that of RBF. Hence, the CF network is quite suitable for the prediction of the data structure selected in this paper.

It can be concluded in general that the prediction results for the four properties are satisfactory and the prediction accuracy is relatively high, with the best prediction effect for the CRI and the most disappointing predictive result for the CSR. The cascade forward network performs best among the three networks with most errors controlled within 5%, while the forward BP behaves the worst, as its most prediction errors vary from 5% to 10% and even over 10%.

6. Conclusions

Three networks for coke quality prediction were proposed based on the forward backpropagation network, the cascade forward backpropagation network, and the radial basis network and the corresponding results were compared, respectively. It is shown that the prediction of the CRI is closest to the actual values with errors within 10% in all three networks, better than that of the CSR with errors between 10% and 15% in 14 sets of samples. The prediction errors for the four properties are all within 15% and it can be said that the three networks all have a relatively high prediction accuracy. When it comes to the network structure, the cascade forward network behaves best with errors mostly controlled within 5%, which acts as an excellent guide during the coke production to some extent. It is the additional connection between each layer and the front ones that ensures the prediction accuracy. The closer the hidden layer is to the output layer, the greater the amount of information of effect on weights adjusting. And the traditional forward network is not suitable here as its most errors are between 5% and 10% and even over 10%. In this way, it is concluded that the cascade forward network matches the data structure selected in this paper.

However, the main budget constraint is computing time for large-scale samples, and the advantage of cascade networks with a large number of internal connections is likely to become no longer obvious; hence, large-scale learning problems need to be further investigated.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51974131), Science and Technology Project of Hebei Education Department (no. BJ2017021), NCST Natural Science Funds for Distinguished Young Scholars (no. JQ201711), Hebei Province Natural Science Fund for Excellent Young Scholars (no.

E2018209248), Hebei Provincial Postgraduate Demonstration Course Project in 2020 (no. KCJSX2020053), and NCST Project Establishment and Construction of Postgraduate Demonstration Course.

References

- [1] H. P. Tiwari, S. K. Haldar, A. Roy, A. Bhattacharjee, and P. Mishra, "Data mining—new perspectives on predicting coke quality in recovery stamp charged coke making process," *Metallurgical Research & Technology*, vol. 112, no. 6, p. 603, 2015.
- [2] X. Tang, S. Snowden, B. C. McLellan, and M. Höök, "Clean coal use in China: challenges and policy implications," *Energy Policy*, vol. 87, pp. 517–523, 2015.
- [3] H. V. Jagadish, J. Gehrke, A. Labrinidis et al., "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, 2014.
- [4] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and challenges of big data research," *Big Data Research*, vol. 2, no. 2, pp. 59–64, 2015.
- [5] S. C. Chelgani, S. S. Matin, and J. C. Hower, "Explaining relationships between coke quality index and coal properties by Random forest method," *Fuel*, vol. 182, pp. 754–760, 2016.
- [6] R. Roest, H. Lomas, K. Hockings, and M. R. Mahoney, "Fractographic approach to metallurgical coke failure analysis. Part 1: cokes of single coal origin," *Fuel*, vol. 180, pp. 785–793, 2016.
- [7] K. Shahzad, S. Kanwal, S. Nawaz, N. Sheikh, and S. Munir, "Effects of moisture and coal blending on the hardgrove grindability index of Pakistani coals," *International Journal of Coal Preparation and Utilization*, vol. 34, no. 1, pp. 1–9, 2014.
- [8] J. P. Mathews, V. Krishnamoorthy, E. Louw et al., "A review of the correlations of coal properties with elemental composition," *Fuel Processing Technology*, vol. 121, pp. 104–113, 2014.
- [9] L. Zhang, W. Liu, and D. Men, "Preparation and coking properties of coal maceral concentrates," *International Journal of Mining Science and Technology*, vol. 24, no. 1, pp. 93–98, 2014.
- [10] R. Morga, I. Jelonek, K. Kruszewska, and W. Szulik, "Relationships between quality of coals, resulting cokes, and micro-Raman spectral characteristics of these cokes," *International Journal of Coal Geology*, vol. 144–145, pp. 130–137, 2015.
- [11] B. D. Flores, A. G. Borrego, M. A. Diez et al., "How coke optical texture became a relevant tool for understanding coal blending and coke quality," *Fuel Processing Technology*, vol. 164, pp. 13–23, 2017.
- [12] M. A. Diez, R. Alvarez, and C. Barriocanal, "Coal for metallurgical coke production: predictions of coke quality and future requirements for coke making," *International Journal of Coal Geology*, vol. 50, no. 1–4, pp. 389–412, 2002.
- [13] C. Liu, Z. Xie, F. Sun, and L. Chen, "Exergy analysis and optimization of coking process," *Energy*, vol. 139, pp. 694–705, 2017.
- [14] J. Chen, G. Liu, H. Li, and B. Wu, "Mineralogical and geochemical responses of coal to igneous intrusion in the Pansan coal mine of the Huainan coalfield, Anhui, China," *International Journal of Coal Geology*, vol. 124, pp. 11–35, 2014.
- [15] S.-F. Zhang, L.-Y. Wen, K. Wang, C. Zou, and J. Xu, "Effects of additives on sulfur transformation, crystallite structure and properties of coke during coking of high-sulfur coal," *Journal of Iron and Steel Research International*, vol. 22, no. 10, pp. 897–904, 2015.

- [16] R. M. Garbarino and R. T. Tonti, "Desulfurization and its effect on calcined coke properties," in *Essential Readings in Light Metals*, pp. 119–122, Springer International Publishing, Cham, Switzerland, 2016.
- [17] S. S. Raaj, S. Arumugam, M. Muthukrishnan, S. Krishnamoorthy, and N. Anantharaman, "Characterization of coal blends for effective utilization in thermal power plants," *Applied Thermal Engineering*, vol. 102, pp. 9–16, 2016.
- [18] R. Loison, P. Foch, and A. Boyer, *Coke: Quality and Production*, Elsevier, Amsterdam, Netherlands, 2014.
- [19] A. Ahmadpour, N. Jahanshahi, S. Rashidi, N. Chenarani, and M. J. D. Mahboub, "Application of artificial neural networks and adaptive neuro-fuzzy inference systems to predict activated carbon properties for methane storage," *Adsorption Science & Technology*, vol. 32, no. 4, pp. 275–290, 2014.
- [20] L. C. Edwards, K. J. Neyrey, and L. P. Lossius, "A review of coke and anode desulfurization. Essential readings in light metals," *Electrode Technology for Aluminum Production*, vol. Volume 4, pp. 130–135, 2013.
- [21] S. S. Matin and S. C. Chelgani, "Estimation of coal gross calorific value based on various analyses by random forest method," *Fuel*, vol. 177, pp. 274–278, 2016.
- [22] S. S. Haykin, *Neural Networks and Learning Machines*, Pearson, Upper Saddle River, NJ, USA, 2009.
- [23] M. R. Patil and P. Sawant, "Corporate social responsibility, performance and sustainability reporting of shree cement company in India: a case study," *Journal of High Technology Management Research*, vol. 3, no. 4, pp. 74–87, 2014.
- [24] E. Hadavandi, J. C. Hower, and S. C. Chelgani, "Modeling of gross calorific value based on coal properties by support vector regression method," *Modeling Earth Systems and Environment*, vol. 3, no. 1, p. 37, 2017.
- [25] L. Ljung, "System identification," in *Signal Analysis and Prediction*, pp. 163–173, Birkhäuser Boston, Basel, Switzerland, 1998.