

Application of BP Neural Network Ensemble Model Based on Bagging Algorithm

Jianxing Liao, Junwei Lv, Yuhan Shao, Peipei Li, and Xuegang Hu

Abstract—The latest coal mine policy of China puts forward new requirements for safe mining in the coalmine. This paper proposes a BP neural network ensemble learning model based on additional momentum term combined with Bagging algorithm for the prediction of coal and gas outburst. The additional momentum term approach is used to optimize the weights and thresholds of BP neural network during the training process. In order to enhance the prediction ability of BP neural network, the Bagging algorithm is adopted to build ensemble neural network. Experiments show that as compared with the standard BP neural network, the ensemble learning model can improve the prediction accuracy of 8%-10% points in the prediction of coal and gas outburst and the prediction accuracy is over 95%.

Index Terms—BP neural network, Bagging algorithm, ensemble learning model, additional momentum terms, coal and gas outburst prediction.

I. INTRODUCTION

Coal and gas outburst (outburst for short) is a major natural disaster in the coal mining process. It is a complicated gas dynamic phenomenon involving coal, rock and gas (individual with carbon dioxide), dominates in coal mine accidents and poses a serious threat to the safe production of coal mines [1]. According to the statistics of coal mine safety accidents from 2005 to 2016, collected by China's State Bureau of Coal Mine Safety Supervision, the number of outburst accidents accounts for 25.5% while deaths accounts for 47.7%, which have caused serious loss of human lives and properties. Therefore, it is becoming one of the focuses of the current research that how to predict outburst effectively.

To predict the outburst, existing methods are mainly divided into regional prediction and working face prediction. The main task of regional prediction is to predict the outburst danger in coal seam area, while the prediction of working face is based on the former, predicting the outburst danger of coal mass near the working face, and carrying out in the process of extraction face [2]. According to the prediction process and continuity, the working face prediction can be further divided into static (or discontinuous) and dynamic (or continuous) categories [3]. The works in [3], [4] are introduced in detail.

Manuscript received December 13, 2018; revised February 28, 2019. This work was supported by The National Key Research and Development Program of China (Grant No. 2016YFC0801406).

Jianxing Liao, Junwei Lv, Yuhan Shao, and Peipei Li are with Hefei University of Technology, Hefei 230601, China (e-mail: ljx4805@163.com, junweilv@hfut.edu.cn, yuhan_shao@yeah.net, peipeili@hfut.edu.cn).

Xuegang Hu is with the Research Association of Computer Education in colleges and University of Anhui Province, and School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China (e-mail: jsjxhuxg@hfut.com).

At present, existing static prediction methods are time consuming and have the trait of inaccurate predictions and susceptibility to artificial influences. The dynamic prediction technology is high costly, and there is still a gap between the reliability of the prediction and the need for the actual production. In view of the shortage of the current prediction methods, this paper introduces the BP neural network ensemble model based on bagging algorithm and improves the BP algorithm with the surplus momentum term method. With the help of the improved BP neural network's highly nonlinear mapping function and the ensemble learning advantage of the Bagging algorithm, the proposed method can investigate the correlation between the outburst and the impact factors hidden in the outburst data to compensate for the inadequacies of traditional predicting methods, and achieve the prediction of working face outburst danger more accurately and reliably.

The rest of the paper is structured as follows: Section II describes the related work; Section III introduces the BP neural network ensemble model based on Bagging algorithm in detail; Section IV gives the experiments and results, and the conclusions are shown in Section V.

II. RELATED WORK

With the deep understanding of the outburst process and its impact factors, researchers have put forward many outburst prediction methods. Existing works in the prediction of working face outburst can be divided into two types: static prediction and dynamic prediction [3], [4].

The static prediction methods are based on some quantitative indexes that affect the properties of gas coal mass and its occurrence conditions. More specifically, there are many indexes used in recent works, such as *drilling cuttings volume S*, *drilling gas inrush initial velocity q*, *initial speed of gas diffusion*, *coal solid coefficient f*, *gas pressure p* and so on^[3]. Static prediction methods include *D and K comprehensive index method*, *R index method*, *drilling gas inrush initial velocity method*, *drill cuttings gas desorption index method*, *drill cuttings comprehensive index method* and so on. The *D and K comprehensive index method* was first proposed by the Fushun branch of the Coal Science Research Institute, which is suitable for the regional prediction of coal seam or the prediction of the outburst danger of cross-measure cutting coal working face [5]. The *R index method* was proposed by the former Soviet Union Eastern Coal Research Institute in 1969, and it is applicable to the outburst danger prediction of coal roadway driving working face [3], [6]. The *drilling gas inrush initial velocity method* is an important index used by coal and gas scientists and

technicians at home and abroad, which is the most widely used daily prediction method in the former Soviet Union [3]. The *drill cuttings gas desorption index method* is widely used because it can reflect the gas content of coal seam better and can be measured easily in the underground, and its determination method was drafted by Chongqing Research Institute of Coal Science Research Institute [7]. The *drill cuttings comprehensive index method* predicts the outburst danger of working face by considering the *maximum drilling chip amount per meter* S_{\max} , *drill cuttings gas desorption index* K_1 , *drill cuttings gas desorption attenuation coefficient* C and *desorption meter reading* Δh_2 at 2min comprehensively.

Dynamic prediction method refers to the method of determining the danger of coal seam near the working face by means of dynamically and continuously monitoring of some indexes which can comprehensively reflect the stress (or deformation) state of the coal containing gas [3], including *Elastic wave technology prediction method*, *gas inrush dynamic index prediction method*, *electromagnetic radiation technology prediction method* and *coal seam temperature prediction method* and so on. The Fushun Branch of China Coal Research Institute has done a lot of work in the prediction of acoustic emission (a type of elastic wave technology) [8]. And countries such as the former Soviet Union, Canada, and France have also conducted some research work on this technology. Gas emission dynamic indexes include V_{30} and K_V . V_{30} refers to the ratio of the amount of gas emission and the amount of coal lost within 30 minutes after blasting on the working face. K_V refers to the coefficient of gas emission fluctuation, which reflects the extent of the increase and decrease of gas emission from the working face [3]. At present, several major coal science branches in our country have studied this technology, where the future research direction is to identify outburst based on the relationship between outburst and gas dynamic emission monitored continuously by environmental monitoring system [3]. *Elastic wave technology* is utilized to predict the outburst based on the principle of outward radiation of electromagnetic energy in the process of loaded coal and rock mass deformation and fracture [9]. China and the former Soviet Union studied earlier in this respect. The coal seam will present different temperature changes due to the effects of ground stress and gas pressure. It would occur precursory phenomena of outburst, such as crown drill, stuck drill, and spray drill during drilling, when the coal mass accumulated high strain energy. The Detailed Regulations on Coal and Gas Outburst Prevention formulated by China have made the coal seam temperature as a prediction index for outburst danger [10].

In recent years, some new theoretical approaches and techniques have been applied in the prediction of outburst danger, including fractal theory, nonlinear theory, rheology-mutation theory and data mining technology. For example, Zhengli Mao used the fractal theory to predict the outburst [11]; Wang Kai *et al.* applied the non-linear theory to the outburst research [12]; Simultaneously, rheological and mutation theory can also be used for outburst research [13], [14]. The application of data mining technology in prediction is becoming more and more extensive. Jiahui Liang *et al.*

applied fuzzy mathematics method to carry out a multi-level fuzzy comprehensive evaluation on outburst prediction [15]; Zhonghui Li *et al.* used the logistic regression model to evaluate outburst comprehensively [16]; Wang, ZY used the support vector machine method to study the prediction of outburst [17]; and Zhang Yu *et al.* proposed a method based on neural network to establish an outburst prediction model [18].

Contrary to the above methods, the BP neural network ensemble model based on Bagging algorithm proposed in this paper combines the advantages of the static and the dynamic prediction method, it can find out the correlation between the outburst and the impact factors hidden in the outburst data to compensate for the inadequacies of traditional predicting methods with the help of the neural network's highly nonlinear mapping function and the ensemble learning advantage of the machine learning. Furthermore, it can be combined with dynamic predicting methods to enrich the prediction model and its functions, for instance, the dynamic prediction method includes electromagnetic radiation method, of which input as a prediction index to the model to make a dynamical and continuous prediction of working face. It aims to achieve the prediction of working face outburst danger more accurately and reliably.

III. BP NEURAL NETWORK ENSEMBLE LEARNING MODEL BASED ON BAGGING ALGORITHM

The work of this paper, inspired by [18]-[20], utilizes the BP neural network ensemble model based on Bagging algorithm to predict the outburst danger in the working face.

BP neural network is one of the most widely used neural network models, which can classify prediction problems and possess highly nonlinear mapping and accurate function fitting, as well as the ability of self-organizing, self-learning, self-adaptive and parallel processing [21]. Bagging algorithm is an ensemble learning algorithm, which is an ensemble method to improve the accuracy of learning algorithms by sampling samples and feature attributes to generate multiple independent models [22]. The problem in this paper can be formally described as follows: Given coal and gas outburst data set D containing m data records $\{(x_i, y_i)_{i=1}^m\}$, where $x_i \in R^d$ indicates a vector sample with d dimension features, $y_i \in Y = \{c_1, c_2, \dots, c_v\}$ indicates the category label, v indicates the number of predicted labels. As shown in Fig. 1, firstly, it is used to divide the data set D into the training set DT and the test set DS by hold-out cross validation [23] to construct the improved BP neural network ensemble learning model. Then k -fold cross validation [23] is used to divide the training set DT into k disjoint subsets and take $k-1$ as the training data set D_i^t , one as the verification data set D_i^v and the test set DS as the test data set D_i^t . Secondly, Bootstrap, a sampling with replacement method, is adopted to sampling l_N samples randomly from the training data set D_i^t each time to train the classifier and determine the model parameters. Thirdly, the classifier is verified by the validation data set D_i^v and a base classifier model E_j of an improved BP neural

network is obtained, subsequently, the same method is used to form the ensemble classifier $E = \{E_1, E_2, \dots, E_k\}$. Finally, the ensemble classifier E is applied to the test data set D_i^t . The work uses the weighted mean method to obtain a strong classifier for testing the classification of unknown samples in the test set. In sum, the problem in this paper is to train an improved BP neural network ensemble learning model on data set D . This model can be used to predict the outburst danger of unknown samples, which can be formalized as below:

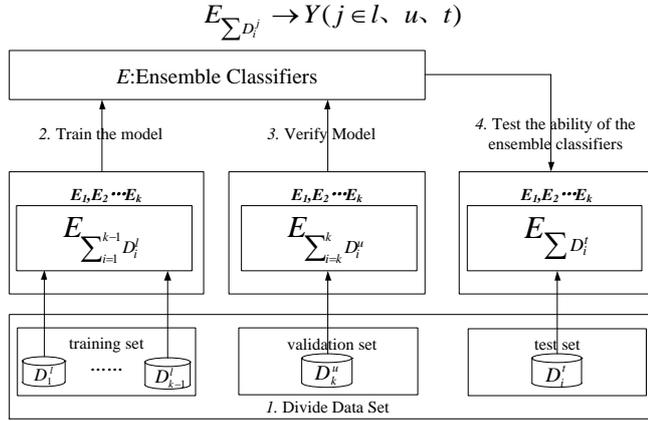


Fig. 1. Framework of the ensemble learning model.

A. Improved BP Neural Network Model

In this paper, the BP neural network is improved with the additional momentum term method. The essence of the additional momentum term is the effect, the change of the weights or the thresholds value, have been passed through by a momentum factor. If the momentum factor is 0, the change of the weights or thresholds is generated by the gradient descent method; While if the momentum factor is 1, the change of the gradient descent method is ignored and the new weight or threshold change is the last one. In this way, when the momentum term is added, the adjustment of the weight value is changed to the average direction of the bottom of the error surface, and the "inertia effect" is used to restrain the possible oscillation in the network training, which plays a buffer and smooth function, and helps to jump out the network from the local minimum of the error surface.

The basic processing framework of the improved BP neural network is shown in Fig. 2. Where $X = (X_1, X_2, \dots, X_n)$ is the set of n values that are input from the outside or other neurons output; $W = (W_1, W_2, \dots, W_n)$ is called the weight, representing the connection strength between the neuron and other neurons; $\sum WX$ is called the activation value that is equal to the total input of the artificial neuron; O refers to the output of the neuron; b refers to the threshold of this neuron, and if the weighted sum of the input signal is greater than b , the artificial neuron is activated. In this way, the output of artificial neuron can be described as below:

$$O = f(\sum WX - b) \quad (1)$$

In Eq.(1), $f(\bullet)$ is called the activation function. The activation function used in this paper is a nonlinear transformation function-bipolar Sigmoid function ($\tanh(x)$

function). In the process of error back-propagation, the problem of derivation with respect to the activation function is involved, and the $\tanh(x)$ function solves the problem of derivative discontinuity and the output problem of zero-centered effectively, so it is used as the activation function of this paper. It is defined as follows:

$$f(x) = (1 - e^{-x}) / (1 + e^{-x}) \quad (2)$$

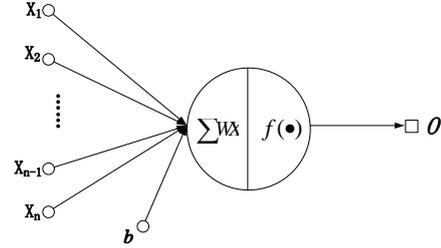


Fig. 2. Artificial neuron model.

Robert Hecht-Nielsen proved that any continuous function in any interval can be approximated by a hidden layer neural network, which is the *universal approximation theorem* [24]. For the convenience of the following discussion, symbols to be used in this paper are defined in Table I.

This paper uses a three-layer BP neural network with a single hidden layer structure to simulate the change of the outburst. The network structure is illustrated in Fig. 3.

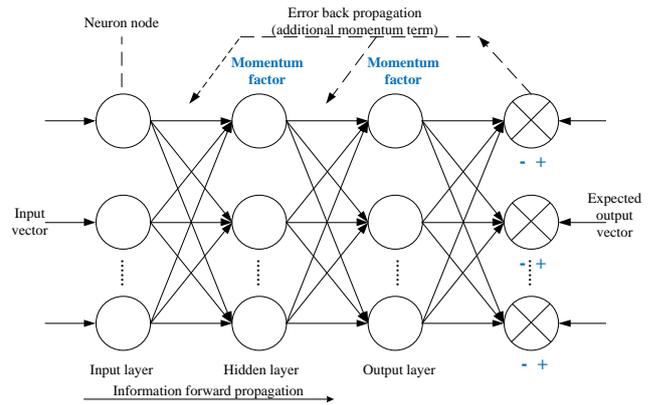


Fig. 3. Improved BP neural network structure.

As the momentum factor acts in the process of error back propagation, the new weights or thresholds are canceled when the updated weights or thresholds make the error too large so that they can avoid the network entering a larger error surface. When the new rate of error change exceeds the maximum rate previous, the calculated value of weights or thresholds is canceled. The maximum rate of error change can be any value, which is greater than or equal to 1, and the empirical value is 1.04. Hence, the program design based on the additional momentum method must add conditions to update the weights and thresholds correctly. The judging conditions are defined as follows:

$$\alpha = \begin{cases} 0 & E(k) > E(k-1) * 1.04 \\ 0.95 & E(k) < E(k-1) \\ \alpha & \text{others} \end{cases}$$

B. The Principle of Improved BP Neural Network Learning Algorithm

The improved BP neural network algorithm adopts the

δ (Delta) learning rule, which belongs to the tutor learning algorithm [25], and completes the input to output mapping through a process that minimizes the loss function. This

process is used to adjust the connection weights and thresholds among the neurons.

TABLE I. SYMBOL DESCRIPTION OF IMPROVED BPNN LEARNING ALGORITHM

Symbol	Description	Symbol	Description
x_i	The input of the i th node of the input layer	b_j^1	The threshold of the j th node of the input layer
w_{ij}^1	The weight of the i th node of the input layer to the j th node of the hidden layer	b_j^2	The threshold of the j th node of the hidden layer
s_j^1	The input of the j th node of the hidden layer	$L(e)$	The loss function
$\theta(\bullet)$	The activation function	δ_i^1	The reverse output value of the i th node of the output layer
w_{ij}^2	The weight of the i th node of the hidden layer to the j th node of the output layer	δ_i^2	The reverse output value of the i th node of the hidden layer
s_j^2	The input of the j th node of the output layer	n	The number of input layer neuron nodes
$\overline{y_j}$	The actual output of the j th node of the output layer	h	The number of hidden layer neuron nodes
y_j	The expected output of the j th node of the output layer	p	The number of output layer neuron nodes
$E(k)$	The error of Output layer	α	momentum factor

Let the *learning rate* be η and the *momentum factor* be α . The input and output of each network layer can be calculated as follows.

The loss function is defined below:

$$L(e) = 1/2SSE = 1/2 \sum_{j=0}^p e_j^2 = 1/2 \sum_{j=0}^p (\overline{y_j} - y_j)^2 \quad (3)$$

The calculation of the input of the hidden layer node:

$$s_j^1 = \sum_{i=1}^n x_i \cdot w_{ij}^1 - x_0 \cdot w_{0j}^1 = \sum_{i=1}^n x_i \cdot w_{ij}^1 + b_j^1 \quad (4)$$

The calculation of the input of the output layer node:

$$s_i^2 = \sum_{j=1}^h \theta(s_j^1) \cdot w_{ji}^2 + b_i^2 \quad (5)$$

The calculation of the reverse output of the output layer:

$$\delta_i^2 = \frac{\partial L}{\partial s_i^2} = \frac{\partial \sum_{j=1}^p \frac{1}{2} (\overline{y_j} - y_j)^2}{\partial s_i^2} = (\overline{y_j} - y_j) \cdot \frac{\partial y_j}{\partial s_i^2} = e_i \cdot \theta'(s_i^2) \quad (6)$$

The calculation of the reverse input of the hidden layer:

$$\delta_j^1 = \partial L / \partial s_j^1 = \theta'(s_j^1) \cdot \sum_{i=1}^p \delta_i^2 \cdot w_{ji}^2 \quad (7)$$

The additional momentum term is used to update the weights and thresholds, which can be calculated according to the following Eq. (8)-Eq. (11).

Update the weights from the input layer to the hidden layer:

$$\Delta w_{ij}^1(t) = (1-\alpha) \cdot \eta \cdot x_i \cdot \delta_j^1 + \alpha \cdot \Delta w_{ij}^1(t-1) \quad (8)$$

Update the thresholds of the hidden layer:

$$b_j^1(t) = (1-\alpha) \cdot \eta \cdot \delta_j^1 + \alpha \cdot \Delta b_j^1(t-1) \quad (9)$$

Update the weights from the hidden layer to the output layer:

$$w_{ij}^2(t) = (1-\alpha) \cdot \eta \cdot \theta(s_j^1) \cdot \delta_i^2 + \alpha \cdot \Delta w_{ij}^2(t-1) \quad (10)$$

Update the thresholds of the output layer:

$$b_j^2(t) = (1-\alpha) \cdot \eta \cdot \delta_j^2 + \alpha \cdot \Delta b_j^2(t-1) \quad (11)$$

The algorithm flow of the improved BP neural network is shown in Algorithm 1.

Algorithm 1. The algorithm flow of the improved BP neural network

Input: The normalized data set D after normalized by min-max function.

Output: The improved BP neural network with connection weight and threshold determination.

1: Network initialization: Initializing of the initial value of each weight and value $W_{ji}(0), b_j(0) \in (0,1)$, learning rate η and momentum factor α , number of nodes in the input layer n , number of nodes in the hidden layer h , number of nodes in the output layer p .

2: for all $(x_i, y_i) \in D$ do

3: Calculate the input and output value of each network layer.

3.1: Calculate the input value of the hidden layer nodes by Eq. (4);

3.2: Calculate the input value of the output layer nodes by Eq. (5);

3.3: Calculate the reverse output value of the output layer nodes by Eq. (6);

3.4: Calculate the reverse input value of the hidden layer nodes by Eq. (7);

4: Update weights and thresholds based on additional momentum term.

4.1: Updating the weights from the input layer to the hidden layer by Eq. (8);

4.2: Update the thresholds of the hidden layer nodes by Eq. (9);

4.3: Update the weights from the hidden layer to the output layer by Eq. (10);

4.4: Update the thresholds of the output layer nodes by Eq. (11);

5: After training all the samples, does the output layer error $E(k)$ meet the accuracy requirements? $E(k)$ meet the accuracy requirements, turn (6), otherwise turn (3) to continue training.

end for

C. Classification Prediction by BP Neural Network Ensemble Model Based on Bagging Algorithm

Leo Breiman proposed Bagging algorithm in 1996 [20] and the deviation variance theory is used to make a complete and reasonable theoretical explanation of its effectiveness [26], [27]. The basic idea of Bagging algorithm is below: Given a weak classifier and a training set, the weak classification algorithm is used to classify the training samples. Because the weak classification algorithm is unstable and the accuracy is low, the weak classification algorithm is used many times to construct a sequence of prediction function. Then they are combined into a predictive function in a certain way. The final predictive result will be obtained by the predictive function

according to an ensemble learning method. The general methods include the voting method and the weighted average method.

Breiman points out that the neural network is unstable, i.e. a weak classifier, and Bagging has a significant effect on the unstable learning algorithm. Therefore, this paper uses the improved BP neural network as a weak classifier and the idea of Bagging ensemble to improve the generalization accuracy of the improved neural network. Fig. 4 shows the diagram of the improved ensemble BP neural network prediction model based on Bagging algorithm.

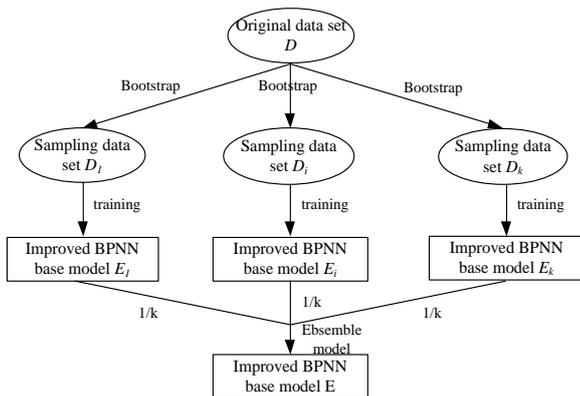


Fig. 4. Improved ensemble BP neural network prediction model based on Bagging algorithm.

Firstly, the improved BP neural network prediction model based on Bagging algorithm generates each training subsets D_1, D_2, \dots, D_k by Bootstrap sampling method (the uniformly sampling with replacement [28]) on the original data D , subsequently, generating an improved BP neural network base model E_i on each training subsets. The model E_i makes the sample instance in the corresponding training set D_i obtain an optimal prediction accuracy. In this way, it is more likely that the sample instances of test set, similar with the training set D_i sample instances, have higher accuracy on the model E_i . Details of the algorithm is shown in Algorithm 2.

IV. EXPERIMENTS AND RESULTS

A. Experimental Data

In this paper, 58 sets of real data were obtained from a coal mine gas monitoring system in Anhui Province, and 142 sets of simulation data were generated on the basis of 58 sets of real data. A total of 200 sets of data are used for our experiments.

200 sets of data consist of the following 7 features: 1. Initial speed of gas diffusion (mmHg); 2. Coal seam gas pressure (Mpa); 3. Coal seam gas content (m3/t); 4. Coal solid coefficient; 5. Coal destruction type; 6. Geological structure type; 7. Coal mining depth (m). The specific classification method is based on the results that have been detected out to classify the outburst danger level, which can be divided into four categories: no outburst danger; small outburst danger; medium outburst danger and large outburst danger. The Coal destruction type are classified into five

categories: I, II, III, IV and V, corresponding to 0.1, 0.2, 0.3, 0.4, and 0.5 in the data, respectively. The Geological structure types include horizontal structure; monoclinical structure; anticline fold structure; syncline fold structure; ascending rock type fault structure and descending rock type fault structure, corresponding to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 in the data, respectively. We give eight sets of sampling data as showed in Table II.

Algorithm 2. The algorithm flow of Bagging algorithm integrated BP neural network

Input: The normalized data set $D = \{(x_i, y_i)_{i=1}^m\}$.

Output: The classification result of this sample x of this test set

1: for $t=1, \dots, k$ do (k is the number of base models)

1.1: Randomly select $k-1$ samples from the training set DT (sampling with replacement);

1.2: Training the improved BP neural network with this samples to obtain the base model E_i ;

2: Averaging all the base model $\{E_i\}_{i=1}^k$ to get a strong learner

$$F(x) \leftarrow 1/b \sum_{i=1}^k E_i(x);$$

3: The strong learner is used to test the data set.

end for

B. The settings of Improved BP Neural Network

In order to make the network converge quickly and avoid numerical problems and neuron output saturation, the training set D is normalized before training. Since bipolar S type activation functions are used in this paper and the input parameters are required to be distributed in $[-1, 1]$, so the data need to be normalized to $[-1, 1]$.

The topological structure of the BP neural network improved by additional momentum term is composed of an input layer, a hidden layer and an output layer. There are 7 neurons in the input layer, that is, 7 features. The output layer neuron node is 4, representing 4 categories, of which (1 0 0 0) indicates no outburst danger, (0 1 0 0) indicates small outburst danger, (0 0 1 0) indicates medium outburst danger and (0 0 0 1) indicates large outburst danger. The number of hidden layer nodes is determined by the empirical value namely $h = \sqrt{n+p} + l$, where h is the number of hidden layer nodes, n is the number of input layer nodes, p is the constant between 0 and 10, and the number of hidden layer nodes is 11 in this paper. The activation function is the bipolar S function ($\tanh(x)$ function). The training approach is gradient descent method with momentum term by initializing weights and thresholds randomly. The maximum number of iterations is 1000, the prediction error is 0.001, and the learning rate and momentum factor are given in Section IV.C.

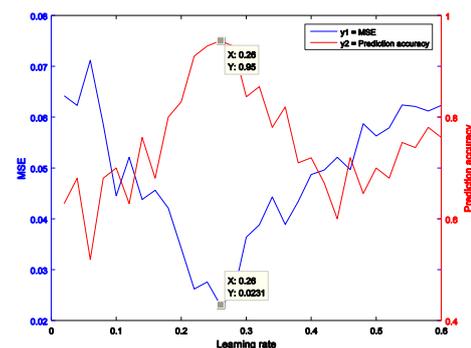


Fig. 5. Relationship between learning rate and forecast accuracy.

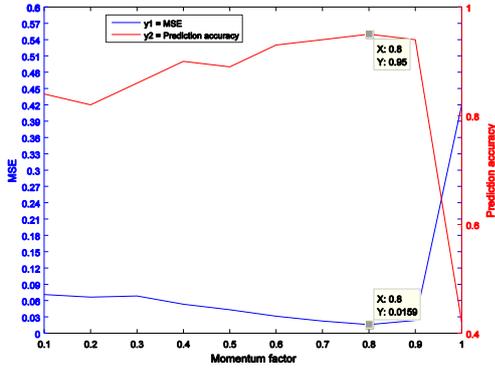


Fig. 6. Relationship between momentum term and forecast accuracy.

C. Results

In order to make the model more stable and the network training period appropriate and the effect convergent,

experiments on the learning rate and momentum factor are first given, the results are shown in Fig. 5 and Fig. 6.

In the experiments, different values of learning rate and momentum factor used in the neural network are simulated many times. It is seen from Fig. 5 and Fig. 6 that a stable high prediction result can be achieved if learning rate is set to $\eta \in (0.22, 0.28)$ and the momentum term is set to $\alpha \in (0.7, 0.9)$. This paper selects the learning rate $\eta = 0.25$ and the momentum factor $\alpha = 0.8$ to conduct subsequent experiments. Table III shows the overall mean square error (MSE), which represents the difference of the mean square error between the actual output value and the target, and also includes the regression value (R). It measures the correlation between the output and the target value for training, verification, and testing of the samples. Detailed analysis is presented in Fig. 7 and Fig. 8.

TABLE II: EXPERIMENTAL SAMPLE DATA

Initial speed of gas diffusion /mmHg	Coal seam gas pressure /Mpa	Coal seam gas content /m ³ /t	Coal solid coefficient	Coal destruction type	Geological structure type	Coal mining depth /m	The outburst danger level
7.4	0.68	7.739	0.93	0.2	0.1	122	No
7.6	0.69	7.186	0.92	0.2	0.2	154	No
12.7	0.92	9.681	0.46	0.3	0.1	329	Small
14.8	0.94	10.618	0.49	0.4	0.1	256	Small
20.7	1.18	15.376	0.38	0.4	0.3	250	Medium
21.2	1.27	17.405	0.36	0.4	0.4	303	Medium
26.8	1.63	22.866	0.23	0.5	0.5	289	Large
29.7	1.50	18.571	0.23	0.5	0.4	227	Large

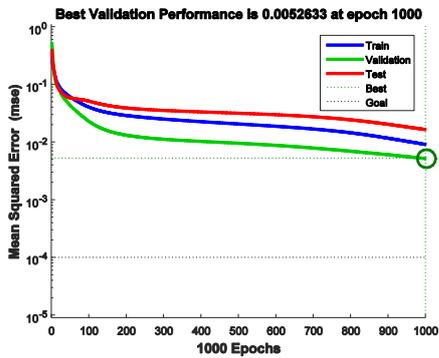


Fig.7. Mean square error varying with the number of steps.

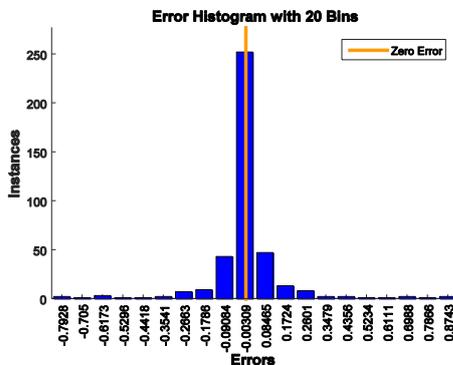


Fig.8. Error histogram.

From the analysis of Fig. 7 and Fig. 8, the test set error and the validation set error are similar, which indicates that the MSE rapidly decreases with time during the training period. And the error is close to the center line of Zero Error, so the model is robust.

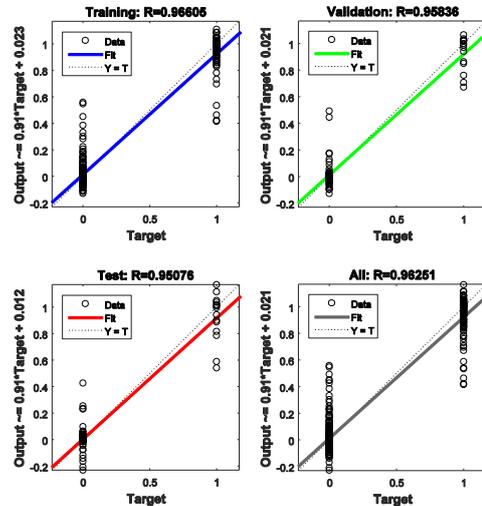


Fig. 9. Regression graph training in Ensemble learning model.

Furthermore, we used the standard BP algorithm to train the network. The comparison results are shown in Fig. 9, Fig. 10, Fig. 11, and Fig. 12. Fig. 9 and Fig. 10 are the training regression graphs. Results show that the improved BP neural network ensemble learning model based on Bagging algorithm presents a good curve fitting effect. The training, test and verification results are all greater than 0.95, which gives the optimal value of our model. Fig. 11 and Fig. 12 are the fitting graphs of the prediction trend and the original tendency. It can be seen that the ensemble learning model of this paper have a significantly better predicting trend than the standard BP neural network. Table IV shows the comparison training results using the standard BP neural network and the

ensemble learning model. Results show that the proposed ensemble learning model can improve the prediction accuracy of 8%-10% points and the prediction accuracy is over 95%.

TABLE III: MEAN SQUARE ERROR AND REGRESSION VALUE

Data set	Samples	Mean Square Error MSE	R
Training	100	0.01463	0.9551
Validation	50	0.01264	0.97322
Testing	100	0.01826	0.93487

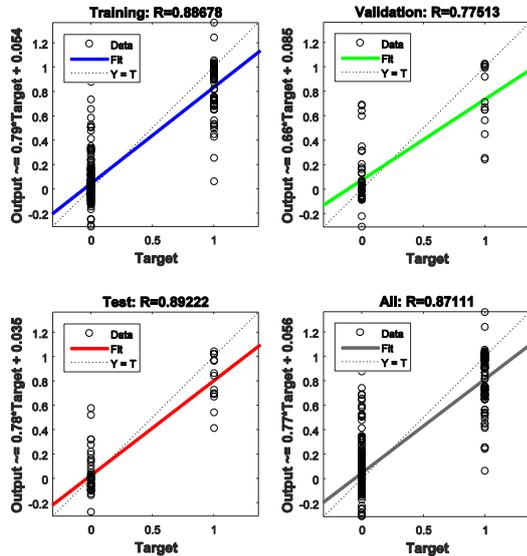


Fig. 10. Regression graph training in standard BP neural network.

TABLE IV: COMPARISON OF TRAINING RESULTS BETWEEN STANDARD BP NEURAL NETWORK AND THE ENSEMBLE LEARNING MODEL

	Mean Square Error MSE	R	Prediction accuracy		
			Training	Validation	Testing
Standard BP neural network	0.0648	0.8798	0.79274	0.88172	86%
The ensemble learning model of this paper	0.0152	0.9551	0.97322	0.93487	95%

V. CONCLUSION

As the outburst is affected by many factors, and the influence principle, the influence mode and the influence range are difficult to be processed quantitatively, thus, the neural network model with highly nonlinear mapping function has played an obvious advantage. The improved BP neural network prediction method proposed in this paper outperforms the traditional prediction method. For example, the solid theoretical basis and rigorous deduction process are more suitable for the comprehensive analysis of outburst compared to most of the baseline methods.

This paper combines Bagging algorithm to construct an ensemble learning model for the improved BP neural network. Compared with a single learner, this ensemble learning model has more significant generalization performance. It is more stable and reliable than the standard BP neural network. Training results are more effective and the prediction accuracy is improved.

The prediction method proposed in this paper is a supplement to the existing prediction method of outburst. It can be combined with other predicting methods to enrich the prediction model and its functions, for example, the dynamic and continuous outburst prediction method, such as *electromagnetic radiation technology prediction method*,

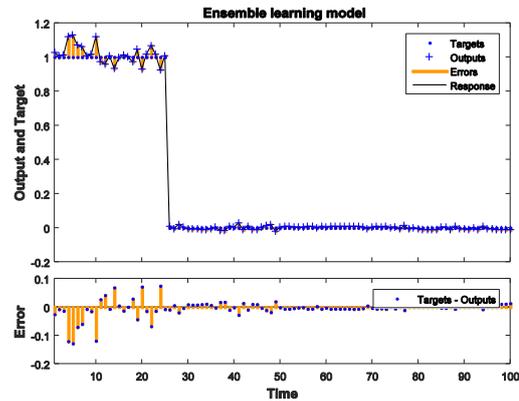


Fig. 11. Comparison between ensemble learning model prediction trend and original trend.

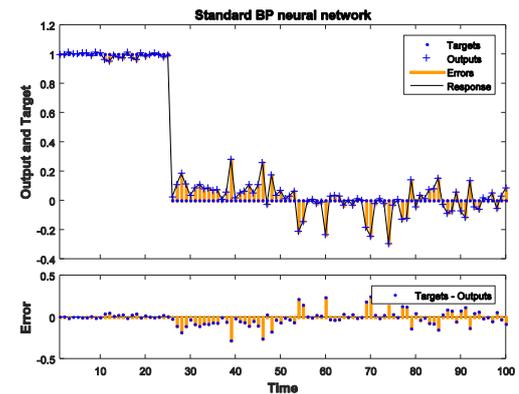


Fig. 12. Comparison between standard BP neural network prediction trend and original trend.

uses a predictive index as an input in the model to make a dynamic and continuous prediction of the working face.

In our studies, there are still some limitations. On one hand, real data are more difficult to obtain and the number of cases selected are not rich enough, which restricts the function of the neural network and makes the established network less stable. On the other hand, although the model established in this paper has a high accuracy, the actual application conditions are more complicated. If the coal mine establishes a multi-index outburst prediction learning sample database, it will undoubtedly make the model more widely and reliably be applied in practice.

REFERENCES

- [1] S. He, L. Jin *et al.*, "Soft coal solid-gas coupling similar material for coal and gas outburst simulation tests," *Journal of Geophysics and Engineering*, vol. 15, no. 5, p. 2033, 2018.
- [2] State Administration of Work Safety, Provisions for Prevention of Coal and Gas Outburst, China Coal Industry Publishing House, 2009.
- [3] B. Nie, X. He *et al.*, "Present situation and progress trend of prediction technology of coal and gas outburst," *China Safety Science Journal*, vol.13, no. 6, pp. 40-43, 2003.
- [4] Q.-B. Mou, "Study status and prospects for regional prediction methods of coal and gas outburst in China," *Coal Science and Technology*, vol. 42, no. 11, pp. 59-63, 2014.
- [5] F. Y. Zhang, D. M. Zhang, and A. A. Office, "Application of comprehensive index method in coal and gas outburst prediction," *Coal Mine Machinery*, vol. 37, no. 2, pp. 137-140, 2016.

[6] K. Fang, Z. Y. Lu, and B. Fang, "The present situation and analysis of the prediction index of coal and gas outburst hazard," *China Science and Technology Information*, 2015, vol. 5, pp. 57-59.

[7] S. L. Kong, L. B. Cheng, H. F. Wang *et al.*, "Determination and application on critical value of drilling cuttings gas desorption indices," *Coal Science and Technology*, vol. 42, no. 8, pp. 56-59, 2014.

[8] Y. Chen, J. Qin, L. I. Jiangong *et al.*, "Application of AE technique in coal and gas outburst prediction," *Mining Safety & Environmental Protection*, vol. 6, pp. 11-16, 2017.

[9] E. Y. Wang, L. I. Zhong-Hui *et al.*, "Application and Pre-warning technology of coal and gas outburst by electromagnetic radiation," *Coal Science & Technology*, vol. 42, no. 6, pp. 53-57, 2014.

[10] X. J. Liao, C.-B. Jiang, M.-K. Duan, and W. J. Tian, "Influence of temperature on the mechanical and deformation properties of containing-gas raw coal under mining," *Journal of Northeastern University (Natural Science)*, vol. 38, no. 9, pp. 1347-1352, 2017.

[11] Z. L. Mao, "Forecast of coal and gas outburst in working face based on fractal theory," *Coal Technology*, vol. 33, no. 6, pp. 43-45, 2014.

[12] K. Wang, Y. Z. Peng *et al.*, "Application of non-linear theory in study of coal and gas outburst," *Journal of Liaoning Technical University*, vol. 22, no. 1, pp. 38-45, 2000.

[13] S.-Q. Ma and W.-J. Ju, "Application research on rheology and mutation mechanism in coal and gas outbursts," *Jiangsu Coal*, vol. 3, pp. 13-15, 2001.

[14] C. Qi, Y. L. Zhu, F. Gao *et al.*, "Safety analysis of lithium-ion battery by rheology-mutation theory coupling with fault tree method," *Journal of Loss Prevention in the Process Industries*, vol. 49, pp. 603-611, 2017.

[15] J. H. Liang, K. Wang, J. Tian *et al.*, "Prediction of coal and gas outburst hazard based on fuzzy mathematics," *Journal of North China Institute of Science & Technology*, vol. 14, no. 1, pp. 40-43, 2017.

[16] Z. Li, E. Wang, J. Ou *et al.*, "Hazard evaluation of coal and gas outbursts in a coal-mine roadway based on logistic regression model," *International Journal of Rock Mechanics and Mining Sciences*, vol. 80, pp. 185-195, 2015.

[17] Z. Wang, "Research on coal mine gas prediction algorithm based on improved svm," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 1729-1733, 2017.

[18] Y. Zhang, W. Yang, X. Yu *et al.*, "Risk assessment method of coal and gas outburst based on T-S fuzzy neural network," in *Proc. International Forum on Mechanical, Control and Automation*, 2017.

[19] R. Bogucki, J. Lasek, J. K. Milczek *et al.*, *Early Warning System for Seismic Events in Coal Mines Using Machine Learning*, vol. 8, 2016.

[20] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.

[21] B. A. Yang, J. I. Hai, X. U. Jing *et al.*, "An application of back propagation neural networks to warning for corporate financial distress," *Rorecasting*, vol. 20, no. 2, pp. 49-54, 2001.

[22] X. Liu, G. Wang, Z. Cai *et al.*, "Bagging based ensemble transfer learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 1, pp. 29-36, 2016.

[23] L. Xu, H. Y. Fu, M. Goodarzi, *et al.*, "Stochastic cross validation," *Chemometrics and Intelligent Laboratory Systems*, vol. 175, pp. 74-81, 2018.

[24] R. Kamimura, "Information-theoretic competitive learning with inverse euclidean distance output units," *Neural Processing Letters*, vol. 18, no. 3, pp. 163-204, 2003.

[25] W. Wang, R. Tang, C. Li *et al.*, "A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights," *Ocean Engineering*, vol. 162, pp. 98-107, 2018.

[26] B. Sun, H. Chen, J. Wang *et al.*, "Evolutionary under-sampling based bagging ensemble method for imbalanced data classification," *Frontiers of Computer Science*, vol. 12, no. 2, pp. 331-350, 2018.

[27] X. H. Shen, Z. H. Zhou, J. X. Wu *et al.*, "Survey of boosting and bagging," *Computer Engineering and Applications*, vol. 36, no. 12, pp. 31-32, 2000.

[28] X. Hu, R. Jiang, K. Yu *et al.*, "Uniformly asymptotic normality of sample quantiles estimator for linearly negative quadrant dependent samples," *Journal of Inequalities and Applications*, no. 1, pp. 196, 2018.



Jianxing Liao received the B.S. degree from the Hefei University of Technology in 2017, where he is currently pursuing the postgraduate degree with the School of Computer Science and Information Engineering at Hefei University of Technology, China. His research interests focus on data mining and machine learning.



Junwei Lv is currently a lecturer at Hefei University of Technology, China. He received his B.S. and M.S. degrees from Northeastern University in 2007, 2009 respectively. He is a Ph.D. student in School of Computer Science and Information Engineering at Hefei University of Technology. His research interests include data mining and machine learning.



Yuhan Shao is currently a postgraduate student at the School of Computer Science and Information Engineering in Hefei University of Technology, China. He received his B.S. degree from Hefei University of Technology in 2016. His research interests include data mining and natural language processing.



Peipei Li received the B.S., M.S., and Ph.D. degrees from the Hefei University of Technology, Hefei, China, in 2005, 2008, and 2013, respectively. She is currently an associate professor with the Hefei University of Technology. She was a research fellow with Singapore Management University, Singapore, from 2008 to 2009. She was a student intern with Microsoft Research Asia, Beijing, China, from 2011 to 2012. Her current research interests include data stream mining and knowledge engineering.



Xuegang Hu received the BS degree from the Department of Mathematics at Shandong University, and the MS and PhD degrees at Hefei University of Technology. He is a professor in the School of Computer Science and Information Engineering, Hefei University of Technology, China, and the director general of Computer Association of Higher Education at Anhui Province. His research interests include data mining and knowledge engineering.