Research and Application of Back Propagation Neural Network-Based Linear Constrained Optimization Method

ZHIGUI DONG[1], CHANGYOU WU[2], XISONG FU[2], AND FULIN WANG[3]

[1]Liaoning Institute of Science and Technology, Benxi 117004, China
[2]School of Management Science and Engineering, Shandong Institute of Business and Technology, Yantai 264005, China
[3]College of Engineering, Northeast Agricultural University, Harbin 150030, China

Corresponding author: Changyou Wu (wuchangyou_81@163.com)

ABSTRACT A back propagation (BP) neural network-based linear constrained optimization method(BPNN-LCOM) was proposed for to solve the problems in linear constraint black box in this paper,hoping to improve the shortcoming of BP neural network-based constrained optimization method (BPNN-COM). In view of minimizing the mathematic model of network output, the basic ideas of BPNN-LCOM wereilluminated,includingmodel design and training, and BP neural network-based global optimization. Firstly, the iteration step size was calculated by optimal step size, and the adjustment step size was calculated by interpolation method, also the iteration speed was accelerated. Secondly, the search direction that iteration point locates on the boundary offeasible region was determined by gradient projection method, which ensured that the iteration process continued along a feasible search direction, and effectively solved the defect of BPNN-COM that sometimes fails to find thetrue optimal solutions. At the same time, the iteration step size along the gradient projection direction was calculated by the optimal constraint step size, which ensured the new iteration point located in the feasible region. Thirdly, the Kuhn-Tucker conditions were introduced to verify whether the iteration point is theoptimization solution that locates on the boundary of feasible region, and it made the termination criterion perfect for BPNN-LCOM.The computation results of two examples showed the effectiveness and feasibility of BPNN-LCOM. The BPNN-LOCM was used to optimize the roller-type bailing mechanism,and the optimal parameters were obtained as follows: round disc diameter was 360 mm, rotationalspeed of the steel rollerwas 250 rpm, feeding quantity was1.7 kg/s, and length-width ratio was 0.8. The corresponding minimum power consumption was 45.8 kJ/bundle. The optimization results were superior to regression analysis and BPNN-COM.The verification test was carried out and the optimization results could improve roller-type bailing mechanism. Verification results showed that the BPNN-LCOM is a feasible method for solving problems in linear constraint black box.

INDEX TERMS BP neural network, optimization method, linear constraint, gradient projection method.

I. INTRODUCTION

Many optimization problems in the fields of scientific research and engineering application belong to black box problem. Their internal structure and interaction relationship are unclear, and the functional relationship between input variables and output variables is unclear or cannot be

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Wang.

expressed. Regression analysis is the traditional method to optimize this kind of problem. Researchers generally use the experimental design method to make a reasonable experimental scheme and carry out experiments, to obtain the corresponding data relationship between input and output, then carry out function fitting, obtain the regression model of input and output, and finally find out an optimal combination through the analysis of the variation law between various factors or the optimization of the model. With the

development of BP neural network algorithm, BP neural network has become the main method of functional relation fitting for black box problem, and theory proved that the three-layer BP neural network structure can approximate complex nonlinear functional relations with any precision [1]–[4]. In recent years, an optimization method has been proposed based on functional relationship fitting of BP neural network [5]–[13], which has expanded the application field of BP neural network and provides a new idea and method for optimization of black box problem. For the constrained black box optimization problem, Zhang *et al.* (2016) selected gradient method to determine the search direction and proposed a constrained optimization method based on BP neural network. Wang *et al.* [6], Dong *et al.* [10]–[12], and Zhao *et al.* [13] applied the method to the actual black box optimization problems in the field of agricultural engineering, and compared it with the regression analysis method. The accuracy and stability of the optimization results are better than the regression analysis method. However, the theoretical research of this method is not systematic enough, and there are still some deficiencies, mainly reflected in the following aspects:

(1) The given constant is used to determine the initial iteration step size. After each iteration, it is necessary to judge whether the output of the new iteration point is better than the current iteration point. According to the change of the output value, the iteration was carried out again to increase or decrease the iteration step size until the relatively optimal iteration point in the iteration direction was obtained. The iteration efficiency of the algorithm is relatively low.

(2) When the new iteration point exceeds the feasible region, the heuristic method is used to reduce the iteration steps, and adjust the new iteration point to the feasible region, which is complicated and inefficient.

(3) The iteration termination criterion is not complete, and Kuhn-Tucker conditions for constrained optimization problems to obtain extreme values is not considered.

(4) The gradient method is used to determine the search direction of the iteration point, the real optimal solution of optimization problem is sometimes not obtained. As shown in Fig.1, when the iteration point moves to the boundary of the feasible region, the search direction determined by the gradient method points out of the feasible region. The iteration is carried out along the search direction with any small step size greater than 0, and the obtained iteration points are all infeasible points, and the iteration process is forced to terminate. The iteration point $X(t)$ is the approximate optimal solution obtained by this optimization iteration. As can be seen from Fig. 1, the approximate optimal solution of the optimization problem obtained is not the real optimal solution, and an appropriate method is needed to determine the search direction and iteration step length of the iteration points on the boundary of the feasible region, so that the optimization iteration can continue, and then

the real approximate optimal solution of the optimization problem can be obtained.

In view of the above-mentioned defects and deficiencies of the neural network-based constrained optimization method(BPNN-COM), a neural network-based linear constrained optimization method (BPNN-LCOM) for constrained function all is linear was proposed. Compared with citation literature, the innovation and merit of proposed method based on the three key elements of the optimization method(search direction, iteration step size and termination criterion) as follow: (1) The optimal step size is used to instead of the constant initial step size, it ensure that each iteration can obtain the optimal point in the search direction. (2) The gradient projection method is applied to determine the search direction that the iterative point located on the boundary of feasible region, it solved the defect that the search direction determined by gradient is not feasible. (3) The adjustment step size calculated by the interpolation method, which is used to adjust the iteration point beyond the feasible region to the boundary of feasible region at a time, it accelerate the iteration speed. (4) The iteration criterion is improved, the Kuhn-Tucker condition for constrained optimization problems is added. The key of proposed method is how to determine the gradient projection direction, the optimal step size and the adjustment step size, Proposed method need to solve several difficulties. (1) Derivation of second order partial derivative of network output versus its input. (2) Determining of the contribute constrained function. (3) How to set the Kuhn-Tucker condition in algorithm. (4) How to determine the iterative step size of gradient projection direction. Combining the above innovations and difficulties, this paper systematically studies the BPNN-LCOM.

II. BACK PROPAGATION NEURAL NETWORK-BASED LINEAR CONSTRAINED OPTIMIZATION METHOD
A. BASIC IDEAS
The BPNN-LCOM is an improved optimization method for the black box optimization problem with linear constraints. The basic ideas of the method are as follows:

First, the BP neural network model structure was established according to the actual optimization problem, and it is trained and fitted by sample data. When the total output error of the network meets the expected accuracy, the BP neural network model parameters of the optimization problem are obtained. Then, taking the fitted BP neural network model as the objective function, an initial iteration point $X(0)$ is artificially or randomly generated in the feasible region composed of constraint conditions, and the gradient of X(0) point is calculated. If the gradient modulus of $X(0)$ point is less than ε_1 ($\varepsilon_1 > 0$), $X(0)$ is the optimal solution, and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, the gradient method is used to determine the search direction $S(0)$ of the $X(0)$ point, and the iterative correction amount is calculated that the $X(0)$ along the $S(0)$ with the optimal step size $\lambda(0)$.
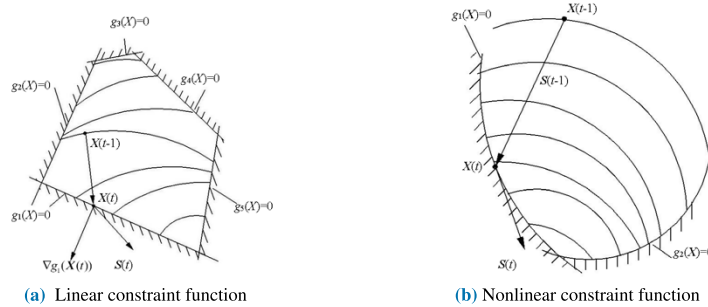
(a) Linear constraint function   (b) Nonlinear constraint function

FIGURE 1. Schematic diagram of the infeasible search direction.

If the correction amount is less than ε_2 ($\varepsilon_2 > 0$), $X(0)$ is the optimal solution, and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, a new iteration point $S(1)$ is obtained by iteration that $X(0)$ along the $S(0)$ with the optimal step size $\lambda(0)$. Then, check whether the point $X(1)$ satisfies the constraint conditions.

If the point $X(1)$ satisfies the constraint condition and locates in the feasible region, the gradient of the point $X(1)$ is calculated. If the gradient modulus of the point $X(1)$ is less than ε_1, $X(1)$ is the optimal solution, and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, the gradient method is used to determine the search direction $S(1)$ of the point $X(1)$, the iterative correction amount is calculated that the $X(1)$ along the search direction $S(1)$ with the optimal step size $\lambda(1)$. If the correction amount is less than ε_2, $X(1)$ is the optimal solution, its corresponding network output is the optimal value, the iteration will terminate. Otherwise, a new iteration point $X(2)$ is obtained by iteration that $X(1)$ along the search direction $S(1)$ with the optimal step size $\lambda(1)$. Then, check whether the point $X(2)$ satisfies the constraint conditions.

If the point $X(1)$ satisfies the constraint condition and locates on the boundary of the feasible region, the effective constraint functionsset is determined.Calculate the objective function and the gradient of the effective constraint of the point $X(1)$, and judge whether the point $X(1)$ satisfies the Kuhn-Tucker conditions. If it satisfies, $X(1)$ is the optimal solution, and its corresponding network output is the optimal value, the iteration will terminate. If $X(1)$ is not satisfied, the gradient projection method is used to determine the search direction $S(1)$ of point $X(1)$, and the iterative correction amount is calculated than the point $X(1)$ along the search direction $S(1)$ with the optimal step size $\lambda(1)$. If the correction amount is less than ε_2, $X(1)$ is the optimal solution, and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, a new iteration point $X(2)$ is obtained by iteration that $X(1)$ along the search direction $S(1)$ with the optimal step size $\lambda(1)$. Then, check whether the point $X(2)$ satisfies the constraint conditions.

If $X(1)$ does not satisfy the constraint condition, the constraint function with the largest constraint violation is determined. The interpolation method is used to calculate the adjusted step size $\lambda_c(0)$, and $X(1)$ is adjusted to the constraint boundary that the constraint function with the largest

constraint violation. The new iteration point $X(1)$ satisfies the constraint condition and locates on the boundary, the iteration will continue.

The calculation continues until the iteration point $X(t)$ $(t > 0)$ satisfies the iteration termination criterion (the gradient module of the iteration point or the correction amount of the network input satisfies the preset accuracy or satisfies Kuhn-tucker condition).

Compared to the methods utilized in the cited literature, the difference between specific methods is described as follow:

(1) The method of determining the iteration step length. a) The method proposed in this paper introduces the optimal step size instead of the constant method to determine the initial step size of each iteration to ensure that each iteration can obtain the optimal value in this direction. b) When the iteration point exceeds the feasible region of the constraint problem, In this paper, interpolation method was used instead of heuristic method to determine the adjustment step length, and the iteration point is adjusted to the boundary of the feasible region. c) Iteration points on the boundary of the feasible region. The method proposed in this paper uses the constraint step method to determine the iteration step length to ensure that the new iteration point is located in the feasible region constituted by the constraints.

(2) How to determine the search direction. The cited literature only used the gradient method as the method of determining the search direction of the iterative point. On the basis, this paper introduces the gradient projection method to improve the method of determining the search direction. When the iteration point is located on the boundary of the feasible region, the gradient projection method was use to instead of the gradient method to determine the search direction of the iteration point, and solved the defect of the iteration termination caused by the infeasibility of the search direction generated by the gradient method.

(3) Setting of termination criteria. The cited literature used the gradient modulus and the adjustment value as the termination criterion of the optimization iteration, and does not consider the Kuhn-Tuck condition for obtaining the extreme value of the constraint problem. For the iterative point on the boundary of the feasible region, the Kuhn-Tuck condition was used to judge whether the iterative point is the optimal point locate on the boundary of the feasible region,

and perfected the termination criterion of the optimization method.

There are two stages in the BPNN-LCOM, the design and training of BP neural network model and the global optimization based on BP neural network.

B. DESIGN AND TRAINING OF BP NEURAL NETWORK MODEL

1) MODEL DESIGN OF BP NEURAL NETWORK

The model design of BP neural network mainly includes network structure design, transfer function selection, normalization interval of training sample, learning rate selection, network weight and threshold initialization. Network structure design is the primary problem in BP neural network model design, since its reasonability directly determines the convergence ability, identification ability, fault tolerance ability and generalization ability of BP neural network model.
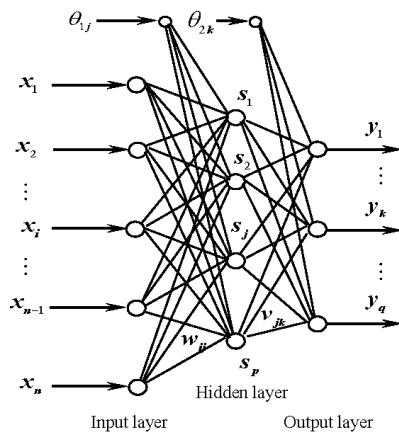

FIGURE 2. Structure chart of three layers of the BP neural network.

In this paper, a three-layer BP neural network model was applied to establish the network structure model, and its structure is shown in Fig.2 [14]. n denotes the number of the neurons in the network input layer, which is determined by the number of input variables of the practical problem, and x_i $(i = 1, 2, \ldots, n)$ denotes the ith neuron of the input layer. q denotes the number of neurons of the network output layer, which is determined by the number of output variables of the practical problem, y_k $(k = 1, 2, \ldots, q)$ denotes the kth neuron of the output layer. p denotes the number of neurons of the network hidden layer, and it is determined according to the empirical formula [14], s_j $(j = 1, 2, \ldots, p)$ denotes the output value of the jth hidden layer neuron; w_{ij} denotes the weight value between the ith input layer neuron and the jth hidden layer neuron, and v_{jk} denotes the weight value between the jth hidden layer neuron and the kth output layer neuron. θ_{1j} denotes the threshold of the jth hidden layer neuron, and θ_{2k} denotes the threshold of the kth output layer neuron.

2) TRAINING OF BP NEURAL NETWORK MODEL

The training process of BP neural network model is the learning process of network weights and thresholds, including the forward propagation of input signals and the back

propagation of error signals. The propagation process follows the chain transmission rule. In the forward propagation process, the input signals (training sample) propagate from the input layer, and the net input and output of each neuron is calculated layer by layer until the output layer. If the error between the actual and the expected output of the output layer fails to meet the expected accuracy, the backward propagation process of the error is transferred. In the backward propagation process, the error signals propagate from the output layer, and the output error of each neuron is calculated layer by layer until the input layer, and the weights and thresholds of each layer is adjusted according to error gradient decline. The learning and training process of forward propagation of signals and backward propagation of errors is carried out repeatedly until the total error between the actual and the expected output of the network is less than the expected accuracy or reaches the preset learning times.At this time, the training process of BP neural network model is completed, and the weights and thresholds of the network are saved. The objective function of constrained optimization problem can be expressed as follows:

$$Y = F(X) = f[V * f(W * X + \theta_1) + \theta_2] \quad (1)$$

where, $f()$ is the transfer function of the BP neural network model, X is the input vector of the BP neural network,and $X = [x_1, x_2, \ldots, x_n]^T$. Y is the output vector of the BP neural network,and $Y = [y_1, y_2, \ldots, y_q]^T$. $F(X)$ is the functional relationship between input and output. W is the weight matrix between the input layer and the hidden layer, θ_1 is the threshold of the hidden layer, V is the weight matrix between the hidden layer and the output layer, θ_2 is the threshold for the output layer.

C. GLOBAL OPTIMIZATION METHOD BASED ON BP NEURAL NETWORK

1) GLOBAL OPTIMIZATION METHOD BASE ON BP NEURAL NETWORK

In order to ensure the generality of the problem, the global optimization method based on BP neural network is expounded by taking the output minimization of BP neural network as an example. Suppose $g_h(X)$, $(h = 1, 2, \ldots, m)$ is the constraint condition of the constraint optimization problem, and m is the number of constraint conditions. Combined with the objective function of the constrained optimization problem, the general mathematical model of the constrained optimization problem based on BP neural network is

$$\begin{cases} \min Y = \min F(X) \\ s.t. g_h(X) \leqslant 0 \ (h = 1, 2, \ldots, m) \end{cases} \quad (2)$$

If solving the optimization problem is to find the maximum output value of the network, the objective function can be transformed by $\max F(X) = -\min[-F(X)]$.

Firstly, the convergence accuracy of the termination criterion is given, and an initial point $X(0)$ is artificially or randomly generated in the feasible region. The iteration search

begins from $X(0)$. Suppose $X(t)$ $(t > 0)$ is the feasible point obtained after the tth iteration, the BP neural network output $F(X(t))$ is calculated by forward propagation, and the gradient vector of $F(X(t))$ is calculated by

$$\nabla F(X(t)) = \frac{\partial F(X)}{\partial X}\mid_{X=X(t)} \quad t \in [0, 1, 2, \ldots] \quad (3)$$

Then verify whether the gradient vector $\nabla F(X(t))$ satisfies the following iteration termination condition

$$\|\nabla F(X(t))\| \leqslant \varepsilon_1 \quad (4)$$

where, $\|\nabla F(X(t))\|$ is the modulus of $F(X(t))$, and ε_1 is the preset accuracy. If meets Eq.(4), $X(t)$ is the optimal solution, the optimal solution is obtained as follows:

$$X^* = X(t) \quad (5)$$

its corresponding output $Y^* = F(X^*)$ is the optimal output value, and the iteration will terminate.

If $F(X(t))$ does not satisfy Eq.(4), then the search direction $S(t)$ of $X(t)$ is generated by the gradient method as follow

$$\begin{aligned} S(t) = -\nabla F(X(t)) &= \frac{\partial F(X(t))}{\partial X(t)} \\ &= \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n} \right) \end{aligned} \quad (6)$$

The optimal step size $\lambda(t)$ of $X(t)$ along the search direction $S(t)$ is calculated by

$$\lambda(t) = -\frac{S(t)^T \nabla F(X(t))}{S(t)^T \nabla^2 F(X(t))} \quad (7)$$

where, $S(t)$ is the search direction of $X(t)$, is the gradient vector of $F(X(t))$, and is the Hessen matrix of objective function at $X(t)$, which can be obtained by the second derivative of the network output to the input. The iteration correction amount $\Delta X(t)$ that $X(t)$ iterates along the search direction $S(t)$ as the optimal step size $\lambda(t)$ is calculated by

$$\Delta X(t) = \lambda(t) S(t) \quad (8)$$

Then verify whether the correction amount $\Delta X(t)$ satisfies the following iteration termination condition:

$$|\Delta X(t)| \leqslant \varepsilon_2 \quad (9)$$

If the correction amount $\Delta X(t)$ meets Eq.(9), it shows that the correction amount $\Delta X(t)$ that $X(t)$ iterates along the search direction $S(t)$ is very small. Then, $X(t)$ is the optimal solution, and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, $X(1)$ iterates along the search direction $S(t)$ with the optimal best step $\lambda(t)$ and a new iteration point $X(t + 1)$ is obtained as follow

$$X(t + 1) = X(t) + \lambda(t) S(t) \quad (10)$$

Calculate the function value $g_h(X(t + 1))$ $(h = 1, 2, \ldots, m)$ of all constraint functions at $X(t + 1)$ and the maximum $g(t + 1)$ can be obtained below:

$$g(t + 1) = \max\{g_h(X(t + 1))\mid(h = 1, 2, \ldots, m)\} \quad (11)$$

For constrained problem, there are three different situations in which the position of iteration point $X(t + 1)$ is relative to the feasible region, which can be judged by comparing the relation that $g(t + 1)$ is related to 0. So there are three different process modes for $X(t + 1)$ in the iteration process.

Case 1: $g(t + 1) < 0$, iteration point $X(t + 1)$ meets the constraint conditions and it locates in the feasible region, let $t = t + 1$. The BP neural network output $F(X(t))$ is calculated by forward propagation, and the gradient vector of $F(X(t))$ is calculated by Eq.(3), then verify the gradient vector $\nabla F(X(t))$ of objective function at new $X(t)$ to see whether it satisfies Eq.(4). If yes, $X(t)$ is the optimal solution and its corresponding network output is the optimal value, the iteration is terminated. Otherwise, the search direction $S(t)$ is determined by Eq.(6), and the optimal step size $\lambda(t)$ that $X(t)$ iterates along search direction $S(t)$ is calculated by Eq.(7), then the iterative correction amount $\Delta X(t)$ is calculated by Eq.(8). If the iteration amount $\Delta X(t)$ meets Eq.(9), $X(t)$ is the optimal solution and its corresponding network output is the optimal value, the iteration is terminated. Otherwise, the iteration is continued from $X(t)$, a new iteration point $X(t + 1)$ is found with optimal step size $\lambda(t)$ along search direction $S(t)$. The function value of all constraint functions at $X(t + 1)$ is calculated and the maximum $g(t + 1)$ is found by Eq.(11), then the process mode of $X(t + 1)$ is selected according to the position $X(t + 1)$, which relates to the feasible region.

Case 2: $g(t + 1) = 0$, iteration point $X(t + 1)$ meets the constraint conditions and it locates on the boundary of feasible region, let $t = t + 1$. The set of contributing constraint functions is found as follows:

$$J = \{h\mid g_h(X(t)) = 0\}(h = 1, 2, \ldots, m) \quad (12)$$

The gradient of objective function $\nabla F(X(t))$ and the gradient of contributing constrained function $\nabla g_h(X(t))$ are calculated at $X(t)$ point, to see whether they satisfy the KKT condition that the gradient of objective function $\nabla F(X(t))$ and the gradient of contributing constraint function $\nabla g_h(X(t))$ $(h = 1, 2, \ldots, J < m)$ satisfy the follow equation:

$$\begin{cases} \nabla F(X(t)) + \sum_{h=1}^{J} \beta_h \nabla g_h(X(t)) = 0 \\ \beta_h \geqslant 0 (h = 1, 2, \ldots, J < m) \end{cases} \quad (13)$$

where, β_h $(h = 1, 2, \ldots, J < m)$ is the Lagrangian multiplier of the hth constraint condition, $\beta_h \geqslant 0$.

If $X(t)$ meets the KKT condition, $X(t)$ is the optimal solution and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, as shown in Fig.3, the negative gradient vector of objective function is projected to the constraint boundary (or the intersection of constraint surface), and the gradient projection vector $S(t)$ is obtained by [15]

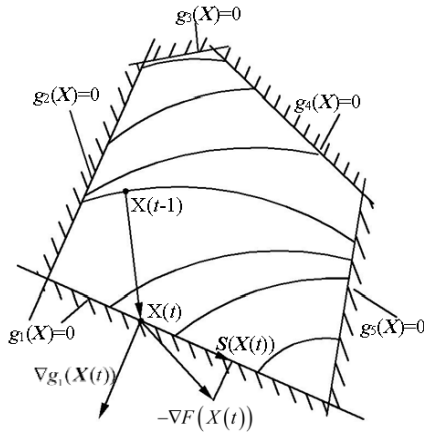$$S(t) = -\frac{P\nabla F(X(t))}{\|P\nabla F(X(t))\|} \quad (14)$$

FIGURE 3. Gradient projection direction on the constraint boundary.

$$P = I - M \left[M^T M \right]^{-1} M^T \tag{15}$$

$$M = [\nabla g_1 (X (t)), \nabla g_2 (X (t)), \dots \nabla g_J (X (t))] \tag{16}$$

where, $\nabla F (X (t))$ is the gradient of objective function at $X (t)$. P is the projection operator, which is an $n \times n$-order matrix. I is an $n \times n$-order unit matrix, and M is a gradient matrix that contributes constraint functions at $X (t)$.

The optimal constraint step size that X(t) iterates along the search direction S(t) is calculated by

$$\lambda_S^* (t) = \min_{\lambda_{th}} \{\lambda_h (t) \mid (h = J + 1, J + 2, \dots m)\} \tag{17}$$

where, $\lambda_h (t)$ is the step size that $X (t)$ iterates along the search direction $S (t)$ to the hth non-contributing constraint boundary. For all $(m - J)$ non-acting constraints, the step size $\lambda_h (t)$ that $X (t)$ iterates along search direction $S (t)$ to the constraint boundary $g_h (X (t)) = A_h X + B_h = 0$ must satisfy

$$g_h (X (t + 1)) = g_h (X (t) + \lambda_h (t) A_h S (t)) \tag{18}$$

which is

$$\lambda_h (t) = \frac{-g_h (X (t))}{A_h S (t)} (h = J + 1, J + 2, \dots m) \tag{19}$$

Let $\lambda (t) = \lambda_S^* (t)$, the correction amount that $X (t)$ iterates along search direction $S (t)$ with the optimal constraint step size $S (t)$ is calculated by Eq.(8), and verifying whether it satisfies Eq.(9). If yes, $X (t)$ is the optimal solution and its corresponding network output is the optimal value, the iteration will terminate. Otherwise, the iteration should continue from $X (t)$, a new iteration point $X (t + 1)$ is found with optimal constraint step size $\lambda (t)$ along search direction $S (t)$. The function value of all constraint functions at $X (t + 1)$ is calculated and the maximum $g (t + 1)$ is found by Eq.(11). Then the process mode of $X (t + 1)$ is selected according the position $X (t + 1)$, which relates to the feasible region.

Case 3: $g (t + 1) (t > 0)$, iteration point $X (t + 1)$ does not satisfy the constraint conditions, and $X (t + 1)$ locates outside the feasible region formed by constraint conditions. Suppose is the constraint condition that have the most constraint function value at $X (t + 1)$. $\lambda_c (t)$ is the adjustment

step size that adjusts $X (t + 1)$ to the boundary of constraint condition $g_h (X) = 0$. It can be calculated by interpolation method through the following equation:

$$\lambda_c (t) = \frac{0 - g_h (X (t))}{g_h (X (t + 1)) - g_h (X (t))} \lambda (t) \tag{20}$$

where, $g_h (X (t))$ and $g_h (X (t + 1))$ respectively denote the function value of constraints function $g_h (X)$ at $X (t)$ and $X (t + 1)$, and $\lambda (t)$ is the iteration step size that $X (t)$ iterates along search direction $S (t)$ to $X (t + 1)$.

Let $\lambda (t) \Leftarrow \lambda_c (t)$, the iteration continues from $X (t)$, and it iterates along the search direction with $\lambda (t)$. The iteration point is adjusted to the boundary of constraint condition. Then the new $X (t + 1)$ satisfies the constraint conditions and locates on the boundary of feasible region, and the iteration process will continue according to case 2.

This iteration continues until the iteration point meets the termination criterion (the modulus of gradient or the correction amount is smaller than preset accuracy, the KKT condition), and the iteration point is the optimal solution of the optimization problem, and the corresponding network output is the optimal value, and the iteration will terminate.

2) PARTIAL DERIVATIVE OF NETWORK OUTPUT VERSUS ITS INPUT

The partial derivative of network output versus its input is the key that determines the search direction and the iteration step size for the iteration process. The following is the procedure to derive the partial derivatives of the BP neural network output versus its input. Suppose $x_i, x_i (i = 1, 2, \cdots, n)$ is the ith and l^{th} network input variable, $y_k (k = 1, 2, \dots, q)$ is the kth network output variable. $I_{1j} (j = 1, 2, \dots, p)$ is the input of the jth hidden layer neuron and $I_{2k} (k = 1, 2, \dots, q)$ is the input of the kth output layer neuron. The first-order partial derivative of network output y_k versus its input x_i can be expressed by Eq.(21):

$$\frac{\partial y_k}{\partial x_i} = \frac{\partial y_k}{\partial I_{2k}} \cdot \sum_{j=1}^{p} \left(\frac{\partial I_{2k}}{\partial x_i} \right)$$

$$= \frac{\partial y_k}{\partial I_{2k}} \cdot \sum_{j=1}^{p} \left(\frac{\partial I_{2k}}{\partial s_j} \cdot \frac{\partial s_j}{\partial I_{1j}} \cdot \frac{\partial I_{1j}}{\partial x_i} \right) \tag{21}$$

$$\frac{\partial I_{2k}}{\partial s_j} = v_{jk} \tag{22}$$

$$\frac{\partial I_{1j}}{\partial x_i} = w_{ij} \tag{23}$$

The unipolar sigmoid function is general transfer function for BP neural network, and it can be expressed by Eq.(24):

$$f (x) = \frac{1}{1 + e^{-x}} \tag{24}$$

The first-order derivative of Eq.(24) is:

$$f' (x) = f (x) [1 - f (x)] \tag{25}$$

If let

$$a_k = \frac{\partial y_k}{\partial I_{2k}} = y_k (1 - y_k) \tag{26}$$

$$b_{jk} = \frac{\partial s_j}{\partial I_{1j}} = s_j (1 - s_j) \tag{27}$$

Then we have

$$\frac{\partial y_k}{\partial x_i} = a_k \sum_{j=1}^{p} (v_{jk} \cdot b_{jk} \cdot w_{ij}) \tag{28}$$

According to the first-order partial derivative of network output versus its input, the second-order partial derivative of network output versus its input x_i, x_l can be expressed by Eq.(29):

$$\frac{\partial^2 y_k}{\partial x_i \partial x_l} = \frac{\partial \left(\frac{\partial y_k}{\partial I_{2k}} \right)}{\partial x_l} \cdot \sum_{j=1}^{p} \frac{\partial I_{2k}}{\partial x_i}$$

$$+ \frac{\partial y_k}{\partial I_{2k}} \cdot \frac{\partial \left(\sum_{j=1}^{p} \frac{\partial I_{2k}}{\partial x_i} \right)}{\partial x_l} \tag{29}$$

$$\frac{\partial \left(\frac{\partial y_k}{\partial I_{2k}} \right)}{\partial x_l} = \left(\frac{\partial y_k}{\partial I_{2k}} \right)' \cdot \sum_{j=1}^{p} \frac{\partial I_{2k}}{\partial x_l} \tag{30}$$

$$\frac{\partial \left(\sum_{j=1}^{p} \frac{\partial I_{2k}}{\partial x_i} \right)}{\partial x_l} = \sum_{j=1}^{p} \left(\frac{\partial I_{2k}}{\partial s_j} \cdot \frac{\partial I_{1j}}{\partial x_i} \cdot \left(\frac{\partial s_j}{\partial I_{1j}} \right)' \cdot \frac{\partial I_{1j}}{\partial x_l} \right) \tag{31}$$

According to the Eq.(24) and Eq.(25), the second-order derivative of unipolar Sigmoid can be expressed by:

$$f''(x) = f(x) [1 - f(x)] [1 - 2f(x)] \tag{32}$$

If let

$$c_k = \left(\frac{\partial y_k}{\partial I_{2k}} \right)' = y_k (1 - y_k) (1 - 2y_k) \tag{33}$$

$$d_{jk} = \left(\frac{\partial s_j}{\partial I_{1l}} \right)' = s_j (1 - s_j) (1 - 2s_j) \tag{34}$$

Then the second-order partial derivative of the network output versus its input is derived according to Eq.(21)-Eq.(25):

$$\frac{\partial y_k}{\partial x_i \partial x_l} = c_k \sum_{j=1}^{p} (v_{jk} b_{jk} w_{ij}) \sum_{j=1}^{p} (v_{jk} b_{jk} w_{lj})$$

$$+ a_k \sum_{j=1}^{p} (v_{jk} d_{jk} w_{ij} w_{lj}) \tag{35}$$

3) ALGORITHM IMPLEMENTATION

Fig. 4 shows the algorithm implementation of the global optimization method based on BP neural network.

III. SAMPLE VERIFICATION AND ANALYSIS

A. EXAMPLE 1

Suppose Eq.(36) is given linear constrained optimization function

$$\begin{cases} \min F(X) = x_1^2 + x_2 \\ s.t. \begin{cases} g_1(X) = x_2 - x_1 - 3 \leqslant 0 \\ g_2(X) = x_1 + x_2 - 1 \leqslant 0 \\ g_3(X) = x_1 - x_2 - 3 \leqslant 0 \\ g_4(X) = -x_2 - x_1 - 3 \leqslant 0 \end{cases} \end{cases} \tag{36}$$

The objective function of the constrained optimization problem is a typical two-dimensional unimodal function. The schematic diagram of objective function is shown in Fig.5, and the contour of the objective function and the feasible region of constraint functions are shown in Fig.6. As shown in Fig.6, the optimal solution of the constrained optimization problem locates at the intersection of the constraint functions $g_3(X)$, $g_4(X)$ and coordinate axis x_2, and the theoretical optimal value is $\min F(X) = F(0, -3) = -3$. Firstly, the independent variable of $F(X)$ was discretized and data sample of constrained black box optimization problem was obtained. The discrete interval of X is $[-5.2, 5.2]$, and the discrete level is equidistant 10 levels. 100 discrete samples are generated, and its corresponding function value $F(X)$ are calculated by Eq.(36). Among them, 90 discrete samples were randomly selected as training samples, and the remaining discrete samples are used as inspection samples.

1) DESIGN AND TRAINING OF BP NEURAL NETWORK MODEL

The structure of the 3-layer BP neural network was chosen as 2-9-1 according to the constrained optimization problem. The number of input layer neuron is 2, it determined by function variable of example 1. The number of output layer neuron is 1, it determined by the function value of example 1. The hidden layer neuron is determined 7 according to the calculation formula and network performance test. The unipolar Sigmoid function was selected as transfer function of the hidden layer and the output layer, the normalized interval of training sample is [0.2, 0.8], the learn rate is 0.8, this two parameters was determined by empirical value. The expectation accuracy is E = 0.00001, it determined by network performance test. The computer program of LM-BP algorithmic was written by MATLAB R2010a, and it was applied to train the BP network neural model. The training was terminated until the output error met expectation accuracy, and the BP neural network model of constrained optimization problem was obtained.

The weight matrix W between input layer and hidden layer is shown at the bottom of the next page.

The weight matrix V between hidden layer and output layer is shown at the bottom of the next page.

The threshold of hidden layer is shown at the bottom of the next page.

The threshold of output layer is
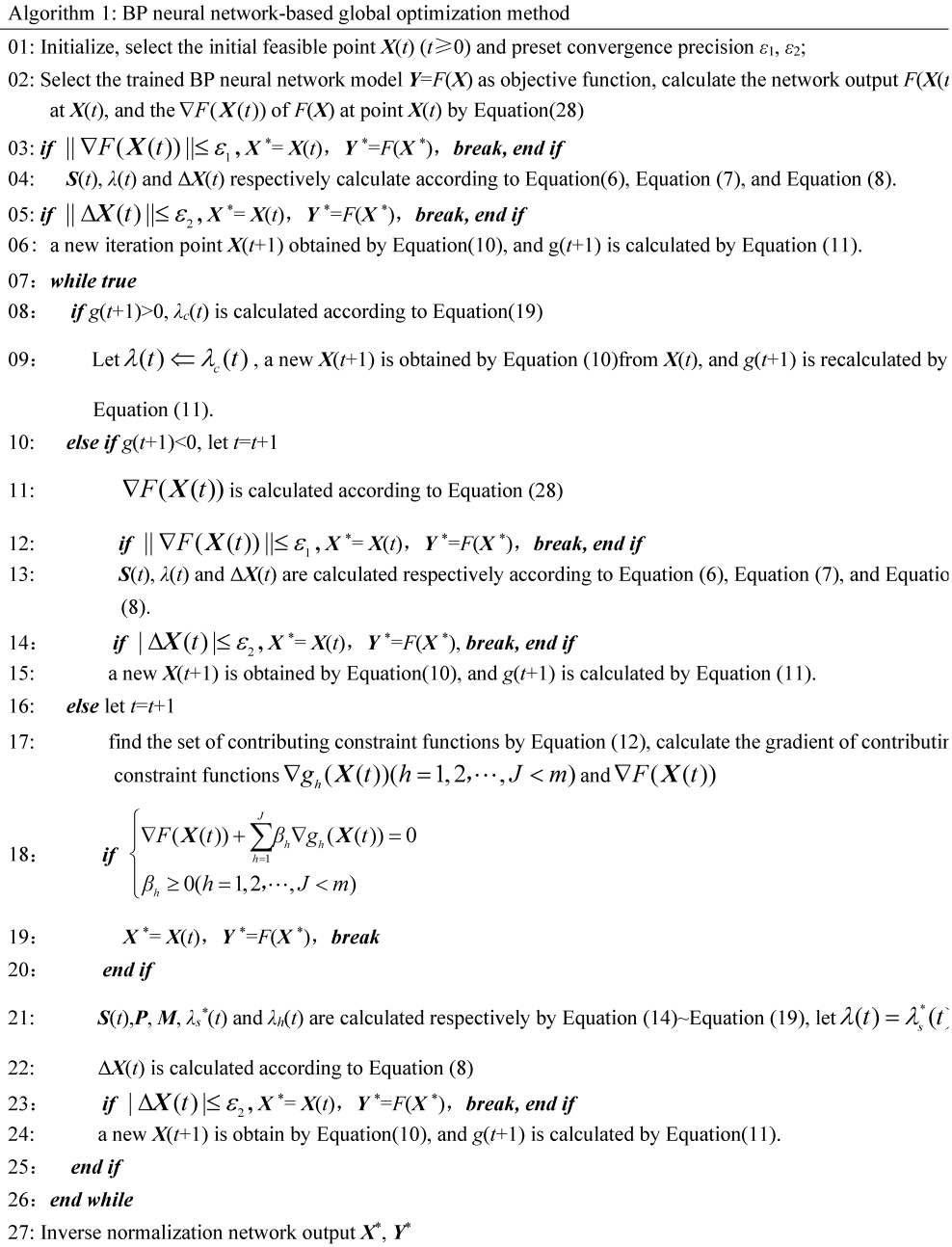
$$\theta_2 = [95.1465]$$

Algorithm 1: BP neural network-based global optimization method

---

01: Initialize, select the initial feasible point $X(t)$ $(t \geqslant 0)$ and preset convergence precision ε_1, ε_2;

02: Select the trained BP neural network model $Y=F(X)$ as objective function, calculate the network output $F(X(t$
    at $X(t)$, and the $\nabla F(X(t))$ of $F(X)$ at point $X(t)$ by Equation(28)

03: **if** $\| \nabla F(X(t)) \| \leq \varepsilon_1$, $X^*= X(t)$, $Y^*=F(X^*)$, ***break, end if***

04:    $S(t)$, $\lambda(t)$ and $\Delta X(t)$ respectively calculate according to Equation(6), Equation (7), and Equation (8).

05: **if** $\| \Delta X(t) \| \leq \varepsilon_2$, $X^*= X(t)$, $Y^*=F(X^*)$, ***break, end if***

06：a new iteration point $X(t+1)$ obtained by Equation(10), and g(t+1) is calculated by Equation (11).

07：***while true***

08：   ***if*** $g(t+1)>0$, $\lambda_c(t)$ is calculated according to Equation(19)

09：   Let $\lambda(t) \Leftarrow \lambda_c(t)$, a new $X(t+1)$ is obtained by Equation (10)from $X(t)$, and g(t+1) is recalculated by
         Equation (11).

10：   ***else if*** $g(t+1)<0$, let $t=t+1$

11： $\nabla F(X(t))$ is calculated according to Equation (28)

12：   **if** $\| \nabla F(X(t)) \| \leq \varepsilon_1$, $X^*= X(t)$, $Y^*=F(X^*)$, ***break, end if***

13： $S(t)$, $\lambda(t)$ and $\Delta X(t)$ are calculated respectively according to Equation (6), Equation (7), and Equatio
         (8).

14：   **if** $| \Delta X(t) | \leq \varepsilon_2$, $X^*= X(t)$, $Y^*=F(X^*)$, ***break, end if***

15：   a new $X(t+1)$ is obtained by Equation(10), and g(t+1) is calculated by Equation (11).

16：   ***else*** let $t=t+1$

17：   find the set of contributing constraint functions by Equation (12), calculate the gradient of contributir
         constraint functions $\nabla g_h(X(t))(h = 1, 2, \cdots, J < m)$ and $\nabla F(X(t))$

18：   ***if*** $\begin{cases} \nabla F(X(t)) + \sum_{h=1}^{J} \beta_h \nabla g_h(X(t)) = 0 \\ \beta_h \geq 0 (h = 1, 2, \cdots, J < m) \end{cases}$

19： $X^*= X(t)$, $Y^*=F(X^*)$, ***break***

20：   ***end if***

21： $S(t)$,P, M, $\lambda_s^*(t)$ and $\lambda_h(t)$ are calculated respectively by Equation (14)~Equation (19), let $\lambda(t) = \lambda_s^*(t$

22： $\Delta X(t)$ is calculated according to Equation (8)

23：   **if** $| \Delta X(t) | \leq \varepsilon_2$, $X^*= X(t)$, $Y^*=F(X^*)$, ***break, end if***

24：   a new $X(t+1)$ is obtain by Equation(10), and g(t+1) is calculated by Equation(11).

25：   ***end if***

26：***end while***

27: Inverse normalization network output X^*, Y^*

FIGURE 4. BP neural network-based global optimization method.

$$W = \begin{bmatrix} -3.46 & 2.46 & -7.50 & -2.28 & -1.97 & -130.51 & -2.77 & -2.11 \\ 0.10 & -0.14 & 1.30 & -0.02 & -0.39 & 0.64 & 0.10 & 0.19 \end{bmatrix}^T$$

$$V = \begin{bmatrix} 76.0 & -410.1 & 0.3 & -103.9 & -117.6 & -0.01 & -179.2 & -226.9 \end{bmatrix}$$

$$\theta_1 = \begin{bmatrix} -3.34 & 0.19 & -2.95 & -2.36 & 5.65 & 26.47 & -0.01 & -0.40 \end{bmatrix}^T$$

The determinate coefficient R^2 of the fitted BP neural network model is $0.9996 (P < 0.01)$; the root mean square error (RMSE) is 0.0048; the average relative error between the fitted value and the theoretical value is -0.0787%. The comparison of function theoretical value with fitted value of the BP neural network model is shown in Fig.7.
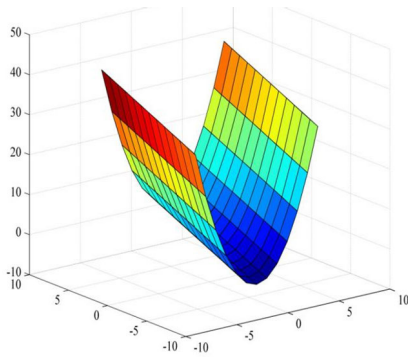
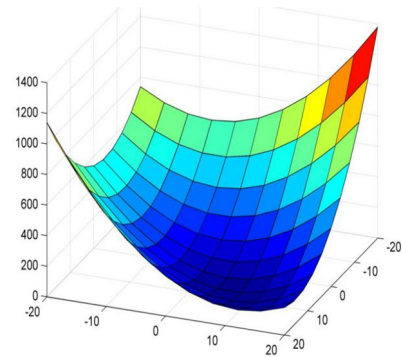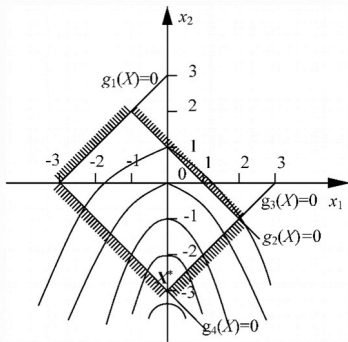**FIGURE 5.** Schematic diagram of example objective function.

FIGURE 6. Contour line of objective function and geometric figure of feasible region.
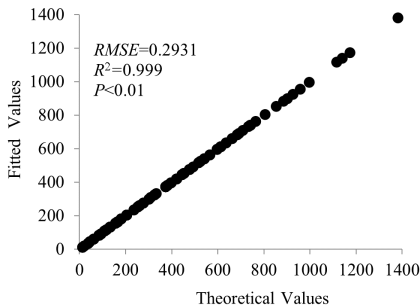

FIGURE 7. Comparison of theoretical value and fitted value of the BP neural network models.

The prediction accuracy of the trained model of BP neural network was verified by independent inspection samples. The predicted values and error analysis of the inspection samples calculated by the BP neural network model are shown in TABLE.1. As can be seen from Fig.7 and Table 1 that the trained BP neural network model by training sample had better fitting effect and higher prediction accuracy, and the function relationship can be accurately expressed between variable and index of constrained optimization problem.

2) BP NEURAL NETWORK-BASED GLOBAL OPTIMIZATION

Taking the trained BP neural network model as the objective function, the constrained optimization problem was optimized by the BPNN-COM and BPNN-LCOM. When $\varepsilon_1 = 10^{-4}$ and $\varepsilon_2 = 10^{-4}$, the optimization results obtained by two optimization methods are shown in Table 2. It can be seen that from Tale.2 that, the two different optimization methods respectively iterate from eight initial point. The optimization

FIGURE 8. Schematic diagram of example objective function.

result has the higher precision and the approximate optimal solution is very close to the theoretical value. The optimization result of BPNN-LCOM is better than BPNN-COM, the approximate optimal solution is $X = [-0.0001, -3.0]^T$, and the approximate optimal value is -2.9998;and the relative error between the approximate optimal value and the theoretical optimal value is 0.007%.

B. EXAMPLE 2

Suppose Eq.(37) is given linear constrained optimization function

$$\begin{cases} \min F(X) = 60 - 10x_1 - 4x_2 + x_1^2 + x_2^2 - x_1x_2 \\ s.t. \begin{cases} g_1(X) = -x_1 \leqslant 1 \\ g_2(X) = -x_2 \leqslant 0 \\ g_3(X) = x_1 - 6 \leqslant 0 \\ g_4(X) = x_2 - 8 \leqslant 0 \\ g_5(X) = x_1 + x_2 - 11 \leqslant 0 \end{cases} \end{cases} \quad (37)$$

The objective function of the constrained optimization problem is a typical two-dimensional unimodal function. The schematic diagram of objective function is shown in Fig.8, and the contour of objective function and the feasible region of constraint functions are shown in Fig.9. As shown in Figure 8, the optimal solution of the constrained optimization problem is located at the intersection of the constraint functions $g_3(X)$ and $g_5(X)$, and the theoretical optimal value is min $F(X) = F(6, 5) = 11$. Firstly, the independent variable of $F(X)$ was discretized and data sample of constrained black box optimization problem were obtained. The discrete interval of X is [-20, 20], and the discrete level is equidistant 11 levels. 121 discrete samples are generated, and its corresponding function value $F(X)$ are calculated by Eq.(37). Among them, 111 discrete samples were randomly selected as training samples, and the remaining discrete samples were used as inspection samples.

1) DESIGN AND TRAINING OF BP NEURAL NETWORK MODEL

The structure of the 3-layer BP neural network was chosen as 2-11-1 according to the constrained optimization problem.The number of input layer neuron is 2, it determined by function variable of example 2. The number of output layer

TABLE 1. Predicted value and error analysis of the BP neural network model.

Serial number	1	2	3	4	5	6	7	8	9	10
x1	-3	-4.1	2.5	4.7	1.4	3.6	-5.2	-1.9	-0.8	0.3
x2	-5.2	-4.1	-3	-1.9	-0.8	0.3	1.4	2.5	3.6	4.7
F(X)PV	3.8003	12.709	3.249	20.1905	1.1594	13.255	28.443	6.1	4.23	4.792
F(X)TV	3.8	12.71	3.25	20.19	1.16	13.26	28.44	6.11	4.24	4.79
RE(%)	0.009	-0.01	-0.016	0.002	-0.05	-0.035	0.011	-0.111	-0.18	0.041

Note: $F(X)_{PV}$ is the predicted value of BP neural network, $F(X)_{TV}$ is the theoretical values of given function, RE is error of predicted values and relative theoretical values, similarly hereinafter.

TABLE 2. Optimization results of two different optimization methods.

| Serial number	X(0) x1	x2	BPNN-COM x1	x2	Min F(X)	RE(%)	BPNN-LCOM x1	x2	min F(X)	RE(%)
1	1	1	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
2	1	1.5	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
3	-1.5	1	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
4	-2	0.8	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
5	-1.5	-1	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
6	-1.2	1	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
7	1	-2	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
8	-1	1	-0.002	-2.9994	-2.9996	0.013	-0.0001	-3	-2.9998	0.007
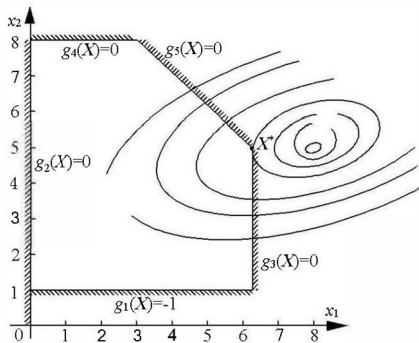
FIGURE 9. Contour line of objective function and geometric figure of feasible region.

neuron is 1, it determined by the function value of example 2. The hidden layer neuron is determined 11 according to the calculation formula and network performance test. The unipolar Sigmoid function was selected as transfer function of hidden layer and output layer, and the normalized interval of training sample is [0.2, 0.8], the initial learn rate is 0.8. The expectation accuracy is $E = 0.00001$, it determined by network performance test. The computer program of LM-BP algorithmic was written by MATLAB R2010a, and it was applied to train the BP network neural model. The training was terminated while the output error met the expectation accuracy, and the BP neural network model of constrained optimization problem was obtained.

The weight matrix W between input layer and hidden layer is as shown at the bottom of the page.

The weight matrix V between hidden layer and output layer is as shown at the bottom of the page.

The threshold of hidden layer is

$$\theta_1 = \begin{bmatrix} -3.34 & 0.19 & -2.95 & -2.36 & 5.65 & 26.47 & -0.01 & -0.40 \end{bmatrix}^T$$

The threshold of output layer is

$$\theta_2 = [2.781]$$

The determinate coefficient R^2 of the fitted BP neural network model is 0.9988 ($P < 0.01$); the root mean square error (RMSE) is 0.2931; the average relative error between the fitted value and the theoretical value is 0.061%. The comparison of function theoretical value with fitted value of the BP neural network model is shown in Fig.10. Fig.10 Comparison of theoretical value with fitted value by BP neural network models The prediction accuracy of the trained model of BP neural network was verified by independent

$$W = \begin{bmatrix} -3.46 & 2.46 & -7.50 & -2.28 & -1.97 & -130.51 & -2.77 & -2.11 \\ 0.10 & -0.14 & 1.30 & -0.02 & -0.39 & 0.64 & 0.10 & 0.19 \end{bmatrix}^T$$

$$V = \begin{bmatrix} 76.0 & -410.1 & 0.3 & -103.9 & -117.6 & -0.01 & -179.2 & -226.9 \end{bmatrix}$$

TABLE 3. Predicted value and error analysis of BP neural network model.

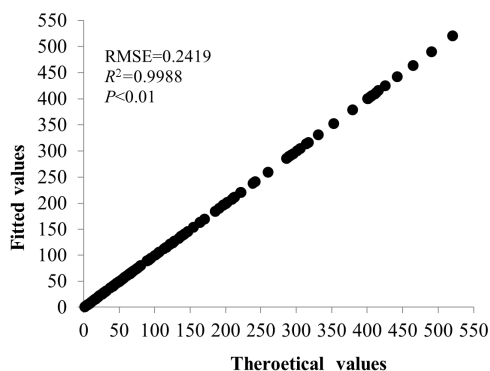| Serial number	1	2	3	4	5	6	7	8	9	10
x1	-20	-16	-12	-8	-4	0	4	8	12	16
x2	-16	8	0	4	-12	0	8	-20	16	-4
F(X)PV	660.5	635.72	324.21	236.1	259.96	60.34	35.91	683.41	84.14	252.24
F(X)TV	660	636	324	236	260	60	36	684	84	252
RE(%)	0.08	-0.04	0.06	0.05	-0.01	0.56	-0.24	-0.08	0.17	0.09

FIGURE 10. Comparison of theoretical value and fitted value by BP neural network models.

inspection samples. The predicted values and error analysis of the inspection samples calculated by the BP neural network model are shown in TABLE.3.

As can be seen from Fig.10 and Table.3 that the trained BP neural network model by training sample had better fitting effect and higher prediction accuracy, and the function relationship can be accurately expressed between variable and index of constrained optimization problem.

2) BP NEURAL NETWORK-BASED GLOBAL OPTIMIZATION

Taking the trained BP neural network model as the objective function, the constrained optimization problem was optimized by BPNN-COM and BPNN-LCOM. when $\varepsilon_1 = 10^{-4}$ and $\varepsilon_2 = 10^{-4}$, the optimization results obtained by two optimization methods are shown in TABLE.4. It can be seen form Table.4 that, the two different optimization methods respectively iterate from eight initial point. The optimization result has the higher precision and the approximate optimal solution is very close to the theoretical value. The optimization result of BPNN-LCOM is better than BPNN-COM, the approximate optimal solution is $X = [6.0, 5.0]^T$, and the approximate optimal value is 11.007, and the relative error between the approximate optimal value and the theoretical optimal value is 0.063%.

The example verification results showed that the stability and accuracy of the optimization results obtained by BPNN-LCOM are better than BPNN-LCOM. The improved ideas for black box problem of linear constraints was correct, and it effectively overcame the disadvantages of BPNN-COM that cannot obtain the true approximate optimal solution.

IV. PARAMETERS OPTIMIZATION FOR ROLLER-TYPE BALING MECHANISM

The roller-type baling mechanism is an important part of the rice straw baler. The power consumption value of bundling mechanism in the bundling process is one of the main indexes to measure the performance of the baler. Disc diameter, steel roller rotation speed, feeding amount and length-width ratio (ratio of rice straw length to bundle chamber length) are the main process parameters of bundle mechanism. The rationality of process parameters directly affects the power consumption value of bundle. In fact, the parameters optimization of the bundling mechanism is a black box optimization problem, which solves the optimal combination of process parameters to obtain the minimum power consumption based on the experimental data in bounding process. Li *et al.* used regression analysis method to optimize the technological parameters of the roller-type baling mechanism, and obtained the best combination of the four parameters, providing a guiding function for the design and improvement of the roller-type baling mechanism [16]–[18]. Zhao *et al.* used the BPNN-COM to carry out optimization research, and compared the optimization results obtained by the regression analysis method. The optimization results obtained by the BPNN-COM is better than obtained by regression analysis method [13]. This paper attempts to use the BPNN-LCOM to optimize the process parameters of roller-type baling mechanism. It also applies to comparing the optimization results with reference [13] and reference [16],to carry out verification experiments.

A. EXPERIMENTAL SCHEME AND RESULTS

In the test, the four parameters, disk diameter, steel roller rotation speed, feeding quantity and length-width ratio were selected as the experimental factors, and power consumption was taken as the experimental influence index, and the experimental goal is as small as possible. The coding schedule of test factors is shown in TABLE.5, and the experiment schemes and results are shown in TABLE.6 [16].

B. DESIGN AND TRAINING OF BP NEURAL NETWORK MODEL

According to the experimental scheme and results (Table.5 and Table.6, a 4-7-1 network structure was adopted in the BP neural network for parameters optimization of the roll-type baling mechanism. The number of input layer neuron is 4,

TABLE 4. Optimization result by two different optimization methods.

Serial number	X(0) x1	x2	BPNN-COM x1	x2	Min F(X)	RE(%)	BPNN-LCOM x1	x2	min F(X)	RE(%)
1	0	1	6	4.6182	11.15	1.36	6	5	11.007	0.063
2	2	1.5	5.9938	4.9952	11.09	0.72	6	5	11.007	0.063
3	4	3	5.9938	4.9952	11.09	0.72	6	5	11.007	0.063
4	5	5	5.9938	4.9952	11.09	0.72	6	5	11.007	0.063
5	1	3	5.9938	4.9952	11.09	0.72	6	5	11.007	0.063
6	2	5	5.9938	4.9952	11.09	0.72	6	5	11.007	0.063
7	3	7	5.8816	5.1184	11.49	4.45	6	5	11.007	0.063
8	2.5	7.5	5.9938	4.9952	11.09	0.72	6	5	11.007	0.063

TABLE 5. Coding schedule of experimental factors.

Factor	Factor level -2	-1	0	1	2
Round disk diameter / mm	220	260	300	340	380
Steel roller rotational speed / rpm	180	224	268	312	356
Feeding quantity / kg · s -1	0.5	1	1.5	2	2.5
Length-width ratio	0.6	0.8	1	1.2	1.4

x_1 denotes the round disc diameter, x_2 denotes the rotational speed of the steel roller, x_3 denotes the feeding quantity, and x_4 denotes the length-width ratio. The number of output layer neuron is 1, y_1 denotes the power consumption. The number of hidden layer neurons was determined 7 according to the calculation formula and network performance test. Unipolar sigmoid function was selected as the transfer function of hidden layer and output layer, the normalized interval of training sample data is [0.2, 0.8], the initial learning rate is 0.8, and the expectation accuracy of network output error is $E = 0.00001$.

The software MATLAB R2010a was used to write LM-BP neural network computer program, and taking the 30 groups test data in Table.6 (except groups 2, 5, 9, 16, 24 and 32) as training sample. The training sample normalization is used to fit the function, the training will stop until the output error of the network meets the expected accuracy. Then the network parameters were saved, and the BP neural network model between the influence index (power consumption) and the experimental factors were obtained. (the round disc diameter, the rotational speed of the steel roller, the feeding quantity, and the length-width ratio). The weight matrix of the input

layer and the hidden layer W is shown at the bottom of the page.

The weight matrix of the hidden layer and the output layer V is shown at the bottom of the page.

The threshold value of the hidden layer θ_1 is shown at the bottom of the page.

The threshold of output layer θ_2 is

$$\theta_2 = [0.5043]$$

According to the experimental data in TABLE.6, the regression equation of power consumption Y (*kj/bundle*) with round disc diameter x_1, rotational speed of the steel roller x_2, feeding quantity x_3, and length-width ratio x_4 are as follows [18]

$$\begin{aligned} Y = {}& 71.97 - 3.78x_1 + 7.76x_2 \\ & -11.17x_3 + 7.11x_4 - 0.23x_1x_2 \\ & -1.79x_1x_3 - 1.46x_1x_4 + 2.07x_2x_4 \\ & +0.83x_1^2 + 1.81x_2^2 + 4.28x_3^2 + 2.59x_4^2 \end{aligned} \quad (38)$$

The fitted values of BP neural network compared with experimental values are shown in Fig. 11a. The fitted values of quadratic regression model with experimental values are shown in Fig. 11b.

It can be seen from Fig.11 that, the determinate coefficient R^2 of the fitted BP neural network model is 0.984 (P<0.01), and the root mean square error RMSE is 2.037 kJ/bundle. The determinate coefficient R2 of quadratic regression model is 0.96(P<0.05), and the root mean square error RMSE is 2.9 kJ/bundle. The prediction accuracy of fitted BP neural network model are better than the quadratic regression model. The fitted BP neural network model has more approximate

$$W = \begin{bmatrix} -5.62 & 3.49 & 5.03 & -0.04 & 4.49 & -2.12 & -4.31 \\ -0.51 & 5.87 & 5.00 & -2.63 & -1.59 & 0.82 & -0.08 \\ 0.91 & -15.38 & -12.93 & 3.81 & -0.08 & -3.49 & -3.25 \\ 9.97 & -0.69 & -3.62 & 0.46 & 1.51 & 9.30 & -2.79 \end{bmatrix}^T$$

$$V = [6.9447 \quad -9.5793 \quad 8.0620 \quad -5.9632 \quad 4.6931 \quad -7.8724 \quad -4.1062]$$

$$\theta_1 = [1.3491 \quad 0.3407 \quad -0.5954 \quad 1.0238 \quad -0.8223 \quad 0.0080 \quad -0.3295]^T$$

TABLE 6. Experimental scheme and results.

Serial number	Round disk diameter /mm	Steel roller rotational speed /rpm	Feeding quantity / kg·s-1	Length-width ratio	Power consump-tion /kJ·Bundle -1
1	260	224	1	0.8	75.51
2	340	224	1	0.8	82.83
3	260	312	1	0.8	88.71
4	340	312	1	0.8	84.32
5	260	224	2	0.8	64.92
6	340	224	2	0.8	51.97
7	260	312	2	0.8	75.06
8	340	312	2	0.8	70.63
9	260	224	1	1.2	99.72
10	340	224	1	1.2	92.94
11	260	312	1	1.2	110.4
12	340	312	1	1.2	102.43
13	260	224	2	1.2	72.01
14	340	224	2	1.2	60.11
15	260	312	2	1.2	91.19
16	340	312	2	1.2	80.06
17	220	268	1.5	1	84.98
18	380	268	1.5	1	75.51
19	300	180	1.5	1	58.49
20	300	356	1.5	1	100.16
21	300	268	0.5	1	113.5
22	300	268	2.5	1	64.93
23	300	268	0.5	0.6	68.48
24	300	268	0.5	1.4	96.4
25	300	268	0.5	1	66.39
26	300	268	0.5	1	71.28
27	300	268	0.5	1	75.6
28	300	268	0.5	1	72.71
29	300	268	0.5	1	69.13
30	300	268	0.5	1	71.05
31	300	268	0.5	1	70.82
32	300	268	0.5	1	72.04
33	300	268	0.5	1	75.82
34	300	268	0.5	1	69.16
35	300	268	0.5	1	74.14
36	300	268	0.5	1	75.48

function relation between the influence index and the experimental factors.

The test data of six groups (groups 2, 5, 9, 16, 24 and 32) in TABLE.6 were selected as independent test samples (more than 10% of the sufficient amount of the test scheme) to test the BP neural network model. The predicted values and error analysis obtained by BP neural network model calculation of the test samples are shown in TABLE.7.

Through independent sample test, it is concluded that the BP neural network model obtained by training and fitting training samples has high prediction accuracy and stable

prediction results, and can obtain more accurate process parameter combination and power consumption value when applied to the optimization of process parameters of roller-type baling mechanism.

C. BP NEURAL NETWORK -BASED GLOBAL OPTIMIZATION

According to the upper and lower limits of each factor level in the experimental design, the constraint conditions for the optimization of power consumption parameters of the
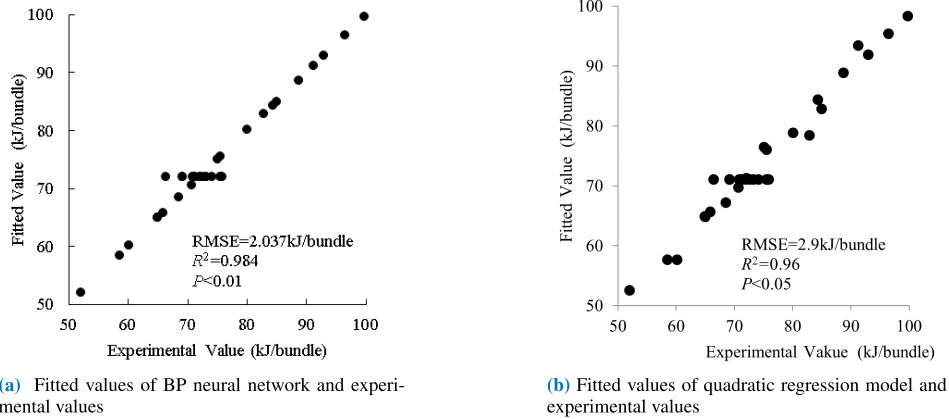
(a) Fitted values of BP neural network and experimental values

(b) Fitted values of quadratic regression model and experimental values

FIGURE 11. Comparison of experimental and fitted values by different models.

TABLE 7. Predicted value and error analysis of BP neural network model at testing sample.

Sample number	Disk diameter/km·h-1	Steel roller speed/rpm	Feed amount/ kg·s-1	Aspect ratio	Winding power consumption / kJ · bundle -1		error/ kg·s-1	Relative error/ %
					Test value	Predictive value		
2	340	224	1	0.8	82.83	82.826	-0.004	0.0048
5	260	224	2	0.8	64.92	64.968	0.048	0.0739
9	260	224	1	1.2	99.72	99.68	0.04	0.0401
16	340	312	2	1.2	80.06	80.092	0.032	0.04
24	300	268	0.5	1.4	96.4	96.407	0.007	0.0073
32	300	268	0.5	1	72.04	72.067	0.027	0.0375

roller-type baliing mechanism are as follows:

$$\begin{cases} 220 \leq x_1 \leq 360 \\ 180 \leq x_2 \leq 356 \\ 0.5 \leq x_3 \leq 2.5 \\ 0.6 \leq x_4 \leq 1.4 \end{cases} \quad (39)$$

Taking the trained BP neural network model as the objective function, the BPNN-LCOM was used to optimize the process parameters of the roller-type bailing mechanism, and the network input that makes the network output obtain the minimum value in the feasible region was solved. When $\varepsilon_1 = 10^{-4}$ and $\varepsilon_2 = 10^{-4}$, the optimization results were respectively obtained from 10 different initial points, and the optimization results are shown in TABLE.8.

It can be seen from Table 8 that the stable optimal solution can be obtained from ten different initial point, the optimal parameter of bailing mechanism are as follows: the round disc diameter is 360 mm, the steel roller rotational speed is 250 rpm, the feeding quantity is 1.7 kg/s, and the length-width ratio is 0.8. Its corresponding minimum bundle power consumption is 45.8 kJ/bundle.

The optimal combination of process parameters of the bailing mechanism obtained by regression analysis is as follows: disc diameter is 380mm, rotational speed of the steel roller is 247rpm, feeding amount is 1.7 kg/s, length-width ratio is 0.75. Under this parameter combination, the minimum bailing power consumption of the roller-type bailing mechanism is 62.7 kJ/bundle [16].

The optimal combination of process parameters of the bailing mechanism obtained by BPNN-COM is as follows:

disc diameter is 360mm, rotational speed of the steel roller 250rpm, feeding amount is 1.8 kg/s, length-width ratio is 0.8. Under this parameter combination, the minimum bailing power consumption of the roller-type bailing mechanism is 50.2 kJ/bundle [13].

Compared with the optimization results obtained by the three methods, the theoretical bundle power consumption obtained by the BPNN-LCOM is less than that in reference [16] and reference [13]. Compared with the regression analysis method, the theoretical bundle power consumption reduced 16.9 kJ/bundle, a relative reduction of 26.95%. Compared with the BPNN-COM, the theoretical bundle power consumption reduced by 4.4 kJ/bundle, and a relative reduction of 8.77%. The optimization research of process parameters of roller-type bailing mechanism belongs to black box problem, and the optimal solution of the black box problem is uncertain, so it is impossible to judge the best combination of process parameters obtained by the three models and the advantages and disadvantages of theoretical bundle power consumption. Theoretically, the better the fitting effect of the model, the higher the prediction accuracy, which can more truly express the functional relationship of the black box problem, and the higher the accuracy of the optimization results obtained, which can be used to optimize the black box problem.

D. VERIFICATION TEST

In order to test the correctness of the combination of process parameters obtained by the BPNN-LCOM, the verification test was conducted on September 30, 2018 in the Agricultural

TABLE 8. Optimization result calculated by BPNN-LCOM.

Serial number	Initial point				Optimal solution				Optimal value
	x1 /mm	x2/rpm	x3 /kg·s-1	x4	x1 /mm	x2/rpm	x3 /kg·s-1	x4	**Y/kJ·Bundle-1**
1	240	210	1.2	0.9	360	250	1.7	0.8	45.8
2	270	300	1.8	1.2	360	250	1.7	0.8	45.8
3	320	200	0.6	0.7	360	250	1.7	0.8	45.8
4	345	285	2.1	1.1	360	250	1.7	0.8	45.8
5	330	320	2	1	360	250	1.7	0.8	45.8
6	285	190	1.6	0.9	360	250	1.7	0.8	45.8
7	310	330	2.4	1.3	360	250	1.7	0.8	45.8
8	360	350	2.3	1.2	360	250	1.7	0.8	45.8
9	220	190	0.6	0.7	360	250	1.7	0.8	45.8
10	250	185	0.8	0.9	360	250	1.7	0.8	45.8

Note: x_1 denotes round disc diameter, x_2 denotes rotational speed of the steel roller , x_3 denotes feeding quantity, x_4 denotes length-width ratio, Y denotes power consumption.

TABLE 9. Statistical description of power consumption values for ten times test.

Method	Index	Power consumption value/kW
	Minimum	44.6
Test	Maximum	48.3
	Average	46.9
BPNN-LCOM	Optimal value	45.8
Relative error/%		2.4

Engineering Laboratory of Northeast Agricultural University in Harbin City, Heilongjiang Province, China. When the technological parameters of the bailing mechanism are:the diameter of the disc was 360mm, rotational speed of the steel roller was 250rpm, the feeding amount was 1.7 *kg/s* of and length-width ratio was 0.8, the power consumption value was repeatedly measured for ten times shown in Table.8.

The average power consumption of ten measurements was 46.9*kW*, the maximum value was 48.3 *kW*, and the minimum value was 44.6 kW. The absolute error was 1.1 *kW* and the relative error was 2.40% (less than 5%) compared with the theoretical power consumption obtained by the BPNN-LCOM. The error of the test results is within the allowable range, and the optimization results obtained by the BPNN-LCOM are accurate and reliable.

V. CONCLUSION
In this paper, an improved BP neural network-based optimization method was proposed for optimization problem with linear constraints, the theoretical and example verification showed the result is stability and accuracy better than the BPNN-COM. the gradient projection method was used to determine the search direction of iterative points on the boundary, it solved the imperfection that BPNN-COM

sometimes cannot obtain the authentic optimal value. The optimal step size, interpolation method, and the optimal constraint step size speed up the iterative speed.

The BPNN-LCOM was applied to optimize the parameters of roller-type baling mechanism, the optimal parameter combination of influencing factors of bailing mechanism under the test conditions is obtained as follows: disc diameter was 360mm, rotation speed steel roller was 250rpm, feeding amount was 1.7 kg/s, length-width ratio was 0.8. The minimum power consumption of the roller-type bailing mechanism under this parameter combination was 45.8 kJ/bundle, which is better than that obtained by regression analysis method and BPNN-COM. The optimization result was verified in the Agricultural Engineering Laboratory of Northeast Agricultural University in Harbin City, Heilongjiang Province, China. The experimental power consumption of the optimization method was 46.9 kJ/bundle, and the relative error between the experimental results and the theoretical results was 2.40%. The experimental results are consistent with the theoretical results. The verification results showed that the BPNN-LCOM was used to optimize the black box problem with linear constraints, and the optimization results have high accuracy and strong reliability. The application of this method in the parameters optimization of roller-type

baling mechanism has important practical significance for guiding the machine optimization design, and provides a new idea for solving similar optimization problems in the fields of scientific research and engineering application.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and Academic Editor for their valuable and constructive comments, which greatly improved the quality and integrity of this manuscript.

REFERENCES

[1] J. de Villiers and E. Barnard, "Backpropagation neural nets with one and two hidden layers," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 136–141, Jan. 1993.

[2] K.-I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Netw.*, vol. 2, no. 3, pp. 183–192, 1989.

[3] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 1, no. 3, 1989, pp. 593–605.

[4] W. M. Tang, "The study of the optimal structure of BP neural network," *Syst. Eng. Pract.*, vol. 10, pp. 95–100, Oct. 2005.

[5] G. L. Liu and F. L. Wang, "The research on optimization of bpnn model and its application," *Math. Pract. Theory*, vol. 6, pp. 161–166, Jun. 2010.

[6] F. Wang, G. Liu, and J. Wang, "An unconstrainted optimization method based on BP neural network," in *Proc. Int. Conf. E-Product E-Service E-Entertainment*, Nov. 2010, pp. 1–4.

[7] L. Zhang, F. Wang, T. Sun, and B. Xu, "A constrained optimization method based on BP neural network," *Neural Comput. Appl.*, vol. 29, no. 2, pp. 413–421, Jan. 2018.

[8] S.-J. Li and F.-L. Wang, "The research of constrained optimization method based on BP neural network and its application," *Int. J. Hybrid Inf. Technol.*, vol. 9, no. 10, pp. 313–324, Oct. 2016.

[9] W. Fulin, D. Zhigui, W. Zhihui, and F. Kun, "Optimization of maize planting density and fertilizer application rate based on BP neural network," *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 6, pp. 92–99, 2017.

[10] Z. Dong, F. Wang, Q. Song, Z. Wu, and K. Fang, "Performance optimization of combine harvester inertia separation chamber based on BPNN," *Int. Agricult. Eng. J.*, vol. 26, no. 3, pp. 19–26, 2017.

[11] Z. Dong, F. Wang, Q. Song, and H. Wang, "Optimization of soybean planting density and fertilizing amount based on BP neural network," *Int. Agricult. Eng. J.*, vol. 27, no. 3, pp. 393–402, 2018.

[12] Z. Dong, Q. Song, F. Wang, Z. Li, and Z. Wu, "Optimize on power dissipation of whole-straw returning device based on BP neural network," *Xitong Gongcheng Lilun Yu Shijian, System Eng. Theory Pract.*, vol. 38, no. 9, pp. 2401–2408, 2018.

[13] S. Zhao, Z. Dong, B. Zhang, L. Qi, and Y. Li, "Power consumption optimization of rice straw round baler based on BP neural network," *Int. Agricult. Eng. J.*, vol. 27, no. 2, pp. 162–168, 2018.

[14] L. Han, "Artificial neural network tutorial," in *Beijing University of Posts and Telecommunications*. Beijing, China, 2006, p. 26.

[15] M. Z. Cohn, "Theory and practice of structural optimization," *Struct. Optim.*, vol. 7, nos. 1–2, pp. 20–31, Feb. 1994.

[16] Y. Li, D. Wang, M. Wang, D. Li, and J. Lei, "Design and parameters optimization of roll-disk baling mechanism for rice straw round baler," *Nongye Gongcheng Xuebao, Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 6, pp. 27–34, 2017.

[17] J. Lei, D. Wang, Q. Zhang, X. Yang, and C. Li, "Experiment on stress relaxation characteristics of intact rice straw during rotary compression," *Trans. Chin. Soc. Agricult. Eng.*, vol. 31, no. 8, pp. 76–83, 2015.

[18] Y. Li, D. Wang, D. Li, M. Wang, and J. Lei, "Theoretical analysis and experiment of baling mechanism of roll-disk round baler," *Trans. Chin. Soc. Agricult. Machinery*, vol. 47, no. 12, pp. 45–52, 2016.

ZHIGUI DONG received the B.E. and M.E. degrees in industrial engineering and management science engineering and the Ph.D. degree in agricultural system engineering and management engineering from Northeast Agricultural University, China, in 2005, 2008, and 2018, respectively.

He is currently an Associate Professor with Liaoning Institute of Science and Technology. He has published over 20 articles in domestic and international academic journals and conference proceedings. His current research interests include neural network theory and its application, optimization algorithm theory and its application, big data technology and application, and agricultural big data.

CHANGYOU WU received the B.E. and M.E. degrees in industrial engineering and management science engineering and the Ph.D. degree in mechanized engineering from Northeast Agricultural University, China, in 2004, 2006, and 2009, respectively.

He is currently an Associate Professor with the Engineering College, Shandong Institute of Business and Technology. He has published over 30 articles in domestic and international academic journals and conference proceedings. His current research interests include genetic algorithm theory and its application, neural network theory and its application, parallel computing, and water resources management.

XISONG FU received the B.E. degree in information management and information system from the Institute of Technology, East China Jiaotong University, China, in 2019. He is currently pursuing the master's degree with the Engineering College, Shandong Institute of Business and Technology.

His current research interest includes brain storm optimization algorithm theory and its application.

FULIN WANG received the B.E. and M.E. degrees in agricultural mechanization and automation and the Ph.D. degree in agricultural equipment engineering technology from Northeast Agricultural University, China, in 1981, 1988, and 1993, respectively.

He is currently a Professor with the Engineering College, Northeast Agricultural University. He has published over 160 articles in domestic and international academic journals and conference proceedings. His current research interests include genetic algorithm theory and its applications, neural network theory and its applications, parallel computing, and image recognition.

• • •