Sales Forecasting Model Based on BP Neural Network Optimized by Improved Immune Genetic Algorithm

Kunyu Cao (✉ cao_kunyu@163.com)

Xi'an Key Laboratory of Modern Intelligent Textile Equipment, School of Mechanical and Electrical Engineering, Xi'an Polytechnic University

Yongdang Chen

Xi'an Key Laboratory of Modern Intelligent Textile Equipment, School of Mechanical and Electrical Engineering, Xi'an Polytechnic University

Xinxin Song

Xi'an Key Laboratory of Modern Intelligent Textile Equipment, School of Mechanical and Electrical Engineering, Xi'an Polytechnic University

Shan Liu

Xi'an Key Laboratory of Modern Intelligent Textile Equipment, School of Mechanical and Electrical Engineering, Xi'an Polytechnic University

Research Article

Keywords: BP neural network, grey relation analysis, immune genetic algorithm, principal component analysis, sales forecasting

Posted Date: June 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-598556/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Sales Forecasting Model Based on BP Neural Network Optimized by Improved Immune Genetic Algorithm

Yongdang Chen¹, Kunyu Cao^{1,*}, Xinxin Song¹⁺, and Shan Liu¹⁺

¹Xi'an Key Laboratory of Modern Intelligent Textile Equipment, School of Mechanical and Electrical Engineering, Xi'an Polytechnic University, Xi'an, 710600 China

*Corresponding author: K CAO (cao_kunyu@163.com).

+these authors contributed equally to this work

ABSTRACT A new sales forecasting model based on an Improved Immune Genetic Algorithm (IIGA), IIGA that optimizes the BPNN (IIGA-BP) has been proposed. The IIGA presents a new way of population initialization, a regulatory mechanism of antibody concentration, and a design method of adaptive crossover operator and mutation operator, which effectively improved the convergence ability and optimization anility of IIGA. And IIGA can optimize the BPNN's initial weights and threshold and improve the randomness of network parameters as well as the drawbacks that lead to output instability of BPNN and easiness into local minimum value. It taking the past records of sales figures of a certain steel enterprise as an example, utilizing the Matlab to construct the BP neural network, Immune Genetic Algorithm that optimizes the BPNN (IGA-BP), IGA-BP neural network, and IIGA-BP neural network prediction models for simulation and comparative analysis. The experiment demonstrates that IIGA-BP neural network prediction model possessing a higher prediction accuracy and more stable prediction effects.

INDEX TERMS BP neural network, grey relation analysis, immune genetic algorithm, principal component analysis, sales forecasting.

I. INTRODUCTION

With the development of enterprise market oriented process, the competitive pressure among enterprises and sales market's instability are growing increasingly. The way to reasonably and accurately predict the market demand has become of great importance for the modern enterprise to control production capacity, to avoid effective operational risks and to achieve precision marketing¹. Therefore, sales prediction is an integral part in modern enterprise management, in which the exactitude of prediction is directly related to the success of the company. In order to accurately predict the sales in the future, scholars home and abroad have once again thrown themselves into the study of sales forecasting model.

In recent years, scholars home and abroad have conducted a lot of researches for sales forecasting model, and have proposed several methods like neural network, extreme learning machine, machine learning, and time series model²⁻⁴. Neural network are widely used in the construction of sales forecasting model for its powerful adaptive learning and preferable nonlinear data fitting⁵. Besides, BPNN as the representative has gradually become the focus of scholars for its complete theoretical system. Such as: In view of sales of make to stock enterprise, Wu *et al.*⁶ build a sales model

based on BPNN by making use of its strong learning ability and powerful information storage. Qin *et al.*⁷, put forward BPNN sales forecasting model by selecting the network types based on the establishment of the network parameter optimization mechanism. By using the logsig activation function, Gao *et al.*⁸ made the BPNN based prediction model in terms of the elements that may affect the data collection and reducing dimension processing. Bibaudalves *et al.*⁹, orienting the data's seasonal quality and value proposed a BPNN prediction model with average monthly best structure. However, it has some drawbacks like it's easy to fall into local minimum value and very sensitive to the initial network weights, which are hard to overcome. In this regard, scholars have proposed many effective methods for optimizing BPNN. Such as: Luo *et al.*¹⁰ used the genetic algorithm to determine the BPNN's network structure and weights search space. Then they integrated the advantages of two algorithms to construct a neural network prediction model with preferred performance. Zhang *et al.*¹¹, proposed an improved bacterial chemotaxis optimization and integrated it into the BPNN to develop a prediction model that used to predict a variety of stock index. Sheng *et al.*¹², took the self calibration of synchronous time series as the reference and took advantages of genetic algorithm optimizing network parameters. They

constructed the neural network sales prediction model that combined the BPNN, time series forecasting model, and genetic algorithm together. According to the inner rules and development trend of the stock market, Liu¹³ made the index prediction of it come to be true by the improved BPNN. Cao *et al.*¹⁴, proposed a vehicle sales forecasting model based on grey neural network with the combination of the advantage that the grey model deals with the small simple data and finding regularity of weak correlation data. Ji *et al.*¹⁵, by using Adaptive particle swarm optimization (APSO) to update weights and thresholds of BPNN raised the agricultural machinery proposed prediction model based on data APSO-BP algorithm. Zhang *et al.*¹⁶, optimized network parameters by using particle swarm optimization and reduced the dimension of the input neurons with Johnson algorithm to propose BPNN prediction model based on data mining. Wang *et al.*¹⁷, came up with an improved BPNN by using cuckoo search algorithm (CS-BP), which the CS-BP model compared with traditional BPNN and multivariable linear regression has a good performance in forecasting lightning. Zha *et al.*¹⁸, using immune algorithm to optimize BPNN's initial weights and thresholds, proposed BPNN prediction method based on Immune Genetic Algorithm (IGA). In order to establish the accurate inventory prediction model, Cao *et al.*¹⁹ adopted the principal component analysis to analyze the index data, used bayesian regularization algorithm to optimize BPNN, which is effective to avoid the overfitting phenomenon during the model training process. Li *et al.*²⁰, with the aim to improve the accuracy and speed of power loaded prediction, proposed a algorithm that mixed the particle swarm optimization and BPNN (PSO-BP), and used the PSO-BP algorithm to predict the power load data. According to the importance of financial markets and the timing characteristics of financial indicators, Lu²¹ improved the weights back propagation model and proposed timing weights back propagation model.

The above literature shows that BPNN has achieved good results in sales forecasting. However, because traditional BPNN uses the principle of gradient descent to adjust network parameters, it has the disadvantages of slow convergence speed and easiness to fall into local minimum value, and the random selection of the algorithm's initial weight and threshold made neurons lack the ability to adjust. When initializing the weights of the neural network, if a random initial is used, two situations are likely to occur²²⁻²⁶:

(1) The result of initialized weights are the same. Then the symmetrical structure will be destroyed. In the forward propagation of the network, the output value of the neural network node of each layer is the same, so when it transfers backward, the partial derivative of the weight is also the same, which will make the ownership values of nodes of each layer to the previous ones are all the same, but the weights layers by layers will not be the

same. In addition, initializing the weights to the same value will make the network convergence very slow.

(2) All weights are initialized to 0. Then the weights cannot be updated. This is because in the forward propagation of the network, the output values of all nodes are the same, and when the output value of each node in the backward propagation seeks the partial derivative of the loss function, result must be 0. Then it is impossible to use gradient descent method to update the weights, nor will training of natural neural network not be able to proceed.

Based on this, this paper proposes an effective neural network sales forecasting model that integrates IIGA and traditional BPNN model. The premature convergence problem of unique population initialization method, antibody concentration adjustment mechanism and crossover operator, mutation operator in the IIGA are utilized to initialize the weights and thresholds of each connection layer of BPNN, and improve the randomness of network parameters that cause the BP neural network the shortcomings of unstable output and easiness to fall into the local extremes. Therefore, the convergence speed of the algorithm and the prediction accuracy of the model will be improved.

II. IMPROVED IMMUNE GENETIC ALGORITHM TO OPTIMIZE BP NEURAL NETWORK DESIGN

A. BP NEURAL NETWORK

BPNN is a multiple layer feed forward neural network with one or several implied layers, which its strong nonlinear data fitting ability is the mathematical model to solve the uncertain problem. It is the most widely used neural network by now through its using the back propagation algorithm based on the principle of gradient descent for supervising learning to get pattern recognition and approximation of functions²⁷. The most typical three layer neural network structure can realize the highly nonlinear mapping of any function's distribution parallel information processing from input to output. Through the forward transmission of data information and the back propagation of error, the weights and thresholds are adjusted according to certain rules to correct the network error until the output error or training times meet the predetermined conditions, as shown in Fig. 1.

FIGURE 1. BPNN structure. The network structure has a hidden layer, the number of input nodes is x_m , and the number of output nodes is y_n ; ω_{ij} and ω_{jk} are the weight vectors between the connected layers; θ_j and ζ_k are the threshold vectors of the hidden layer and the output layer.

B. IMPROVED IMMUNE GENETIC ALGORITHM TO OPTIMIZE BP NEURAL NETWORK PRINCIPLE

Being a new improved and refined algorithm, IGA has the immune function, which considers both the search speed and ability by adding the genetic operator into the immune operator whereas it also avoids the degeneracy phenomenon and premature convergence of genetic algorithm. It encodes the parameters that need to be optimized to form a series population. Through immune genetic evolution operation, it evolves the population to a better search space, so that the individuals in the population reach the optimal²⁸. The IGA has outstanding global search ability and population diversity maintenance mechanism which just make up for the deficiency of BPNN. However, the standard IGA still has the problems of slow convergence speed and local minimum value in dealing with complex optimization problems. Therefore, the IIGA is considered to optimize the BP network.

In this paper, the IIGA is used to optimize BPNN from the following three aspects:

- (1) Determine the network topology structure.
- (2) Optimize the initial weights and thresholds of BPNN.

Guided by the fitness function of the optimal antibody in the population, the IIGA is used to search for the optimal solution and a group of network parameters with the highest global fitness are obtained as the initial weights and thresholds.

- (3) Do a simulation prediction.

The obtained initial weights and thresholds are used in the BPNN again, to find the optimal weights and thresholds during the training period, and then the optimal prediction output is realized.

C. PROCESS OF BP NEURAL NETWORK OPTIMIZED BY IMPROVED IMMUNE GENETIC ALGORITHM

The following is the flow that used the IIGA to optimize the BPNN algorithm, as shown in Fig. 2.

The specific steps are as follows:

- (1) Determine the network topology structure.
- (2) Antigen recognition.

The target function of antigen's corresponding problem: the mean square error of the network prediction is the least.

$$E(x) = \min E = \min \frac{1}{2m} \sum_{i=1}^m \sum_{j=1}^n (Q_{ij} - Y_{ij})^2 \quad (1)$$

where m is the amount of the samples, n is the total number of output nodes; Q_{ij} and Y_{ij} are the expect output results and predictable results of the neural network respectively which the i antibody corresponds.

- (3) Initial antibody production.

In order to avoid the disadvantages of randomization of initial weights and thresholds and low computational efficiency, the convergence speed of the algorithm is accelerated whereas ensuring the diversity of population. In this paper, the Nguyen Widrow method²⁹ is used to generate 70% of the initial population, and the remaining 30% is randomly generated by multiple evolution. In each generation of antibody group, only some antibodies with higher fitness were retained, and the antibody with weak fitness was replaced in the next evolution process, which was a cycle until the randomly generated population size was reached.

the Nguyen Widrow method:

$$w = 0.7 \times s^{1/r} \text{normr}(2\text{rand}(s, r) - I(s, r)) \quad (2)$$

$$\theta = 0.7 \times s^{1/r} \text{normr}(2\text{rand}(s, 1) - I(s, 1)) \quad (3)$$

where w is the weight matrix, θ is the threshold matrix; s and r are the number of nodes in the hidden layer and the dimension of input index respectively; $\text{rand}(s, r)$ is random matrix that is well distributed in the row s and the line r ; $I(s, l)$ is the whole 1 matrices in row s and line r ; $\text{normr}(J)$ is the matrix that the J matrix is developed into 1.

- (4) Antibody coding.

In this paper, real coding is used to encode the initial connection weights and thresholds of neural networks, and the random distribution of gene code strings is obtained. Each gene code string corresponds to a set of weight threshold.

The length of the antibody code is:

$$S = n \times m + m \times l + m + l \quad (4)$$

where n , m , and l are nodes of input layer, hidden layer and output layer respectively.

- (5) Affinity calculation.

- ① The affinity between antibody and antigen.

The affinity between the antibody and the antigen is the criterion for determining the quality of the antibody. The reciprocal of the mean square error of the training data is

used as the original affinity function of the IGA, that is the fitness function.

The fitness function is as follows:

$$f(x) = \frac{1}{\alpha + E(x)} \quad (5)$$

where α is the small positive quantities that is to avoid zero denominator; $E(x)$ is the objective function of the network.

In order to prevent the initial stage of evolution, special antibodies in the initial antibody population mislead the population's development direction. This paper uses a linear adjustment strategy to adjust the fitness function³⁰.

$$f' = \beta \frac{f + f_{\min}}{f_{\max} + f_{\min}} \quad (6)$$

where $f_{\max}$ and $f_{\min}$ are the largest and smallest fitness values in the contemporary population respectively; f is the current fitness values of the antibody; f' is the revised antibody fitness values; β is the adjustment coefficients.

② Affinity between antibody and antibody.

This article uses euclidean distance as an important indicator to measure the similarity between antibodies. The Euclidean Distance between antibody $v = \{v_1, v_2, \dots, v_n\}$ and antibody $r = \{r_1, r_2, \dots, r_n\}$ is:

$$d_v = \sum_{1 \leq i \leq n} \sqrt{(v_i - r_i)^2} \quad (7)$$

The affinity function is as follows:

$$S_{v,r} = \frac{1}{1 + d_v} \quad (8)$$

where $S_{v,r} \in (0,1)$ reflects the degree of similarity between antibodies v and antibodies r .

The antibody concentration function is as follows:

$$C_v = \frac{1}{N} \sum_{r=1}^N S_{v,r} \quad (9)$$

where C_v reflects the proportion of similar antibodies in the response population; N is the total number of antibodies;

$S_{v,r} = \begin{cases} 1 & \alpha \cdot S_r \leq S_v \leq \beta \cdot S_r \\ 0 & \text{other} \end{cases}$, α and β are the

adjustment coefficient. α is a little less than 1, so 0.9 to 1 is usually taken, and β is a little more than 1, so 1 to 1.1 is usually taken. In this way, the calculation of antibody concentration considers both the same antibody and similar antibodies.

(6) Population regulation based on concentration.

In order to avoid simple selection operations, antibodies with high fitness are gradually accumulated, resulting in excessively high antibody concentration. According to the principle of mutual inhibition of antibodies due to different concentrations in the immune mechanism, the introduction of concentration regulators adjusts the probability of antibodies being selected P_s . For any antibody v_i , v_j , if

$|f(v_i) - f(v_j)| \leq \xi, \xi > 0$, the probability of the antibody v_i being selected is:

$$P_s(v_i) = \alpha C_i \left(1 - \frac{f(v_i)}{\text{Maxf}(v)} \right) + \beta \frac{f(v_i)}{\text{Maxf}(v)} \quad (10)$$

where α and β is the variable parameter between 0 to 1; C_i is the concentration of antibody; $f(v_i)$ is the fitness value of the antibody v_i ; $\text{Maxf}(v)$ is the maximum fitness value in the population. This not only retains antibodies with high adaptability, but also inhibits antibodies with high concentrations, maintains the diversity of the population, and improves the convergence ability of the algorithm.

(7) Immune manipulation.

① Clonal selection.

The clone selection is performed according to the strategy of combining elite retention by the combined way of maintaining the best solution found over time before selection and roulette. In this way, every time the memory bank is updated, the antibodies with higher fitness are stored first, and then the selection probability is calculated remanent referring to "(10)," and the excellent antibodies in the population are stored in the memory bank.

② Adaptive crossover and mutation.

In order to make the algorithm's crossover and mutation probabilities to be adjusted by itself in terms of the degree of antibody fitness in the population and evolutionary algebra, the crossover and mutation probabilities need to be parameterized.

Crossover probability P'_c , mutation probability P'_m are:

$$P'_c = \begin{cases} \frac{P_c(T-t) \sin\left(\frac{\pi}{2} + \frac{f' - f_{\text{avg}}}{f_{\max} - f_{\text{avg}}}\right)}{T} & f' > f_{\text{avg}} \\ \frac{P_c(T-t) \sin\left(\frac{\pi}{2} \frac{f_{\text{avg}} - f'}{f_{\text{avg}} - f_{\min}}\right)}{T} & f' \leq f_{\text{avg}} \end{cases} \quad (11)$$

$$P'_m = \begin{cases} \frac{P_m(T-t) \sin\left(\frac{\pi}{2} + \frac{f - f_{\text{avg}}}{f_{\max} - f_{\text{avg}}}\right)}{T} & f > f_{\text{avg}} \\ \frac{P_m(T-t) \sin\left(\frac{\pi}{2} \frac{f_{\text{avg}} - f}{f_{\text{avg}} - f_{\min}}\right)}{T} & f \leq f_{\text{avg}} \end{cases} \quad (12)$$

where f' is the maximum fitness of the antibody to be crossed; $f_{\max}$ is the maximum fitness of the current population; f_{avg} is the average fitness of the current population; $f_{\min}$ is the minimum fitness of the current population; f is the fitness of the antibody to be mutated; t is the current evolution algebra of the algorithm; T is the total evolution algebra of the algorithm; P_c and P_m are the initial value of the crossover and mutation probability.

(8) Determination of initial weights and thresholds of network.

It means to take the fitness of the optimal antibody in the population as the guide cycle to perform operations such as

evaluation, selection, crossover, and mutation, etc. until the set evolutionary algebra is reached and select a group of network parameters with the highest adaptability as initial weights and thresholds.

(9)After obtaining the initial weights and thresholds, use the BP algorithm to learn and continuously adjust until the established network error is met or the maximum evolution algebra is reached, and finally the optimal weights and thresholds are obtained, and the prediction output is achieved.

FIGURE 2. Flow chart of IIGA optimized BPNN algorithm. It is firstly to improve the population initialization of the IGA, the regulation mechanism, and crossover operator of antibody concentration as well as the adaptive adjustment strategy of mutation operator. Then it regards the fitness function value of the optimal antibody in the population as the guide to conduct global search and locate the parameter space of the optimal solution. And the local search ability of BP algorithm through bayesian function's fine tune to the network weights and thresholds is used to make the network output increasingly close to the expect results and finally the predicted value is obtained by antinormalization.

III. ANALYSIS OF MAIN INFLUENCING FACTORS OF STEEL SALES

A. IDENTIFICATION OF INFLUENCING FACTORS OF STEEL SALES

The sales status of steel is affected by various complex factors. Taking the historical sales data of a steel company as an example, select from the many factors that affect steel sales the Gross domestic product (X_1), Steel export volume

(X_2), Steel imports volume (X_3), Real estate development investment (X_4), Metal cutting machine tool output (X_5), Tractor output (X_6), Total social fixed asset investment (X_7), Vehicle output (X_8), and Total output value of the national construction industry (X_9) are used as primary indicators. The nine selected influential data are selected from the 1994 to 2018 China Statistical Yearbook, as shown in Table 1.

Year	Gross domestic product (10 ⁸ yuan)	Steel export volume (10 ⁴ tons)	Steel import volume (10 ⁴ tons)	Real estate development investment (10 ⁸ yuan)	Metal cutting machine tool output (10 ⁴ sets)	Tractor output (10 ⁴ sets)	Total social fixed asset investment (10 ⁸ yuan)	Vehicle output (10 ⁴ vehicle)	Total output value of the national construction industry (10 ⁸ yuan)	Steel sales volume (10 ⁴ tons)
1994	48637.50	151.00	2245.00	2554.00	20.65	4.67	17042.10	163.56	4653.32	3101.56
1995	61339.90	567.00	1397.00	3149.00	20.34	6.33	20019.30	178.97	5793.75	3311.23
1996	71813.60	422.00	1599.00	3216.40	17.74	8.37	22974.00	185.81	8282.25	3448.45
1997	79715.00	462.00	1323.00	3178.40	18.65	8.24	24941.10	206.85	9126.48	3681.89
1998	85195.50	357.00	1242.00	3614.20	11.91	6.78	28406.20	213.71	10061.99	4070.25
1999	90564.40	368.00	1486.00	4103.20	14.22	6.54	29854.70	240.30	11152.86	4503.76
2000	100280.10	621.00	1596.00	4984.10	17.66	4.10	32917.70	267.70	12497.60	4885.48
2001	110863.10	474.00	1722.00	6344.10	25.58	3.82	37213.50	304.53	15361.56	6104.69
2002	121717.40	545.00	2449.00	7790.90	30.86	4.54	43499.90	434.30	18527.18	7205.60
2003	137422.00	696.00	3717.00	10153.80	30.58	4.88	55566.60	651.47	23083.87	9161.04
2004	161840.20	1423.00	2930.00	13158.30	48.72	11.38	70477.40	736.74	29021.45	11140.66
2005	187318.90	2052.00	2582.00	15909.25	51.14	16.33	88773.60	847.50	34552.10	13352.03
2006	219438.50	4298.00	1850.00	19422.92	57.30	19.93	109998.20	1114.83	41557.16	16719.84
2007	270092.30	6265.00	1687.00	25288.84	64.69	20.31	137323.90	1368.67	51043.71	21496.63
2008	319244.60	5923.00	1543.00	31203.19	71.73	28.44	172828.40	1434.40	62036.81	22899.96
2009	348517.70	2460.00	1763.00	36241.81	58.55	37.13	224598.80	2128.01	76807.74	25989.76
2010	412119.30	4256.00	1643.00	48259.40	69.73	33.68	251683.80	2784.12	96031.13	30798.90
2011	487940.20	4888.00	1558.00	61796.89	88.68	40.19	311485.10	2854.31	117059.65	33675.41
2012	538580.00	5573.00	1366.00	71803.79	88.23	52.73	374694.70	3004.62	137217.86	35916.98
2013	592963.20	6233.00	1408.00	86013.38	87.55	66.56	446294.10	3422.52	159312.95	41126.31
2014	641280.60	9378.00	1443.00	95035.61	85.80	64.37	512020.70	3620.83	176713.42	42579.88
2015	685992.90	11240.00	1278.00	95978.85	75.50	68.82	561999.80	3613.32	180757.47	39308.26
2016	740060.80	1083.00	1322.00	102580.61	67.28	61.84	606465.70	4023.03	193566.78	39829.11
2017	820754.30	7541.00	1330.00	109798.53	60.85	34.44	641238.40	4096.35	213943.56	39874.46
2018	900309.50	6933.00	1317.00	120263.51	48.86	24.35	645675.00	3941.97	235085.53	42036.63

TABLE 1. Data of steel sales and influencing factors of a steel company.

B. GREY RELATION ANALYSIS

Grey relational analysis is a method to judge the correlation between factors in the system by weighing the similarity of the sequence curves of various factors in the same coordinate system. It has no special requirements for the sample size and the regularity among samples, and has the characteristics of small amount of calculation and high precision. It is a new method to study the small sample and poor information analysis and prediction³¹. In this paper, the grey factor correlation analysis method is used to identify the main influencing factors of steel sales volume, and find out the correlation degree between the input factors and the output volume, so as to extract the main control factors affecting the output index.

The steps of grey relation analysis are as follows:

(1) Determine reference sequence and comparison sequence.

According to research needs, take the steel sales volume from 1994 to 2018 in Table 1 as the dependent variable Y , and record the reference sequence as $Y = \{y(k) | k = 1, 2, \dots, n\}$; the 9 steel sales influencing factors as the independent variable groups, and the 9 sets of comparison sequences as $X_i = \{x_i(k) | k = 1, 2, \dots, n\}, i = 1, 2, \dots, m$.

(2) Nondimensionalization of data.

In order to avoid large prediction error caused by too large order of magnitude difference, dimensionless processing is needed before data calculation. This paper adopts the Mean

conversion method to disabuse the dimension to clear up the dimension differences of data. And the calculating equation is:

$$x'_i(k) = \frac{x_i(k)}{\frac{1}{m} \sum_{k=1}^n x_i(k)} \quad (13)$$

(3) Calculate the relationship coefficient between each factor and steel sales volume.

$$\eta_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)} \quad (14)$$

where $\Delta_i(k) = |y(k) - x_i(k)|$; the resolution coefficient ρ is generally taken 0 to 1.

(4) Calculate the correlation degree between each factor and steel sales volume.

$$r_i = \frac{1}{n} \sum_{k=1}^n \eta_i(k) \quad (15)$$

The above process is based on the Matlab 2018a platform to write a calculation program, the resolution coefficient ρ is set to 0.5, then the gray correlation analysis results are shown in Table 2.

Influencing factor	Grey relational degree
Gross domestic product (X_1)	0.9399
Steel export volume (X_2)	0.8583
Steel import volume (X_3)	0.6284
Real estate development investment (X_4)	0.6214
Metal cutting machine output (X_5)	0.7666
Tractor output (X_6)	0.8416

Total social fixed asset investment (X_7)	0.8757
Vehicle output (X_8)	0.9390
Total output value of the national construction industry (X_9)	0.9638

TABLE 2. Gray relation degree table of influencing factors and Steel.

According to the calculated relational degree results in Table 2, the threshold is equal to 0.7, and the factors greater than 0.7 are used as the selected indicators. The order of gray correlation is: Total output value of the national construction industry (X_9) > Gross domestic product (X_1) > Vehicle output (X_8) > Total social fixed asset investment (X_7) > Steel export volume (X_2) > Tractor output (X_6) > metal cutting machine tool output (X_5) > 0.7.

C. PRINCIPAL COMPONENT ANALYSIS

It can be seen from Table 1 that the original data is multidimensional data. Due to the coupling between the data, not all the data has a high contribution rate to the sales of steel. If the indicators selected using grey relation analysis are directly used as the input nodes of the prediction model, due to the complexity and redundancy of the experimental data, the performance of the network and the prediction accuracy of the model will inevitably be affected³². For this reason, this paper chooses the principal component analysis method to perform dimensionality reduction processing on the indicators selected by the gray correlation analysis, and determines the input indicators to reduce the coupling relationship between the variables and retain the data that is valuable for the prediction results.

The steps of principal component analysis are as follows

(1) Establish a correlation matrix R .

Assuming that the input original data set is X and each sample corresponds to P influencing factors, i.e. $X = (X_1, X_2, \dots, X_p)$. Then the Correlation Matrix R is:

$$R = (r_{ij})_{p \times p} \quad (16)$$

$$r_{ij} = \text{cov}(x_i, x_j) = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i) \sum_{k=1}^n (x_{kj} - \bar{x}_j)}{n-1} \quad (17)$$

where r_{ij} is the correlation coefficient between factor x_i and factor x_j , $i, j = 1, 2, \dots, p$, n is the total number of samples; $\bar{x}$ is the average of x .

(2) Calculation correlation matrix R eigen value of $\lambda_1, \lambda_2, \dots, \lambda_p$ and the corresponding eigen vector $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ip})$.

(3) Determine the number of principal components m ($m < p$).

Generally speaking, when the cumulative contribution rate is reached 75% to 95%, the corresponding previous m principal component can contain most of the information in

the original data set. Variance contribution rate and cumulative variance contribution rate are:

$$\eta_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad (18)$$

$$\eta(p) = \sum_{i=1}^m \eta_i \quad (19)$$

(4) Calculate the principal component score.

$$Y = (f_{bv})_{n \times m} \quad (20)$$

$$f_{bv} = \alpha_{v1}x_{b1} + \alpha_{v2}x_{b2} + \dots + \alpha_{vp}x_{bp} \quad (21)$$

where $b = 1, 2, \dots, n$, $v = 1, 2, \dots, m$.

The above process is based on the Matlab 2018a platform to write a calculation program, and the princomp function is used to obtain the contribution rate of each indicator factor, as shown in Table 3.

Principal component	Eigen value	Contribution rate (%)	Cumulative contribution rate (%)
1	6.0981	87.1159	87.1159
2	0.5126	7.3231	94.4390
3	0.2682	3.8320	98.2710
4	0.1097	1.5668	99.8378
5	0.0082	0.1177	99.9555
6	0.0019	0.0270	99.9825
7	0.0012	0.0174	100.0000

TABLE 3. Eigen value and Variance contribution rate of influencing factors.

It can be seen from Table 3 that the first three principal components already contain more than 95% of the information of all influencing factors. Therefore, the first three principal components are selected to replace the factors selected by the original gray correlation analysis as the final input indicators. The score value of each principal component, as shown in Table 4.

Year	First principal component	Second principal component	Third principal component
1994	-2.6038	-0.1997	-0.0343
1995	-2.5023	-0.1639	0.0406
1996	-2.4830	-0.2430	-0.0127
1997	-2.4440	-0.2371	-0.0087
1998	-2.5474	-0.4471	0.0603
1999	-2.4968	-0.4104	0.0450
2000	-2.4314	-0.3587	0.1279
2001	-2.3001	-0.2258	0.0206
2002	-2.1332	-0.1324	-0.0363
2003	-1.9896	-0.1988	-0.0190
2004	-1.4505	0.3185	-0.1459
2005	-1.1421	0.4198	-0.1178
2006	-0.5703	0.7382	0.3011
2007	-0.0188	0.9687	0.7112
2008	0.3801	1.0403	0.3916
2009	0.3765	0.1982	-0.6048
2010	1.0723	0.3459	-0.2263
2011	1.8387	0.6743	-0.3690
2012	2.4422	0.6479	-0.4543
2013	3.1632	0.5759	-0.5956
2014	3.7665	0.5977	0.2633
2015	4.0745	0.4729	0.7509
2016	3.0988	-1.1998	-1.6321
2017	3.5499	-1.2541	0.6671
2018	3.3506	-1.9273	0.8771

TABLE 4. Principal component score table.

IV. SIMULATION EXPERIMENT

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTING

In order to fully verify the performance of the algorithm, a steel sales data set of a steel company is used as an example for verification. Taking the three principal component scores obtained by principal component analysis as a new sample, among the 25 data of existing steel sales statistics, 18 are randomly selected as the simulation training set, and the remaining 7 are used as the simulation verification set. Used to verify the error. The algorithm program is implemented on the basis of MATLAB 2018a, and the simulation test is carried out on a personal computer with an Intel(R)Core(TM)i5-5200U CPU, a RAM of 4 GB, and a running environment of Windows 10 (64-bit). BPNN hidden layer transfer function adopts tansig function; output layer transfer function adopts purlin function; training function adopts trainlm function; activation function adopts Sigmoid function.

To determine the number of hidden layer nodes, empirical formulas are usually used to calculate the range of nodes, and then BPNN trial calculations are performed to determine the exact number of nodes.

$$q = \sqrt{m + n + a} \quad (22)$$

where q , m , and n are the number of nodes in the hidden layer, the input layer, and the output layer; a is generally taken 1 to 10 as the adjustment factor. By continuously increasing the number of hidden layer nodes and taking the minimum network output error as the standard, the optimal number of hidden layer nodes is obtained. After repeated adjustments, when the number of hidden layer nodes is 10, the output error is the smallest, and the network topology is 8-10-1.

The experimental parameters are shown in Table 5 and table 6.

Maximum training times	Expected error	Learning rate	Step interval
1000	0.0001	0.1	5

TABLE 5. Parameter setting of BPNN.

Population size	evolutionary algebra	Initial crossover probability	Initial mutation probability
100	100	0.75	0.2

TABLE 6. Parameter setting of IGA.

B. NETWORK PERFORMANCE ANALYSIS

The evolution curve of fitness function of IGA model and IIGA model, as shown in Fig. 3.

FIGURE 3. Fitness function evolution curve. (a) is IGA model fitness function evolution curve, (b) is IIGA model fitness function evolution curve.

It can be seen from Fig. 3 that after 54 generations of the IIGA model evolved, the fitness function value of the optimization result gradually stabilized, about 2.2961, indicating that every individual is in Near the optimal solution. IGA model fitness function value gradually stabilized at around 1.7294 after the 76th generation, and the curve has oscillation phenomenon in the later stage of evolution. It shows that the IIGA model is better than the classic IGA model in optimizing network parameters.

When the target error is 10^{-4} , the three network model training error curves of BP, IGA-BP, and IIGA-BP are shown in Fig. 4.

It can be seen from Fig. 4 that the IIGA-BP algorithm has the fastest convergence speed and the steepest curve in the early stage of training. After getting rid of the local optimal solution around the 5th generation, it quickly enters a stable state. The design of adaptive adjustment strategy of crossover operator and mutation operator in this article is related to the design. Moreover, the IIGA-BP neural network reaches the target error after 23 evolution, the IGA-BP neural network reaches the target error after 65 evolution, and the BPNN has a training error after 100 evolution. Converges to 6.253×10^{-3} and falls into the local optimum and cannot escape. It can be seen that the BPNN optimized by the IIGA can increase the convergence speed of the network and reduce the possibility

of the BPNN falling into a local minimum value which has a strong superiority.

FIGURE 4. Training error curves of different network models.

C. PREDICTION ACCURACY ANALYSIS OF ALGORITHM MODEL

In order to further compare the performance differences of the three models and verify the accuracy and applicability of the prediction model, 7 sets of data with years 1995, 1996, 1997, 2002, 2004, 2009, 2010 in Table 4 as verification samples are randomly selected. and the remaining 18 sets of data are as training samples. The BP, IGA-BP, and IIGA-BP neural network models are conducted with performance simulation experiments, and run them independently for 20 times to obtain the simulation results shown in Figure 5 and Figure 6. The predicted values of the three network models are compared with the true values, as shown in Table 7.

FIGURE 5. Predicted outputs of different network models.

FIGURE 6. Relative error curves of the network model prediction.

From Fig. 5 and Fig. 6, we can see that the BP neural network model and IGA-BP neural network model predict the dispersion and the error fluctuation range are relatively large; the IIGA-BP neural network model predicts the dispersion is small, the error is basically maintained Between $[-0.25, 0.25]$, the prediction effect is more stable, and the prediction accuracy is greatly improved.

Year	Actual value (10 ⁴ tons)	Predictive value by BP (10 ⁴ tons)	BP average error (10 ⁴ tons)	Predictive value by IGA-BP (10 ⁴ tons)	IGA-BP average error (10 ⁴ tons)	Predictive value by IIGA-BP (10 ⁴ tons)	IIGA-BP average error (10 ⁴ tons)
2004	11140.66	11246.90	106.24	11548.39	407.73	11823.16	682.50
1997	3681.89	5039.38	1357.49	5404.13	1722.24	4572.67	890.78
2010	30798.90	15368.12	15430.78	23763.54	7035.36	31544.44	745.54
2009	25989.76	7564.83	18424.93	17948.37	8041.39	24061.47	1928.29
1995	3311.23	5013.78	1702.55	5386.53	2075.30	3969.28	658.05
2002	7205.60	6193.65	1011.95	7959.45	753.85	7740.85	535.25
1996	3448.45	4776.73	1328.28	5959.06	2510.61	4435.11	986.66

TABLE 7. Comparison table of predicted and actual values of different network models.

In order to display the prediction accuracy of the three models more intuitively, the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Nash-Sutcliffe efficiency coefficient (NSE) are used to further evaluate the error indicators of the three network models.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad (23)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (24)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (25)$$

where y_i and $\hat{y}_i$ are the expected output value and the predicted output value respectively; $\bar{y}$ is the average value of the measured value; N is the number of predicted samples.

The prediction accuracy comparison of the three models is shown in Table 8

Neural Network	MAPE	RMSE	NSE
BP	0.3754	9142.7421	0.2596
IGA-BP	0.3574	4284.0378	0.8374
IIGA-BP	0.1372	1016.0461	0.9909

TABLE 8. Comparison table of prediction accuracy of different network models.

It can be seen from Table 7 that the MAPE of the IIGA-BP neural network model is 0.1372, which is 23.82% higher than the prediction accuracy of the BP neural network model, which is also better than the prediction accuracy of the IGA-BP neural network model. Compared with the other two models, the RMSE value of the IIGA-BP network is the smallest, the error of the model is small, and the generalization performance is strong. The NSE of the IIGA-BP neural network is 0.9909. Compared with the other two models, this model has better nonlinear data fitting ability and higher prediction stability.

V. CONCLUSION

This article focuses on the shortcomings of BPNN in sales forecasting. An IIGA is proposed to optimize the sales forecast model of BPNN. The IIGA proposes a new way of population initialization, antibody concentration, adjustment mechanism, crossover operator, mutation operator, adaptive strategy design method, which guarantees the diversity of the population and improves the convergence ability and global search ability of the algorithm. The IIGA is used to expand the weights search space of the BP network, obtain the optimal weights and thresholds, and improve the shortcomings of the BP network that it is easy to fall into the local minimum value. Finally, through the construction of IIGA neural network steel sales forecasting model and horizontal evaluation with BP and IGA-BP neural network prediction models, experiments show that the BPNN optimized by the IIGA has strong nonlinear data fitting And generalization performance, whereas improving model prediction accuracy and convergence speed, it also simplifies the network structure, reduces the evolutionary algebra, the stability of the prediction output is better. However, due to the lack of network training data and imperfect network training, the model prediction accuracy is not too high. For in depth research, the training data set can be further expanded, the learning effect of the model can be strengthened, and the prediction accuracy can be improved.

REFERENCES

[1] S. S. Crawford, Y. Huang, N. Li, and Z. Yang, "Customer Concentration and Public Disclosure: Evidence from Management

Earnings and Sales Forecasts," *Contemporary Accounting Research*, vol.37, no. 1, pp. 131–159, Apr. 2019, DOI:10.1111/1911-3846.12526.

[2] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, Mar. 1998, DOI:10.1016/S0169-2070(97)00044-7.

[3] A. Martinez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *European Journal of Operational Research*, vol. 281, no. 3, pp. 588–596, Mar. 2020, DOI:10.1016/j.ejor.2018.04.034.

[4] X. Zhou, H. Xu, J. Meng, and X. Su, "Research Progress of Garment Sales Forecasting Methods," *Knitting Industries*, no. 3, pp. 68–72, Mar. 2020, DOI:10.1016/j.ejor.2018.04.034.

[5] C. Frank, A. Garg, L. M. Szandera, and A. Raheja, "Forecasting women's apparel sales using mathematical modeling," *International Journal of Clothing Science and Technology*, vol. 15, no. 2, pp. 107–125, Apr. 2003.

[6] Z. Wu, W. Wang, and J. Zhou, "Application of BP neural network to sale forecast for make to stock enterprise," *Industrial Engineering Journal*, vol. 13, no. 1, pp. 105–107, Feb.2010.

[7] Y. Qin, and H. Li, "Sales forecast based on BP neural network," in *Proc. IEEE International Conference on Communication Software and Networks*, Xian, China, May. 2011, pp. 186–189, DOI:10.1109/ICCSN.2011.6014419.

[8] W. Gao, F. Li, and Y. Niu, "Prediction of house price based on BP neural network," *Journal of Yanan University(Natural Science Edition)*, vol. 37, no. 3, pp. 37–40, Sep.2018.

[9] J. Bibaudalves, P. Thomas, and H. B. E. Haouzi, "Demand Forecasting using Artificial Neuronal Networks and Time Series: Application to a French Furniture Manufacturer Case Study," in *Proc. International Conference on Neural Computation Theory and Applications*, Vienna, Austria, vol. 2, Sep. 2019, pp. 502–507, DOI: 10.5220/0008356705020507.

[10] R. Luo, S. Liu, and C. Su, "Garment sales forecast method based on genetic algorithm and BP neural Network," *Journal of Beijing University of Posts and Telecommunications*, vol. 37, no. 4, pp. 39–43, Aug.2014.

[11] Y. Zhang, and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network," *Expert Systems With Applications*, vol. 36, no. 5, pp. 8849–8854, Jul.2009, DOI:10.1016/j.eswa.2008.11.028.

[12] W. Sheng, H. Zhao, and Y. Sun, "Research on sales forecasting model based on optimized BP Neural Network," *Journal of Liaoning University of Technology(Natural Science Edition)*, vol. 40, no. 2, pp. 117–120, Apr.2020.

[13] X. Liu, and X. Ma, "Based on BP Neural Network Stock Prediction," *Journal of Curriculum and Teaching*, vol. 1, no. 1, pp. 45–50, May.2012.

[14] S. Cao, and Y. Zhu, "Research on Prediction Model of Car Sales Based on Grey Neural Network," *Advanced Materials Research*, vol. 889–890, pp. 1569–1573, Feb.2014, 10.4028/www.scientific.net/AMR.889–890.1569.

[15] P. Ji, W. Meng, B. Yang, and D. Wang, "Research on Prediction of the number of agricultural machinery data based on an improving method of BP neural network," *Journal of Chinese Agricultural Mechanization*, vol. 41, no. 2, pp. 200–205, Feb.2020.

[16] Y. Zhang, "Application of improved BP neural network based on e-commerce supply chain network data in the forecast of aquatic product export volume," *Cognitive Systems Research*, vol. 57, no. 10, pp. 228–235, Oct.2019, DOI: 10.1016/j.cogsys.2018.10.025.

[17] J. Wang, Z. Sheng, B. Zhou, and S. Zhou, "Lightning potential forecast over Nanjing with denoised sounding-derived indices based on SSA and CS-BP neural network," *Atmospheric Research*, vol. 137, pp. 245–256, Feb.2014, DOI:10.1016/j.atmosres.2013.10.014.

[18] L. Zha, and C. Xie, "Application of BP neural network in yam evenness prediction based on immune genetic algorithm," *Journal of Silk*, vol. 56, no. 2, pp. 19–26, Dec.2018.

[19] J. Cao, and J. Wang, "Exploration of stock index change prediction model based on the combination of principal component analysis and artificial neural network," *soft computing*, vol. 24, no. 11, pp. 7851–7860, Jun.2020.

- [20]X. Li, K. Hua, L. Yuan, and J. Shu, "The Research of BP-PSO Optimization Method in Electric Power Market Analysis and Forecast," *Advanced Materials Research*, vol. 860-863, pp. 2522-2525, Dec.2013, DOI:10.4028/www.scientific.net/AMR.860-863.2522.
- [21]M. Lu, "A monetary policy prediction model based on deep learning," *Neural Computing & Applications*, vol. 32, no. 10, pp. 5649-5668, May.2020.
- [22]M. Liu, G. Chen, X. Liu, and J Zhan, "Research on Initialization Algorithms of Weights and Biases of BP Neural Network," *Journal of Southwest China Normal University(Natural Science Edition)*, vol. 35, no.6, pp. 137-141, Dec.2010.
- [23] Elminir H K. Azzam Y A. and Younes F I. "Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models," *Energy*, vol. 32, no.8, pp. 1513-1523, Aug.2007.
- [24] Pasini A, "Artificial neural networks for small dataset analysis," *Journal of Thoracic Disease*, vol. 7, no.5, pp. 953-960, Apr.2015.
- [25] Kalogirou S A, "Applications of artificial neural-networks for energy systems," *Applied Energy*, vol. 67, no.1, pp. 17-35, Apr.2000.
- [26]Y. Chen, L. Chang, C. Huang, and H Chu, "Applying Genetic Algorithm and Neural Network to the Conjunctive Use of Surface and Subsurface Water," *Water Resources Management*, vol. 27, no.14, pp. 4731-4757, Oct.2013.
- [27]R. D. Hill, and J. M Conner, "Transient Heat Transfer Model of Machine Gun Barrels," *Materials and Manufacturing Processes*, vol. 27, no. 8, pp. 840-845, 2012, DOI:10.1080/10426914.2011.648694.
- [28]G. Li, Z. Li, and H. Xie, "Model and application of wavelet neural network based on IGA-BP," *Systems Engineering*, vol. 30, no. 10, pp. 112-117, Oct. 2012.
- [29]X. Li, L. Zhao, and Q. Meng , "Calculation of line losses in distribution systems using artificial neural network aided by immune genetic algorithm," *Proceedings of the CSU-EPSCA*, vol. 21, no. 5, pp. 87-91, Oct. 2009.
- [30]L. Wu, "A research of fault diagnosis algorithm based on genetic neural network," M.S. thesis, Dept. Computer Software and Theory, Liaoning Univ., Liaoning, China, 2012.
- [31]M. Cao, "Research on grey incidence analysis model and its application," M.S. thesis, Dept. College of Economic and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2007.
- [32]K. Yang, F. Meng, and Z. Liang, "Adaptively weighted PCA algorithm," *Computer Engineering and Applications*, vol. 48, no. 3, pp. 189-191, Jan. 2012.

ACKNOWLEDGMENTS

This work was supported in part by the Science and Technology Research Program Funded by the China National Textile and Apparel Council under Program Nos.2016091 and Program Nos.2019064, in part by the Science and Technology Research Program in Xi'an City under Program Nos.201805030YD8CG14-16, in part by the Xi'an Science and Technology Innovation Platform Construction Project/Key Laboratory Construction Project under Program Nos.2019220614SYS021CG043, and in part by the Youth Academic Backbone Support Program of Xi'an Polytechnic University.

AUTHORS CONTRIBUTES

Y.D.Chen initiated this work and performed the algorithm design, K.Y.Cao performed modeling and designed the experiment, S.Liu build experimental set up and performed experiments, X.X.Song wrote the paper.

Figures

Figure 1

Please see the Manuscript PDF file for the complete figure caption

Figure 2

Please see the Manuscript PDF file for the complete figure caption

Figure 3

Please see the Manuscript PDF file for the complete figure caption

Figure 4

Please see the Manuscript PDF file for the complete figure caption

Figure 5

Please see the Manuscript PDF file for the complete figure caption

Figure 6

Please see the Manuscript PDF file for the complete figure caption