

Application of GA Optimizing Neural Network to Coal Sales Forecasts

Shuang Zhang
Software College
Northeastern University
Shenyang, 110819, P.R.China
E-mail: zhangs@swc.neu.edu.cn

Qinghe Hu
Information Science and Engineering College
Northeastern University
Shenyang, 110819, P.R.China
E-mail: huqinghe@ise.neu.edu.cn

Abstract—With the rapid development of the market, coal enterprises predict sales by subjective experience, which is far from accurate. In order to minimize decision-making errors, to avoid warehouse inventory shortages or backlog and to increase prediction accuracy of coal sales forecast, the study of forecast methods is particularly important. In the paper, improved BP algorithm is adopted based on some large coal enterprises' practical characters. Connection weights are optimized by generic algorithm. The forecast method is implemented in the paper. Theoretical analysis and experimental results show that neural network is feasible and effective for coal sales prediction, with a bright future. Genetic algorithm optimizing neural network increases speed calculation and reliability.

Keywords—coal sales forecast; neural network; BP algorithm; LM algorithm; Genetic algorithm

I. BACKGROUND INTRODUCTION

With the rapid development of the market, enterprises are more serious about sales forecast. Coal enterprises predict sales by subjective experience, which is far from accurate. It is necessary to study deeply coal sales forecast to help coal enterprises to judge more accurately, scientifically and rapidly^[1,2].

Coal sales is influenced by many factors, such as market changes of downstream coal industry, change of season, development of coal industry, strategy changes of state policy and so on^[3]. Coal enterprises can not accurately grasp each factor's influence by subjective experience. And nonlinear coal sales change is difficult to set up an accurate mathematical model. So, it is necessary to select a suitable forecast method for coal sales^[4].

In recent years, neural network and generic algorithm is widely used to forecast^[5]. In the paper, genetic algorithm optimizing neural network is adopted to help coal enterprises to forecast coal sales accurately and scientifically to master market trends, to speed capital turnover, to optimize resources, and to avoid decision mistakes.

II. GA OPTIMIZING BP NEURAL NETWORK

Learning is one of the most important characteristics of neural network. So, during the process of neural network

development, algorithm learning is very important. GA is strong in global search; and neural network can learn by itself. Their combination can complement each other's advantages, increase network's generalization ability^[6,7]. There are 3 ways of GA optimizing neural network: connection weight of optimizing network, topology of learning network, learning rules of GA evolving neural network.

BP algorithm (BP) enables weight modification for gradient error descent, resulting in unavoidable tendency being entrapped in local minimum. LM method combining gradient descent algorithm and Newton method is adopted in the paper^[4]. Based on research experience of neural network identifying structural parameters, weights of BP are initialized based on generic algorithm (GA).

Network weights are initialized in a certain area. According to genetic algorithm of the survival of the fittest evolution theory, with minimum network errors as criteria, a group of weights are obtained after iterations. The weights have minimum network errors among the iterated ones by GA, close to minimum global energy under given structure. The group of weights is taken as the initial weights for BP network training^[10]. Combination of GA and BP algorithm achieves the complementary advantages, and solves the problem better.

The basic theory of GA optimizing BP (GA-BP) network weights is to find a weight matrix to get minimum sum of error square. Each chromosome in GA population is correspondent to weight matrix one to one. After iterations, the optimal chromosome is achieved and transferred into neural network weight. Then, with forecast factors as the network input, network output is the forecast result.

III. GA-BP NEURAL NETWORK MODEL DESIGN

A. BP Network Design

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this template and download the file for A4 paper format called "CPS_A4_format".

- Design network layers number and activation function

The author composes a multiple nonlinear complicated function based on BP algorithm, the number of independent variables equals to the number of network input nodes, the

number of dependent variables (functions) equals to the number of output nodes. A sample's learning time should not be too long, for the sake of making network less complicated, reducing requirement for computer and guaranteeing learning speed. There are three layer, input, single hidden and output layer. Hidden layer adopts hardness adjustable *sigmoid* type function; output layer adopts *purelin* type function.

- Design input and output nodes number

The number of BP network input nodes depends on dimensions of data source, with each node standing for a data source. Input nodes are not the more the better; more nodes will increase calculation exponentially. System size should be decreased as possible, on the basis of information guarantee. So, all the factors should be fully analyzed and strictly filtered^[9]. As we can see that factors influencing coal sales include market changes of downstream customers, season changes and coal mining. So, the author selects the factors as 9 input nodes: coal consumption ratio

of steel customers, chemical customers, electric power customer, life-canton and other customers, coal ore prices, coal mining, coal purchase price, coal purchase quantity.

In the model, it is coal sales to be forecasted, so output nodes number is 1.

- Design hidden nodes number

The number of nodes influences network performance greatly, so they must be selected properly. Normally the below formulas are used to determine the number of hidden nodes.

$$n = \sqrt{n_0 + n_1} + a$$

therein, n is the number of hidden layer neural cells, n_0 is the number of input neural cells, n_1 is the number of output neural cells, a is an integer between 1 and 10. according to the formula, the number of layers in the paper is between 5 and 20.

TABLE I. Network Error with Hidden Nodes Number Change

Number of hidden nodes	Network error
5	0.924912e-004
6	0.961323 e-004
7	0.889735e-004
8	0.751640 e-004
9	0.926443 e-004
10	0.852767 e-004
11	0.553252 e-004
12	0.461352 e-004
13	0.414382 e-004
14	0.374656 e-004
15	0.199721e-004
16	0.114840 e-004
17	0.265632 e-004
18	0.478594 e-004
19	0.562122 e-004
20	0.828352 e-004

First, scope of hidden nodes numbers is determined based on formula, one sample set is used to train the network one by one, to find the hidden nodes number when minimum network error. Training result is shown in Table1.

It can be seen from Table 1 that network training error is minimum when the number of nodes is 16. So the number of hidden nodes is set as 16.

- Design other parameters

Experiments show that forecast is most effective when error accuracy is 0.000001, training function is *trainlm*, learning ratio is $lr=0.5$, error of training expectation is $goal=0.00001$, maximum training times is $epochs=200$.

B. GA Design

- Coding Design

GA's advantages can show up only in effective coding, so coding is its basis. In the paper, coding objects are continuous weights and thresholds. In the paper, real coding is adopted for it's much shorter than binary coding and work saving of code conversion.

Among the three layers in the BP network, X and Y are network input and output respectively, R and $S2$ are numbers of nerve cells of input and output respectively. The beginning part of the coding is weight between input and hidden layer, middle part is weight between hidden and output layer, and the ending part is thresholds of hidden and output layer. Suppose there is $S1$ hidden nodes, so the length of the coding is $R \times S1 + S1 \times S2 + S1 + S2$.

For the case, coding length is $9 \times 16 + 16 \times 1 + 16 + 1 = 177$. Its contents are shown in Figure 1.

1...144	145...160	161... 176	177
Weight between input and hidden layer	Weight of hidden and output layer	Threshold of hidden layer	Threshold of output layer

Figure 1. Chromosome gene chain section

- Design of Population Size

Population size directly affects convergence and calculation efficiency. Too small size cannot provide enough samples; it brings immature convergence. Too big size can prevent too early convergence to local optimal solution. But, the bigger size, the more calculation for each generation; it may lead to unacceptable slow convergence. Normal size is 20~200 for experience [9]. Based on experiment training, initial population size is 100 in the paper.

- Design of Fitness Function

Because fitness is the only deterministic index of individual survival, it directly affects population evolution. In order to directly connect fitness with individual's quality measure, fitness value is rated non-negative, and the bigger the better in any case [10]. The smaller is the error between coal prediction sales and real ones, the better it is. In the BP network, performance index is the sum of error squares, the smaller the better performance. The objective function is

$$\min(f_i) = E(i) \quad (1)$$

$$E(i) = \sum_p (d - o)^2 \quad (2)$$

Therein, i is training times, p is number of data, d is forecast sales, o is real sales. So, the to-be-solved problem is a minimization of objective function. $E(i)$ is sum of error squares, non-negative, so individual fitness function $F(i)$ is set as the reciprocal of objective function., i.e.,

$$F(i) = \frac{1}{E(i)} \quad (3)$$

- Selecting operators

According to above analysis, operators are selected according to proportionate fitness calculation and roulette.

- Crossover operator and probability

Real coding is adopted in the paper, so crossover operators are adopted real number crossover. Crossover probability is determined according to trainings.

- Mutation operator and probability

Because real coding is adopted, mutation operators such as basic variation and uniform mutation are not suitable. Because the purpose of mutation is to change GA's local search capability, non-uniform mutation is introduced into GA for real coding mutation design. Mutation probability is

determined in the following experiments as crossover probability.

- Termination condition

It is easy to set the maximum iteration. But it is exactly determined based on real problem's complication and requirements. In real simulation, it is done by repeated experiments. According to optimum weight threshold characters, the maximum iteration is 200 based on training result.

C. Algorithm Steps

GA optimized neural network process is divided into training, testing and forecast process. Evolution is a cyclic process. When training samples reach objective accuracy or meet the maximum iteration, evolution terminates. The process includes generating initial population, calculating individual fitness, selection calculation, crossover calculation and crossover calculation.

Steps are described below.

Step1. Each weight and threshold of a random distribution is coded into a string by real coding. Based on the determined network structure and learning rules, the string corresponds to a neural network with a certain weight and threshold.

Step2. With input training samples, fitness is determined by calculate error function of neural network. The bigger the error is, the smaller the fitness is.

Step3. The individuals with maximum fitness are selected to the next generation.

Step4. The current generation is processed by crossover and mutation operators to generate next generation.

Step5. A new population is generated.

Step6. Neural network error is calculated.

Step7. Judge whether the training objective is achieved. If yes, testing step is taken; otherwise, step2~step6 are repeated to have the initial weight continuously evolved until training objective is achieved.

Step8. Testing samples are inputted to calculate neural network error.

Step9. Judge whether the training objective is achieved. If yes, forecast step is taken; otherwise, step1~step8 are repeated to have the initial weight continuously evolved until training objective is achieved.

Step10. Forecast samples are input, network forecast result is taken, and training is terminated.

GA-BP network takes advantages of GA's global search and BP's strong local search to optimize structure and parameters. It avoids the flaws of poor accuracy and easy sticking in local minima of single BP.

IV. APPLICATION OF COAL SALES FORECAST

A. Determine Factors to Coal Sales

It has been pointed out that coal sales are affected by many factors. The downstream coal enterprises are iron and steel companies, chemical enterprises, power supply departments and others. Their market changes and development will affect coal industry. National strategy for

downstream customers and international financial crisis indirectly affect coal sales. The most direct manifestation is the sales change to downstream industry. So, the author takes coal consumption sales proportion of steel customers, chemical customers and power customers, and other types of customers as input.

Season changes influence coal industry are mainly changes of heating and coal mining output. Winter is coal industry's peak period, because coal requirements for life surely increase; while in other seasons, it surely decreases. The coal for life is also an important factor.

Selling coal resource has two parts, one is enterprise-affiliated coal mine, and another is outside purchasing. So, mining, ore prices, purchasing amount and price will affect coal sales. If ore price or purchasing price is high, sales price

will be higher; more mining will cause to less purchasing. So, ore price, mining, purchasing quantity and price are all sales factors.

B. Sample Data

Based on the above analysis, the model uses data from coal enterprises in Liaoning Province from January 2004 to December 2010 as training and forecast samples. 72 sets of data are taken as training samples, 12 sets of data are taken as testing samples. Due to limited space, forecast samples sets are depicted in Table 2.

C. Training Experiment and Result Analysis

Samples are trained by GA optimizing traditional BP

TABLE II. PREDICTION DATA TABLE

Sr. No.	Coal consumption ratio of steel (%)	Coal consumption ratio of chemical industry (%)	Coal consumption ratio of power (%)	Coal consumption ratio of life (%)	Coal consumption ratio of others (%)	Ore price (yuan/ton)	Minining (ton)	Purchasing price (yuan/ton)	Purchasing amount (ton)	Sales amount (ton)
1	40.2	22.68	16.2	12.01	8.91	701	99049	692.5	876954	304491.5
2	41.8	22.67	16.55	11.68	7.3	719.5	77105	713	123957	225300.5
3	48.6	23.06	14.78	11.55	2.01	690	91405	684	116495	280248.5
4	58.8	23.09	12.98	3.08	1.95	732.5	54432	732.5	236950	169328
5	60.08	23.89	13.75	1.9	0.38	769	59038	780.5	107050	180742
6	59.2	23.65	14.88	1.3	0.97	734	59895	735	129654.5	183638
7	56.6	23.78	14.36	2.8	2.46	728.5	76338	720	103589	238163.5
8	55.6	23.77	13.62	2.5	4.51	732	57896	726.5	178680	174110
9	54.35	23.32	14.33	2.6	5.4	728.5	46530	760	132440	142424
10	55.2	22.24	14.66	3.2	4.7	735	90894	753	50693.5	282580.5
11	44.6	24.56	14.24	8.76	7.84	765.3	70687	745	214760	218068
12	41.2	23.31	14.08	11.99	9.42	742	62780	732.5	149850	191001

network (traditional GA-BP) and GA optimizing improved BP network (improved GA-BP) respectively. Both network training errors are shown in Figure 2.

SSE=4.078950e-005 is achieved. It can be seen that GA embedded improved BP network has better training results than GA embedded traditional BP network.

Forecast sets are tested by traditional and improved BP network respectively. The comparison is shown in Figure 3.

It can be seen that GA optimizing improved BP network has better forecast results, but its accuracy is not as high as GA optimizing traditional BP network.

D. Network Model Comparison

From above cases it can be seen that GA optimizing neural network has better forecast effect. In order to further demonstrate its effect and accuracy, forecast results of general BP network, improve BP network and GA-BP network are compared more accurately by mean absolute error (MAE), square error (SSE), mean square error (MSE) respectively.

Comparison result is presented in Table 3. It proves that real coding of GA optimizing improved BP network has better forecast effect and accuracy than traditional and improved BP, avoiding BP algorithm's inherent deficiency of easy plunging into local minimums and slow iterative convergence.

TABLE III. GA-BP AND BP ALGORITHM RESULTS CROSS-REFERENCES

Testing index	MAE	SSE	MSE
General BP	0.0099	0.5134	0.0071
Improved BP	1.1682e-004	7.0750e-005	9.8264e-007
GA-BP	0.9079e-004	4.2731e-005	5.9349e-007

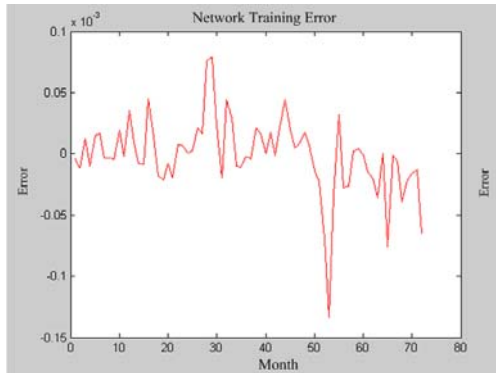
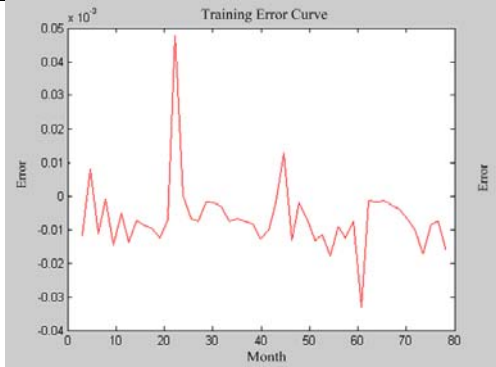


Figure 2. Network training error curve

V. CONCLUSION

The paper introduces GA optimizing improved BP network, which is used to coal sales forecast. Its comparison to general BP and improved BP shows that GA optimizing neural network increases accuracy and convergence for nonlinear time sequence forecast like coal sales. It is capable of learning, mapping and generalizing. It possesses very important value in theory and practice.

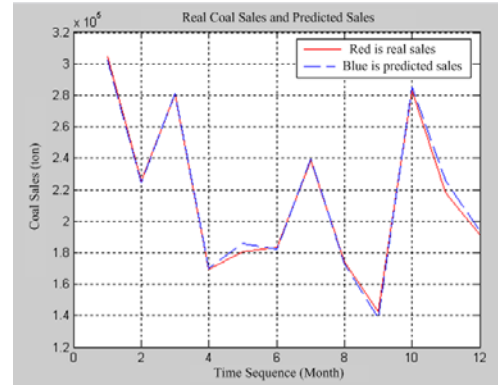
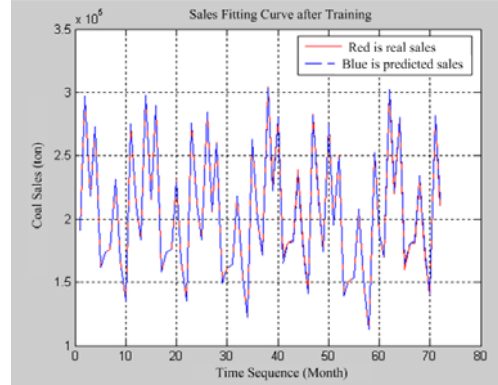


Figure 3. Test sample of fitting curve

ACKNOWLEDGMENT

This work is supported by the State Key Program (Grant No. 70931001) and the Innovative Community Project (Grant No.71021061) of National Natural Science of China.

REFERENCES

- [1] Wu C, ShaSha Y. Coal Demand Forecast Model Analysis, Open mining technology, (4), 2009, pp.91-91.
- [2] Xiyu C. Application of Market Forecast to Coal Sales, China Coal, 33(2), 2007, pp.17-19.
- [3] Li T. Preliminary Study of Coal Demand, Coal Engineering, 5(12), 1991, pp.30-33.
- [4] Yan W, Jingwen L. China's Present Situation of Coal Consumption and Future Coal Demand Forecast, China

- Population Resources and Environment, 18(3), 2008, pp. 2-156.
- [5] Dechun A. Application of Neural Network for Domestic Coal Demand Forecast, Journal of Liupanshui Modeling College, 21(2), 2009, pp. 34-36.
 - [6] A. S Chen, LeungM. T, Daouk. H. Application of Neural Network to an Emerging Financial Market Forecasting and Trading the Taiwan Stock Index, Computers and Operations Research, 30(6), 2003, pp: 901-924.
 - [7] Xin W, Lu Z. Simulation and Application of Matlab Neural Network, Beijing Science and Technology Press, 2003, pp: 258-284.
 - [8] ZhongZhi S. Neural Network, Beijing, High Education Press, 2008, pp.56-63.
 - [9] Lu Y. Theory and Application of Stock Forecast based GA-BP Neural Network, Electric Power University (Beijing), 2010.
 - [10] Jun Y, Shouyu C, Ju Q. Forecast Model and Application of Neural Network based on GA and BP, Journal of Dalian University of Technology, 42(5), 2002, pp. 594-59.