Study on Gold Price Forecasting Technique Based on Neural Network Optimized by GA with Projection Pursuit Algorithm

Guixia Yuan
Jiangsu Radio and Television University, Nanjing, Jiangsu 210036, China

Abstract
Gold price has significant nonlinearity and time-variance with many indeterminate influencing factors. In order to improve the forecast accuracy of gold price, this paper puts forward a gold price forecast model combing projection pursuit with neural network. At first, projection pursuit algorithm is used to screen the influencing factors, and then the influencing factors are used as the input variables of BP neural network to learn. Meanwhile, this paper applies genetic algorithm to optimize BP neural network and build gold price forecast model. At last, the forecast performance of the model is tested through simulation experiment. Experimental results show that the combined model can well describe the variation trend of gold price, simplify the network structure and speed up network convergence, which effectively improves the forecast accuracy and operating efficiency of gold price and provides a new forecast method for gold price.

Keywords: Genetic Algorithm, Gold Price, Neural Network, Projection Pursuit, Forecast Model

1. Introduction

Because that gold price is influenced by various factors with dramatic fluctuations, great randomness, mutability and nondeterminacy, gold price forecast has long been the hot spot in focus of the authors and investors all around the world [1].

The traditional forecast methods of gold price include: autoregressive conditional heteroskedasticity model, time series analysis method and grey forecasting model. These methods all assume that gold price is determinate and linearly-changed, while the actual gold futures market is a complex nonlinear dynamical system which is difficult to build accurate mathematic model. Therefore, these traditional forecasts can not grasp the nonlinear phenomena in gold market, which is difficult to reflect the changing rules of gold price, resulting in great forecast errors [2, 3]. Along with the development of nonlinear dynamical system, some nonlinear forecast methods of gold price emerge, especially the neural network method which does not need to build the mechanism model for the target and can reasonably describe its features with the characteristics of parallel processing, strong learning ability and high robustness. It has been widely used in the modeling of complex systems and become the main research method for gold price forecast [4-6]. However, in the actual forecast of gold price, many factors influence gold price, such as rate of inflation, effective exchange rate of dollar, oil spot price and the world gold reserve. If all the data of influencing factors of gold price are directly input into the neural network to learn, calculated amount and input dimension will increase exponentially and network structure will become very complex, resulting in that network convergences become slower and problems such as “curse of dimensionality” are easy to appear. Therefore, the reliability and accuracy of gold price forecasting result is influenced [7].

Projection pursuit (PP) is a new nonlinear and high-dimensional method of data analysis, which projects high-dimensional data to low-dimensional subspace and seek the projections with high-dimensional data features in low-dimensional subspace, decreasing the dimension of feature space and preserving necessary information at the same time [8-10]. In order to improve the forecast accuracy of gold price, this paper puts forward a gold price forecast method on the basis of combining projection pursuit with BP neural network optimized by genetic algorithm (PP-BPNN) and conducts empirical analysis through simulation experiment to test the effectiveness and superiority of the forecast method.

2. BP neural network algorithm

BP network is a kind of feedforward neural network based on gradient decent algorithm, which is
mainly constituted by input layer, hidden layer and output layer. The neuron nodes in the same layer
do not have interactions, while the neuron nodes in different layers interconnect.

The basic operating process of BP neural network is as follows: at first, conduct forward calculation
through network and calculate the error between actual output and expected output; then dynamically
regulate the link weight and threshold between the nodes of the neural network according to the actual
error values to generally decrease the error value between actual output and expected output; finally,
make network actual output and ideal output lower than the given error value through repeatedly
conducting the process.

Set that the output of the j-numbered neuron in the first layer is \( O_p^{(l+1)} \), then \( W_{ij}^{(l)} \) refers to the
neuron weight of the two layers and the input and output relations of each neuron are as follows:

\[
\begin{align*}
O_p^{(l+1)} & = f_s[I_p^{(l+1)}] \\
I_p^{(l+1)} & = \sum_{j=1}^{N_i} W_{ij}^{(l)} \cdot O_j^{(l)} - \theta_i^{(l+1)}
\end{align*}
\]

(1)

Where, \( f_s[\cdot] \) is the function of network nodes.

\[
f_s[I] = \frac{1}{1 + \exp(-I/I_0)}
\]

(2)

Set \( -\theta_i^{(l+1)} \cdot W_{i,N_i+1}^{(l)} \cdot O_p^{(l)} = 1 \), then

\[
\begin{align*}
O_p^{(l+1)} & = f_s[I_p^{(l+1)}] \\
I_p^{(l+1)} & = \sum_{j=1}^{N_i} W_{ij}^{(l)} \cdot O_j^{(l)}
\end{align*}
\]

(3)

\( E_p \) is the quadratic sum of the error between the expected output and actual output, then

\[
E_p = \frac{1}{2} \sum_{j=1}^{N_i} (d_{pj} - y_{pj})^2
\]

(4)

Minimize \( E_p \) through changing the weight coefficient \( W_{ij}^{(l)} \) and make the actual value
approximate the expected output. Apply steepest decent algorithm to make the weight value shift
towards the direction of the negative gradient of the error function, and the adjustment formula is
shown in Equation (5):

\[
\Delta_p W_{ij}^{(l)} = -a \frac{\partial E_p}{\partial W_{ij}^{(l)}}
\]

(5)

Where, \( a \) is the study step.
It can be gained:
\[
\frac{\partial O_{pt}^{(i+1)}}{\partial H_{ij}^{(i+1)}} = \frac{O_{pt}^{(i+1)} (1 - O_{pt}^{(i+1)})}{I_q}
\]  

(7)

The optimization of BP network mainly focuses on the optimization of network algorithm and BP algorithm has two important problems in actual application: the convergence speed is slow and the target function has local minimum. Therefore, this paper does some significant attempts to optimize it and the following are parts of the improvement measures: that is, apply genetic algorithm to optimize neural network.

At first, the gradient of the transfer function can be changed. Add constant factor \( \beta \) to the input value of the node function of the following function to change the nonlinear feature of the function:

\[
o[j] = f(\text{net}[j]) = \frac{1}{1 + e^{-\beta \text{net}[j]}}
\]  

(8)

\( \beta \) can be used to adjust the gradient of the curve since the smaller \( \beta \) is, the bigger the curve’s gradient is. Increasing the gradient of the activity function appropriately can change the gradient of the node function and so that change the convergence of the algorithm.

Secondly, introduce exponential energy function. For the situation with large quantity calculating data, the convergence speed of ANN is slow and ANN is easy to oscillate. In order to speed up the convergence of forward neural network and avoid the oscillation in convergence process, the exponential energy function in the following form can be used:

\[
J = A^{BE} (A > 0 \text{ and } A \neq 1, \ B > 0)
\]  

(9)

Where, \( J \) is the energy function; \( E \) is the quadratic sum of the output errors of the neural network; \( A \) and \( B \) are the parameters of the exponential energy function. Applying energy function to update the weight can speed up learning. The network convergence process can be improved as long as appropriate parameters \( A \) and \( B \) are selected.

3. Determination of BP Neural Network Structure

3.1 Normalization processing of influencing factor

Set the influencing factor of gold price sample set as \( \{x(i,j), i = 1, 2, \ldots, n; j = 1, 2, \ldots, p\} \), where \( x(i, j) \) is the \( j \)-numbered influencing factor of \( i \)-numbered sample; \( n \) is the number of samples and \( p \) is the number of factors. In order to eliminate the effects of the difference in factor dimensions on the training speed of BP neural network, the following equation is used to conduct normalization processing:

\[
x(i, j)' = \frac{x(i, j) - x_{\text{min}}(j)}{x_{\text{max}}(j) - x_{\text{min}}(j)}
\]  

(10)

Where, \( x(i, j)' \) is the influencing factor value after the normalization, \( x_{\text{min}}(j) \) and \( x_{\text{max}}(j) \) are the maximum value and minimum value of the \( j \)-numbered influencing factor.

3.2 Determine the number of nodes of the network input layer by projection pursuit
(1) Construct projection indicator function $Q(\alpha)$

Projection pursuit algorithm integrates $p$-dimensional data \( \{(i,j), j = 1, 2, \ldots, p\} \) to one-dimensional projection value $Z(i)$ whose projection direction is $\alpha = \{\alpha(1), \alpha(2), \ldots, \alpha(p)\}$, then:

$$Z(i) = \sum_{j=1}^{p} \alpha(j)x(i, j), i = 1, 2, \ldots, n$$

Projection indicator function can be expressed as:

$$Q(\alpha) = S_z D_z$$

Where, $D_z$ is the local density of projection value $Z(i)$; $S_z$ is the standard deviation of projection value $Z(i)$, then:

$$S_z = \sqrt{\frac{\sum_{i=1}^{n} (Z(i) - E(z))^2}{n - 1}}$$

$$D_z = \sum_{i=1}^{n} \sum_{j=1}^{n} (R - r(i, j)) \ast u(R - r(i, j))$$

Where, $R$ is the window radius of local density; $E(z)$ is the mean of projection sequence $Z(i)$; $r(i, j)$ is the distance between the samples, then:

$$r(i, j) = |Z(i) - Z(j)| = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}$$

(2) Optimize projection indicator function

Different projection directions show different characters of data structure and the best projection direction is the projection direction that reflects a kind of structure characters of high-dimensional data as much as possible. Therefore, the best projection direction can be estimated by solving the maximum problem of projection indicator function, that is:

$$\max Q(\alpha) = S_z D_z$$

The constraint condition is:

$$\sum_{j=1}^{p} \alpha^2(j) = 1$$
(3) Determine the best projection direction
Solve Equation 14 with genetic algorithm and the best projection direction is gained when its target function reaches the extreme value.

3.3 Determination of the number of hidden layers and nodes in output layer of BP neural network

A number of researches have shown that for the hidden layer of BP neural network, the neural network with only one hidden layer can approximate a nonlinear function in any precision as long as there are enough hidden layers, so the number of BP neural network hidden layer in this paper is 1. Meanwhile, because that gold price is a single-output forecast problem, the number of nodes in the output layer of BP neural network is 1.

3.4 Procedure of gold price forecast

Gold price forecast includes two stages of learning and forecast:
(1) Learning stage. Optimize BP neural network parameters and construct corresponding gold price forecast model.
(2) Forecast stage. Forecast the future gold price with the constructed forecast model.

The gold price forecast procedure on the basis of combining projection pursuit and neural network is shown in Figure 2.

4. Simulation Experiment

4.1 Data source

In order to test the performance of the gold price forecast model combining projection pursuit with neural network, the price data of Au999.5 exchanged in Shanghai Gold Exchange is selected to conduct simulation experiment. The time range of the sample is from July 15th in 2010 to October 14th in 2010 (the data come from RESSET http://www.resset.cn) with 100 sample data. The front 80 data are selected as training sample set to build the gold price forecast model and the rear 20 data constitute the training sample set to test the feasibility and effectiveness of the built forecast model, the data is shown in Figure 3.
4.2 Determination of influencing factors of gold price

(1) Preliminary selection of the influencing factors. In order to explain and correctly predict the changing rules of gold price, it is necessary to select as many influencing factors of gold price fluctuation as possible at first. The preliminary influencing factors selected by this paper are: dollar index, oil price, gold spot price, quantity of gold supplied by the world, quantity of gold demanded by the world, Dow Jones index, US treasury bonds yield curve in ten years and exchange rate of dollar for RMB.

(2) Further select the projection pursuit algorithm of the factors. Projection pursuit algorithm is applied to gain the best projection direction:

\[ \alpha^* = [0.145, 0.114, 0.346, 0.355, 0.447, 0.0569, 0.015, 0.075] \]

(3) Rank the influencing factors according to the best projection direction and select the five factors of quantity of gold supplied by the world, quantity of gold demanded by the world, gold spot price, and dollar index and oil price.

(4) Select the factor which is most influencing to gold price from the above five factors as the input node of BP network, that is, gain five optimal input nodes with projection pursuit.

4.3 Simulation result of BP neural network

The convergence process of BP neural network to the training sample set is shown in Figure 4. It can be known from Figure 4 that the studying accuracy of BP neural network reaches the requirement of expected accuracy after 80 generations of iteration. Then, this model is applied to fit the gold price training sample and the gained actual value and fitted value of gold price are shown in Figure 5.
It can be seen from Figure 5 that the gained fitted value and actual value of gold price forecast model are very approximate with little error and the result shows that the suggested combined gold price forecast model is feasible and effective, which can forecast the future gold price.

4.4 Comparison with results of other models

The forecasting ability is what should be mainly examined to evaluate a forecast model, but not its back substitution fitting result. Meanwhile, in order to make the forecasting result of the suggested model more persuasive, PP-MLR, BPNN and the suggested model (PP-BPNN) are applied respectively to conduct independent forecast, where PP-MLR refers to apply projection pursuit algorithm to conduct factor screening at first and then apply multiple linear regression to conduct modeling forecast; BPNN refers to apply BPNN to conduct modeling forecast directly without factor screening. The forecasting results on gold price of these models are shown in Figure 6.

![Figure 6. Forecasting Results of Gold Price](image)

It can be known from Figure 6 that applying PP-BPNN to build gold price forecast model can accurately forecast gold price, which well describes the changing trend of gold price, controlling the forecasting error within 8% and meeting the demand of gold price actual forecast. The simulation result shows that the gold price forecast model combining projection pursuit with BP neural network is a gold price forecast model with high forecasting accuracy.

Conduct statistics on the root mean square errors (RMSE) of the forecasting results of three forecast models and the gained result is shown in Table 1.

<table>
<thead>
<tr>
<th>Forecast model</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNN</td>
<td>12.33</td>
</tr>
<tr>
<td>PP-MLR</td>
<td>27.45</td>
</tr>
<tr>
<td>PP-BPNN</td>
<td>12.27</td>
</tr>
</tbody>
</table>

It can be seen from Table 1 that the RMSE of PP-BPNN is much lower than that of PP-MLR, which shows that for the gold price presenting high nonlinear changing rules, it is difficult to build accurate mathematical model with linear forecast model MLR and the forecasting error is bigger; while comparing PP-BPNN with BPNN, the forecasting accuracy does not decrease, but the studying speed obviously accelerates, which shows that it can greatly increase the efficiency of solving problems by neural network and be beneficial to increase the forecasting accuracy and efficiency of gold price to apply PP to process and screen the influencing factors of gold price, which reduces the dimension of gold price, extracting the most useful information and eliminating the influence of relevant or repeated information among the influencing factors.
5. Conclusion

Aiming at the characters of gold price such as randomness, mutability and nondeterminacy, the gold price forecast model combining projection pursuit with BP neural network is put forward. At first, projection pursuit algorithm is applied to screen the nonlinear factors which have great influence on gold price changes and decrease the number of nodes in the input layer of BP neural network. Furthermore, genetic algorithm is used to optimize BP neural network, speeding up the training and building the simulation model of gold price forecast. The simulation result shows that the combined model can fit the changing trend of gold price change and effectively improves the forecasting accuracy and efficiency of gold price, which provides a new effective method for gold price forecast, so the suggested model has wide application prospect in gold price forecast.

6. References