Computational intelligence model based on GA-BP neural network

Chengzhao Zhang[*]

Chengdu Polytechnic, Chengdu, PR China

Abstract. Since the birth of the secondary stock market, the prediction of the stock price trend has become a research direction concerned by many people. Aiming at the problem of non-stationary and non-linear stock price forecasting, this paper builds a computational intelligence model to improve the neural network with genetic algorithm. The results show that, compared with other models, the GA-BP neural network model proposed in this article can effectively improve the prediction of the rise and fall of the HS300 index, and the withdrawal range is small when the market falls. The research of this paper enriches the method of financial time series data analysis, which can not only provide decision-making reference for investors, but also help to enhance the cognition of financial market rules. The model can significantly reduce the forecast error and improve the model fitting ability.

Keywords: Stock market, Genetic algorithm, BP neural network, Prediction error.

1 Introduction

Stock market forecasting is a very demanding problem. Many factors influence stock market prices, for example, company news and results, sector performance, consumer sentiment, social media sentiment and financial factors. The characteristics used for prediction are diverse: price data, technical indicators of historical data, other market price fluctuations related to individual stocks, currency exchange rate, oil prices, etc. Automatic feature extraction and accurate market prediction have attracted more and more attention. The existing financial market analysis methods are generally divided into two categories. One is fundamental analysis and the other is technical analysis.

In the technical analysis, the historical data of the target market and some other technical indicators are important factors in the prediction. According to the efficient market hypothesis, stock prices can reflect all their information. Technical analysis predicts the future trend of market price by analyzing the previous price data and technical indicators. The fundamental analysis of the intrinsic value of securities investment can be achieved by obtaining the detailed information of the financial statements of related assets,

---
[*]Corresponding author: 2595643823@qq.com

such as cash flow, growth rate, discount factor, annual discount and capitalization rate, so as to deeply understand the future development of the company. Fundamental analysis is often used to decide which stock to choose for trading.

Deep learning is a kind of method suitable for automatic feature extraction and prediction. In fields like machine vision, natural language processing, and speech signal processing, deep learning methods can gradually construct useful and complex features from raw data or simple features. Since the actions of stock markets are complex and nonlinear, and the stock data has noise, it is facing a great challenge to extract accurate features to predict.

The stock market is one of the most important parts of the investment market. Compared with other industries, the stock market has a higher rate of return on investment, but there are also higher risks. Stock market forecasting has always been a problem that investors pay close attention to. Therefore, the research on the stock market prediction methods and how to reduce the error of stock price prediction has become a research hotspot of many scholars at home and abroad. At present, the common stock forecasting methods include econometric method and machine learning method.

Econometric method is based on statistical theory, the application of statistical analysis model to predict the stock price. The common models are ARMA, ARIMA, and GARCH. Although the econometric model is objective and has good theoretical support, its advantages lie in the processing of linear data and the need for strict basic assumptions. The stock price series is a complex nonlinear time series, so the traditional econometric model can not achieve the best prediction effect.

Compared with the econometric model, machine learning method can directly mine valuable information from the data without making assumptions in advance, and machine learning method can better deal with nonlinear data, so it is widely used in stock price forecasting. Common machine learning methods include decision tree, SVM, SVR, neural network and so on. Among them, the learning mechanism of neural network makes it have more powerful learning ability and nonlinear fitting ability, and has more advantages in the field of stock price prediction.

The latest literature related to GA-BP neural network is as follows. The GA-BP neural network model proposed by Wang[1] et al. provide a new practical method for knowledge fusion risk assessment of Creative ecosystem. Li[2] et al. construct an evolutionary neural network model by combines GA with BP neural network to   solve the global optimization problem. Song[3] et al. point out that GA-NN model has a good application prospect in ionosphere research. A BP neural network prediction model optimized on the basis of genetic algorithm is proposed by Cao[4] et al. and Wang[5] et al. The experimental results achieved by Zou[6] et al. show that the GA-BP algorithm is accurate and effective for online identification. The experimental results of Peng[7] et al. show that GA-BP model is superior to other competitors. The BP neural network based on PEC and GA proposed by Wang[8] et al. has been proved to detect cracks in ferromagnetic steel quantitatively. An optimized GA-BP neural network model is established by Zheng[9] et al. to predict the long-term skid resistance. A wind speed prediction method based on EEMD and GA-BP neural network is proposed by Wang[10] et al..

Other chapters are summarized as shown below. Section 2 introduces the fundamental knowledge of genetic algorithms. Moreover, Section 3 presents the introduction of BP neural network and Section 4 introduces BP neural network model optimized by GA. Furthermore, Section 5 presents the empirical result analysis and comparative analysis. Finally, Section 6 gives the conclusions.

2 Basic knowledge of GA

Genetic algorithm is a stochastic worldwide search and optimization method that simulates biological world based on the natural genetic mechanism of natural selection and species evolution. Optimization of BP neural network using genetic algorithm can obtain better initial weights and thresholds, avoid network training into partial minima and improve convergence speed.

Genetic algorithms were first proposed by John Hollander in the USA in the 1970s. The algorithm was designed and proposed based on the evolutionary laws of organisms in nature. Genetic algorithm is a kind of random search algorithm. It is a process based on Darwin's theory of biological evolution and genetic mechanism to obtain an optimal solution by simulating the process of natural selection and biological evolution. It is widely used in areas like function optimization and picture processing production scheduling, machine learning and combinatorial optimization. Different from the traditional search algorithm, genetic algorithm is an optimization method based on probability search. It does not need to determine the search rules in advance, and can adjust the search direction adaptively to achieve automatic search space and optimization guidance.

The genetic algorithm used in this paper starts from the randomly generated partial solution set of the problem to be solved, which contains a certain number of individuals. These individuals generally use a string of data or array to represent the coding of the solution, which is called chromosome. Then the quality of these chromosomes is measured using the fitness function. On this basis, crossover and mutation operations are performed on the chromosomes in the initial population using genetic operators to form a new population. According to the evolutionary principle of survival, this evolutionary process will make the fitness value of chromosomes in the population gradually increase after several generations, that is, the algorithm will converge to the best chromosome, which is the optimal solution we seek. Therefore, the main contents of genetic algorithm include: initial coding design, determination of fitness function, selection, crossover, mutation, setting termination conditions.

2.1 Selection

Selection refers to the selection of several chromosomes from the population according to a certain probability, which is a process of converging the population to excellent individuals based on fitness value. In this paper, the strategy of combining fitness proportion method and best individual retention method is used to select.

The main idea of fitness proportion method is that the probability of selecting an individual from a group is related to its fitness evaluation. The higher fitness of the individual's corresponding solution, the larger value of objective function and the greater probability of selection. Suppose that the initial population size is H, that is, the chromosome number is 1, 2,..., h. The fitness value of individual r is f_r, the probability of individual r being selected is set to p_r, and the cumulative probability is set to q_r, then:

$$p_r = \frac{f_r}{\sum_{r=1}^{h} f_r}$$

(1)

$$q_r = \sum_{s=1}^{r} p_s$$

(2)

After selecting chromosomes from male parents for cross mutation, the offspring population will be formed. According to the fitness, 50% of the individuals with the highest fitness will be selected from the parents and offspring to form a new population, which can not only make the excellent individuals participate in the cross mutation operation, but also keep the excellent individuals in the male parents and avoid missing the better solution. At the same time, the convergence speed is accelerated.

2.2 Crossover

Crossover means that two individuals selected from the parents exchange partial genes to form new offspring, which is used to guarantee the diversity of individuals in the population. This paper uses a two point crossover method to select sires from the mating pool, and then two gene positions are randomly assigned in each pair to cross, so as to produce a new pair of individuals until the end of the crossover, forming a new population. For example, individuals y and z are selected for crossover, and the chromosome codes before and after crossover are as follows.

Before crossover:

$$R_y = \left\{ y_1, y_2, y_3, \cdots, y_n \right\} \tag{3}$$

$$R_z = \left\{ z_1, z_2, z_3, \cdots, z_n \right\} \tag{4}$$

After crossover:

$$R_y = \left\{ y_1, y_2, y_3, \cdots, z_i, \cdots, z_j, \cdots, y_n \right\} \tag{5}$$

$$R_z = \left\{ z_1, z_2, z_3, \cdots, y_i, \cdots, y_j, \cdots, z_n \right\} \tag{6}$$

2.3 Mutation and setting termination conditions

Mutation is to select some genes in some chromosomes to make changes, which enhances the diversity of the population and effectively prevents the algorithm from entering the convergence state early.

After selection, crossover and mutation, we have a new population. In the case of successive choices, crossing and variation, if the termination condition is not set, the genetic algorithm will continue to calculate, so we need to set a reliable termination condition.

The termination conditions set in this paper are as follows: When the iteration count hits its maximum value, the algorithm stops running. When the algorithm matures and converges, it stops running when there is only one chromosome type left in the population.

3 BP neural network

Artificial neural network model has strong nonlinear mapping power, and can process plenty of data at one time. The idea of neural network model is: the information input, the parameters are adjusted according to the set error expectation, and the output conforms to the expected error result.

Neural network data is a classical machine learning method when dealing with large data. The model has strong nonlinear mapping capability, self-adaptive capability and self-

learning capability. It has good generalization ability and fault tolerance when dealing with multi-source heterogeneous data prediction because of its wide applicability and good fitting effect, multilayer feedforward neural network realizes the feedforward learning and training of model threshold through the feedback learning mechanism of hidden layer, which has been widely used in classification, prediction and other fields.

BP neural network is a kind of feedforward neural network model. Its core idea is to analyze the error according to the comparison between the output of each training and the actual result, and to correct the weight and bias of BP neural network depending on the error. Finally, the output of the model is consistent with the expected result step by step. Generally, A BP neural network is composed of an input layer, a hidden layer and an output layer.

BP neural network is a kind of multi-layer perceptron, which is used to solve the linear indivisible problem in prediction. Neural network is a kind of model similar to black box, which includes several hidden layers besides input layer and output layer. A neural network model, on the other hand, has three or more layers of neurons. Among them, the neurons between the upper and lower layers of the adjacent BP neural network in the fully connected layer achieve full connectivity, while the neurons in the same layer are not connected to each other. In this case, the connection between the input layer and the hidden layer is based on the weights of the network, i.e., the strength of the connection between two neurons. Input layer transfers information contained in data into hidden layer, while hidden layer integrates all information from all neurons in the previous layer and continues to pass down until it is passed to output layer.

Although adding the number of network layers can reduce the error, it also complicates the neural network and increases the network weights' learning time. Actually, improving the error accuracy can be achieved by increasing the number of neurons in the hidden layer, and the training effect is better than increasing the number of network layers. Meanwhile, in accordance with Kosmogorov's theorem, the three-layer neural network can approximate any continuous function under the condition that the neural network structure is reasonable and the node weights are appropriate. Therefore, based on the above analysis, the number of hidden layers is set to 1, that is to build a three-layer neural network.

4 Genetic algorithm for optimizing BP neural network models

BP neural network is a multi-layer forward neural network trained using the error back propagation algorithm. BP neural network can achieve arbitrary non-linear mapping of input and output, self-learning and simple structure, but the speed of training of BP neural network is slow and easy to fall into local minima. Genetic algorithm is a biological natural genetic mechanism based on imitation genetic algorithm. It is a randomised global search optimization method based on natural biological genetic mechanisms, mimicking natural selection and species evolution. Genetic algorithms are a stochastic global search and optimization method based on biological natural genetic mechanisms that emulate natural selection and species evolution. Optimizing BP neural networks with genetic algorithms can obtain better initial weights and thresholds, avoid network construction falling into local minima and improve convergence speed.

The neural network model divides the samples as training samples and testing samples. The training samples are used to train the optimal simulation network model, and the test samples are employed to test the prediction results of the model. In this paper, a 3-layer neural network structure is constructed, in which the input layer, hidden layer and output layer are all one layer, and the structure is shown in Figure 1.

The weights and thresholds of the BP neural network are used as the initial populations of the genetic algorithm. The next generation population is obtained by genetic action, and

the individual fitness value is obtained. The weight and threshold are adjusted and optimized by selection, crossover and mutation until the individual learning error of fitness value is less than the specified value.
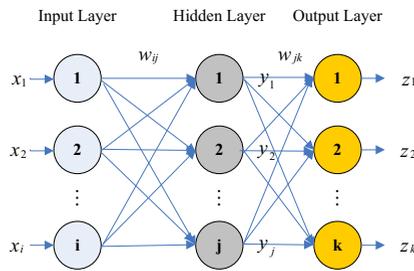
Fig. 1. Architecture of BP Neural Networks.

Finally, training the neural network and calculating the mean square error after each iteration. The iteration is stopped when the error between the outcome and the desired output reaches a predetermined level of error convergence. The sequence of optimal network weights is obtained. The trained simulation network is used to simulate and predict the test samples corresponding to the trained simulation network, and the prediction results are obtained. The specific implementation process of the GA optimized BP neural network is shown in Figure 2.
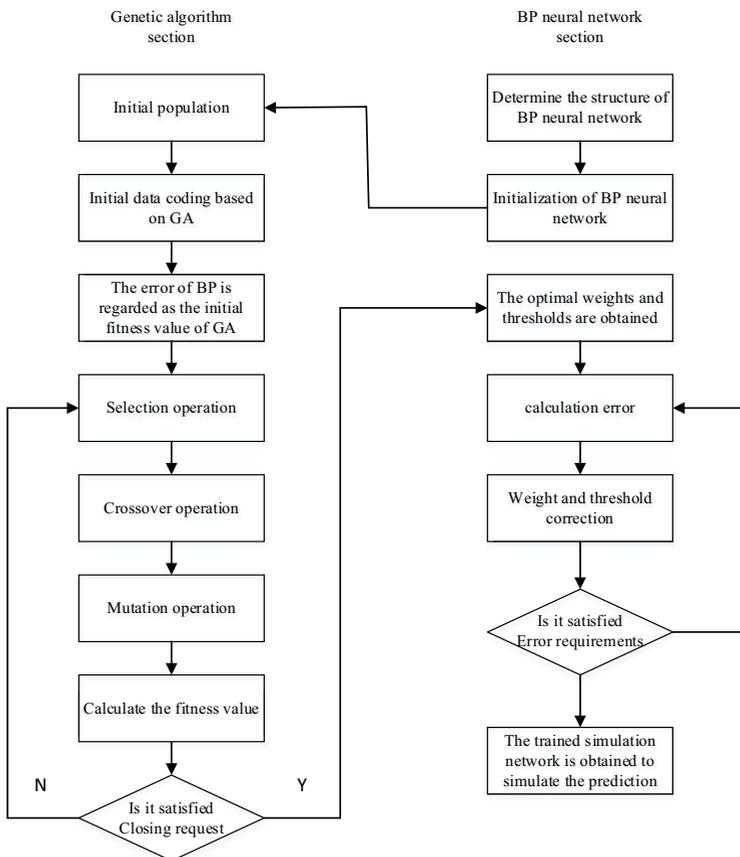
Fig. 2. Flow chart of BP neural network optimized by GA.

## 5 Empirical result analysis and comparative analysis

This paper uses the 975 data points of HS300 close price index from June 9, 2017 to June 9, 2021 as the sample set. 725 data points from June 9, 2017 to May 29, 2020 were used as training sets, and 250 data points from June 1, 2020 to June 9, 2021 were used as test sets to implement the test. GA-BP neural network, BP neural network and GARCH model are used for prediction, and root mean square error (MSE) and mean relative error (MAPE) are used for comparative analysis.

where y_n is the real value, $\overline{y}_n$ is the simulation predicted value, and $n = 1, 2, \cdots, N$ is the number of test samples.

$$MAPE = \sum_{n=1}^{N} \frac{\left| \dfrac{y_n - \overline{y}_n}{y_n} \right|}{N}$$

(7)

$$MSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (y_n - \overline{y}_n)^2}$$

(8)

The flow chart of HS3000 close price index sequence is shown in Figure 3.
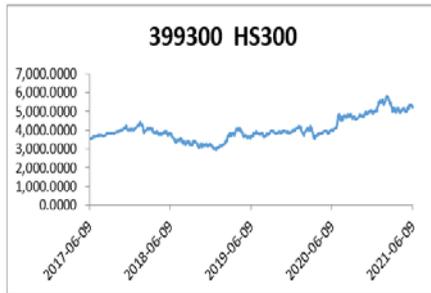
Fig. 3. Process flow diagram HS3000 close price index sequence.

The simulation errors of the three models are calculated by formula (7) and formula (8), and the results are shown in Table 1. The MAPE of GA-BP neural network model is 23.48, the MAPE of BP neural network is 69.78 and the MAPE of GARCH model is 75.84. The empirical findings show that the simulation prediction effect of GA-BP neural network is better and can reveal the trend of HS300 index better. Comparing the results of the three models, the mean square error and MAPE error of the GA-BP neural network model are smaller than those of the BP neural network model and the GARCH model, indicating that the simulation accuracy of the GA-BP neural network model is higher.

Table 1. Simulation performance comparison of three models.

Model	MSE	MAPE
GA-BP	3806.54	23.48
BP	23854.36	69.78
GARCH	24567.23	75.84

GA-BP neural network model has stronger simulation ability. The output value of the network is nearer to the actual value. The prediction error, convergence speed and accuracy are better than BP neural network and GARCH model, which can be used as the basis for the determination of the trend of HS300 index.

6 Conclusion

This paper uses three models including GA-BP neural network model to forecast the HS300 index and compare the results based on 975 data points from 2017 to 2021. The empirical results show the following conclusion.

On the one hand, GA-BP neural network inherits the self-learning and nonlinear mapping ability of BP neural network, and on the other hand, it solves the problems of slow convergence and easy to fall into local optimum of BP neural network. It has the benefits of stable output, fast speed of convergence and high prediction accuracy. The GA-BP neural network outperforms the BP neural network in terms of simulation capability, error level, convergence accuracy and number of iterations. It has high prediction accuracy, strong simulation ability and good data fitting ability.

We developed a GA-BP neural network tool to test the HS300 index and use the daily data for analysis in this paper. The results show that the MAPE of the simulation prediction is 23.48 when the daily close price of the HS300 index is used as the input layer variable of the neural network. It shows that the GA-BP neural network model based on daily data can accurately predict the trend of daily close price data of HS300 index. In the future research, we will consider the trend of the highest price and the lowest price.

References

1. Lin Wang, Xinhua Bi. Risk assessment of knowledge fusion in an innovation ecosystem based on a GA-BP neural network [J].Cognitive Systems Research, 2021(66):201-210.

2. LI Hui, HU Cai-xia, LI Ying. Application of the Purification of Materials Based on GA-BP [J].Energy Procedia, 2012(17): 762-769.

3. Rui Song, Xuemin Zhang, Chen Zhou, Jing Liu, Jianhui He. Predicting TEC in China based on the neural networks optimized by genetic algorithm [J].Advances in Space Research, 2018(62):745-759.

4. Huajun Cao, Lei Liu, Bo Wu, Yuan Gao, Da Qu. Process optimization of high-speed dry milling UD-CF/PEEK laminates using GA-BP neural network [J].Composites Part B, 2021(221), 1-17.

5. Hongyang Wang, Zixin Zhang, Liming Liu. Prediction and fitting of weld morphology of Al alloy-CFRP welding-rivet hybrid bonding joint based on GA-BP neural network [J]. Journal of Manufacturing Processes, 2021(63):109-120.

6. Meng Zou, Long Xue, Hongjian Gai, Zhaolong Dang, Song Wang, Peng Xu. Identification of the shear parameters for lunar regolith based on a GP-BP neural network [J].Journal of Terramechanics, 2020(89): 21-29.

7. Yanni Peng, Wanli Xiang. Short-term traffic volume prediction using GA-BP based on wavelet denoising and phase space reconstruction [J].Physica A, 2020(549):1-14.

8. Zhenwei Wang, Yuan fei, Pengxin Ye, Fasheng Qiu, Guiyun Tian, Wai Lok Woo. Crack characterization in ferromagnetic steels by pulsed eddy current technique based on GA-BP neural network model [J].Journal of Magnetism and Magnetic Materials, 2020(500):1-9.

9. Dong Zheng, Zhen-dong Qian, Yang Liu, Chang-bo Liu. Prediction and sensitivity analysis of long-term skid resistance of epoxy asphalt mixture based on GA-BP neural network [J].Construction and Building Materials, 2018(158):614-623.

10. Shouxiang Wang, Na Zhang, Lei Wu, Yamin Wang. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method [J].Renewable Energy, 2016(94): 629-636.