Research on investment portfolio model based on neural network and genetic algorithm in big data era

Wei Zhou¹, Yuanjun Zhao², Weiwei Chen³, Yanghui Liu⁴, Rongjun Yang¹* and Zheng Liu⁵

Abstract

With the maturity of neural network theory, it provides new ideas and methods for the prediction and analysis of stock market investment. The purpose of this paper is to improve the accuracy of stock market investment prediction, we build neural network model and genetic algorithm model, study the law of stock market operation, and improve the effectiveness of neural network and genetic algorithm. Through the empirical research, it is found that the neural network model can make up for the shortcomings of the traditional algorithm through the optimization of genetic algorithm.

Keywords: Neural network, Genetic algorithm, Stock investment, Big data

1 Introduction

In the western countries with developed financial markets, the stock market has been running for more than 300 years, which to a certain extent reflects the development and trend of social economy in a specific period of time, and has a profound impact on economic development. In the early 1990s, China established its own securities market, and the listed companies in Shanghai and Shenzhen are increasing day by day. After nearly 30 years of accumulation, the market scale has reached a high level [1]. With the rapid development of the financial market, the enthusiasm of the public for financial investment is constantly improving, and the knowledge of financial investment is constantly increasing. However, the stock market investment is a high-risk and high-yield investment activity. The level of return and the degree of risk have a significant positive correlation. In other words, investors need to take a greater risk to get a higher return. In this context, a more reasonable stock market forecasting method can effectively reduce risk and increase returns. Therefore, it is of great significance to strengthen the research of stock market forecasting methods. In reality, the number of investors in the stock market is increasing, and investors also hope to obtain an effective analysis method to maximize the ratio of return to risk [2]. In recent years, with the general investors more and more understanding of the basic rules of stock market investment, as well as the increasing number of stock market investment researchers, stock market forecasting methods...
are also increasing [3]. It needs to be emphasized that at present, some stock market prediction methods emphasize ideal state, but due to the complex internal and external environment, various uncertain factors always impact the stock market investment market, which to a certain extent improves the prediction difficulty of the stock market, and greatly reduces the prediction effectiveness of the stock market investment [4]. Therefore, even the prediction methods with high popularity often fail in market prediction [5]. At present, the rapid development of science and technology provides a new way for stock market investment prediction and analysis, especially the growing maturity of neural network theory, which has been well applied in many aspects, such as signal processing, pattern recognition and so on. By analyzing the theory of neural network, it can be found that neural network has great advantages in self-adaptive and self-learning, has the characteristics of nonlinear approximation ability, and has a high degree of agreement with the stock market prediction [6]. Therefore, it is a good attempt to apply neural network to stock market prediction [7].

The specific contributions of this paper include:

1. A literature survey about various existing neural network and genetic algorithms, and analyze their advantages and disadvantages.
2. An effective neural network model and genetic algorithm model for improve the accuracy of stock market investment prediction is proposed.
3. Performance analysis of the proposed algorithm and an evaluation of the algorithm with respect to other existing algorithms.

The remainder of this paper is organized as follows: Related work will be introduced in Sect. 2. Neural network model and genetic algorithm model is explained in Sect. 3. Experimental results and discussion will be presented in Sect. 4 and conclusion will be drawn in Sect. 5.

2 Related work
In the process of social practice, people need to find the optimal solution in the complex system in order to solve the problem efficiently. However, because the solution space is relatively large, the correlation between the parameters and the target value is difficult to determine, and there are relatively many factors to be considered, so how to deal with the optimization problem must be highly valued. In many cases, people determine the approximate optimal solution by comparing and analyzing the random effective solution [8]. The essence of this method is to randomly extract the parameters of the domain of definition to obtain the optimal solution. This method is simple and easy, but it is only suitable for the field with small search space, but for the field with large search space, it cannot solve the problem simply by exhaustive method. More advanced optimization techniques are needed to solve the problem [9]. In contrast, the genetic algorithm with ‘survival of the fittest’ as the core has great advantages. By introducing competition mechanism into the algorithm, the search efficiency can be improved. The basic process of genetic algorithm is to determine a group of initial solutions in a random way, and then conduct individual search to obtain an independent solution, which is defined as a ‘chromosome’. Through the ‘fitness value’ index, the adaptability of chromosomes in
the population can be effectively evaluated, and then whether to select them to enter the next stage can be judged [10]. According to the principle of survival of the fittest, on the basis of continuous crossing, selection and variation, the evolution selection of chromosomes forms a chromosome group with higher adaptability. After reaching a certain number of iterations, the chromosome convergence is completed and the optimal solution of genetic algorithm is obtained [11]. By analyzing the process, we can find that the whole process of genetic algorithm is essentially similar to the genetic principle in biological sense [12].

At the operational level, genetic algorithm is not complex. According to the above discussion, it is essentially an iterative process, that is, it starts from the initial group of individuals, and obtains the approximate optimal solution through continuous cross selection and mutation operation [13].

Markov chain analysis is an important part of genetic algorithm, and its core is the convergence theory [14]. It can be found that the convergence of traditional genetic algorithm is generally based on the Markov chain limit theory. In the practice of solving problems, the ultimate goal of genetic algorithm is to determine the global optimal solution. The essence of the whole process is random search, which has great uncertainty [15]. The operation process of genetic algorithm is continuously optimized under the expected value of the optimal solution, and it is regarded as the initial sequence [16]. Through the convergence theory of genetic algorithm, its convergence can be effectively verified. Not only that, in order to achieve good convergence of genetic algorithm, we must focus on two parameters, one of which is the possibility of breaking away from the satisfactory solution set on the premise of determining the satisfactory solution; the other is the possibility of still not obtaining the satisfactory solution on the premise of not obtaining the satisfactory solution, and the convergence of genetic algorithm is formed on the basis of the above two parameters matching General theory [17]. The convergence research based on the two parameters is pure probability research, which is intuitive and simple in the convergence verification of genetic algorithm [18].

3 Methods
The purpose of this paper is to improve the accuracy of the stock market investment prediction. By combining the neural network model and the genetic algorithm model, we can predict the operation law of the stock market [19]. This paper relies on the existing theoretical research results to optimize the real number coding scheme and improve the effectiveness of neural network algorithm and genetic algorithm. In this paper, the real number coding method is adopted, the sample segmentation is optimized, the training is strengthened, the training speed and convergence speed of neural network are improved, the local minimum value is obtained, so as to avoid falling into, the three-layer neural network is constructed to determine the global optimal solution, so as to effectively solve the problem.

3.1 Genetic algorithm to optimize the learning rules of neural network
In the training process, neural network learning rules need to be set in advance [20, 21]. However, whether the learning rules are reasonable or not is uncertain [22, 23]. Therefore, it is necessary to design and optimize neural network learning rules with the help
of genetic algorithm, so as to improve the ability of neural network algorithm to solve complex problems and the adaptability of the algorithm to uncertain environment [24]. Research results show how to design coding by learning rules is the core problem in the evolution process, so far there are no cases with application value [25, 26]. Therefore, the study of learning rules is only the initial stage, and its process is as follows:

Step 1 The effective coding method is determined, the learning rules are coded, and the matching between individual and single learning rules is realized.

Step 2 To construct a training set, the elements of the training set are determined firstly, and then the corresponding learning training is carried out according to the matching learning rules.

Step 3 Calculate the fitness of all learning rules.

Step 4 Select and determine learning rules that meet the requirements.

Step 5 Cross selection, individual variation processing, analysis of individual attributes, to determine the next generation of population.

Step 6 Repeat the above steps until the goal of evolution is achieved.

In this paper, after the optimization of genetic algorithm, the connection weight of neural network is improved. By solving the existing problems of neural network, the generalization function of neural network is enhanced [27]. On this basis, the learning model of neural network is constructed, and the global optimal solution is obtained to achieve the ability of solving specific problems.

3.2 GA-BP algorithm design

Parameters play a decisive role in the performance of the algorithm model. In this paper, finite length coding is chosen. After the design of the algorithm coding scheme is completed, the parameter coding is transformed into the genetic algorithm coding, and then the function used to accurately evaluate the algorithm performance is determined, and the global search is completed in the parameter space. In this way, not only the space can be expanded, but also the target of regional search can be realized, and a balance state can be achieved between the two. In the initial stage of genetic search, due to the uncertainty brought by cross variation, the search scope has been expanded to a certain extent. After obtaining the high fitness solution, the crossover operation completes the search near the above solution. Therefore, through the genetic operation, we can determine the best combination of parameters to meet the requirements of practical application and solve the problem. In terms of algorithm implementation, the specific process is as follows:

1. Step 1 Randomly forming $n$ codes and forming initial set $s$.
2. Step 2 Complete the coding in sequence, decode the coding, determine a parameter combination $P$ reflecting BP model, determine BP, evaluate the BP and obtain its corresponding fitness value.
3. Step 3 According to the appropriate value determined in the previous step, determine $n$ individuals, and enter the next generation to obtain the next generation group. In this step, some individuals may need to be selected multiple times.
(4) **Step 4** According to the probability $P$ and the fitness value of different codes, the parent generation is determined, then cross inheritance is carried out, and the next population is entered after random pairing.

(5) **Step 5** According to the probability $P$ and fitness, select the parent population that meets the requirements, insert new individuals through mutation inheritance, and achieve the goal of population iteration.

(6) **Step 6** By repeating the above steps repeatedly, the search target can be achieved on the premise of meeting the standard requirements.

\[ w < x - 1 \]  
\[ w < \sqrt{(x + y) + a} \]  
\[ w = \log_2 x \]

In which, $x$ is the number of input layer nodes, $w$ is the number of hidden layer nodes, $a$ is a constant between 0 and 10.

### 4 Results and discussion

#### 4.1 Neural network toolbox

As a highly complex and comprehensive algorithm model, neural network model has relatively high requirements for toolbox. Through the application of neural network toolbox, the goal of activation function can be realized. At the same time, through the algorithm training, the network designer can complete the specific subroutine, and based on this, promote the learning training, complete the corresponding call requirements to the greatest extent, and improve the effectiveness of neural network learning.

In the process of building algorithm model, different types of algorithms are integrated into neural network toolbox, so as to improve the convenience of algorithm design.

#### 4.2 Genetic algorithm toolbox

All kinds of algorithms of genetic algorithm ultimately need to act on chromosomes. Chromosomes are essentially vector types, which can be reduced to specific matrices, and matrix operations form operators. It can be seen that the basic data unit has the infinite feature of matrix dimension. Based on this recognition, users can ignore the low-level problems related to matrix algorithm, so as to improve the operation efficiency on the basis of programming. In the application process of genetic algorithm, the toolbox can provide the necessary algorithm with the characteristics of scalability and standardization. Through its matrix computing ability, it can improve the efficiency of genetic algorithm, reduce the difficulty of chromosome programming, and improve the difficulty of solving problems.

#### 4.3 Sample data

The number of samples is closely related to the accuracy of genetic algorithm, it is closely related to the complexity of mapping relationship, and also to a certain extent depends on data noise. With the increase of the complexity of mapping relationship, the number
of training samples is also required to be higher, and with the increase of noise, the number of samples is also increased synchronously. In the aspect of sample selection, we need to adhere to the following three principles: (1) meet the requirements of sample quantity; (2) meet the requirements of sample accuracy; (3) meet the requirements of sample representativeness; (4) meet the requirements of sample distribution.

Stock price return is a common index in the quantitative analysis of stock market investment. However, if the stock market is positioned as a nonlinear dynamic system, the return is not the optimal price alternative transformation, and the factor of forecasting price should be fully considered. Not only that, with the increase of the number of training samples, the amount of calculation will be increased to a large extent, and the convergence speed in the training process will decline, which requires a longer convergence time. If the number of samples does not meet the requirements, the network will not be able to fit the corresponding stock index curve. In this paper, ‘gzmt’ is chosen as the representative of empirical analysis, and its stock number is 600519. In terms of data selection, this paper adheres to the basic principles of representativeness, continuity, uniform distribution and accuracy to improve the adaptability of this algorithm.

If we want to reduce the prediction error and improve the prediction accuracy, we need to choose a reasonable number of samples to meet the training requirements of neural network. On the basis of fully considering the prediction characteristics of neural network, the training samples are optimized to improve the convenience of detection.

Before the formal learning, the effectiveness of the network largely depends on data processing. Data processing will affect its accuracy and speed. Generally speaking, the network training cannot directly apply the samples obtained, it needs to complete the necessary processing before it can be put into application. In other words, the acquired data samples usually need to be normalized before they can be applied to network training.

4.4 Closing price network model

In this paper, 277 historical closing prices are selected as data samples to build the basic data model of stock research. In order to ensure that the data samples meet the use requirements, the data samples must be normalized before they can be put into use. The input node of the network is \( p = 5 \), and the output node is \( t = 1 \), that is, the closing price data of five trading days before \( t + 1 \) is used for prediction. At present, the hidden layer of neural network cannot be determined. In this paper, on the basis of ensuring that the error meets the requirements, in order to reduce the calculation difficulty to the greatest extent, the best number of hidden layers is verified through experiments, so as to obtain the number of hidden layers in a reasonable range. The numerical results show that only through the L-M training method can a high-speed three-layer network be established. At present, 5–12–1 network structure is the most widely used, which uses 5 input nodes, 1 output node and 12 hidden layer nodes. Because BP network security has the characteristics of network generalization, 159 training samples and 122 test samples can be selected according to the data samples. Figure 1 is schematic diagram of stock market forecast.

The main goal of this paper is to optimize BP network and build an efficient and accurate operation model based on genetic network. In the process of forecasting the closing
price of GZMT, Guizhou Province, after the introduction of the optimized network model, necessary testing and training are needed. Figure 2 is test chart of BP network model.

In this paper, based on the operation tools provided by neural network toolbox, combined with the algorithm discussed above, and through programming, the calculation of network closing price is completed. After completing the training, we need to rely on the test set to carry out the necessary tests, and then we can determine the samples that integrate with the stock index curve. Figure 3 is neural network fitting curve.

Through sample training, the number of iterations is determined when the target error meets the requirements. According to the network model constructed in this paper, the error curve can be calculated. Figure 4 is error curve.

By training 150 groups of sample data, the error can be reduced to the greatest extent, and most of the errors are reduced to about 0. Therefore, the learning training in this
paper has basically achieved the expected goal. We found that the neural network model has high accuracy in prediction. Figure 5 is fitting curve of test sample.

After three times of training, the SSE of sample data can be determined as $11.5857 \times 10^{-4}$, which achieves a relatively good fitting effect, and the error analysis also shows that it achieves a good effect. Therefore, we choose L-M back propagation algorithm to carry out the learning and training, which not only has a good approximation effect, but also can achieve a high speed of operation. The most important thing is to avoid the local minimum problem, which needs to occupy a relatively large memory space. It should be emphasized that in the process of learning and training, it is necessary to track and observe the learning rate and target error, so as to reduce the impact on the convergence speed as much as possible. To sum up, this paper can effectively improve the efficiency of stock market prediction through artificial neural network algorithm, which has high application value.

4.5 Prediction and analysis of stock market based on GA-BP network model

The specific process of GA-BP network model is as follows:
Step 1 Determine the initial function.
Step 2 Complete the fitness training.
Step 3 Obtain the initial population.
Step 4 Call the genetic function.
Step 5 The weight and threshold of neural network are determined.
Step 6 To construct the targeted network training.
Step 7 The network is determined by weight and threshold.
Step 8 Tracking and observing the network performance.
Step 9 Solve the prediction problem according to the network.

The neural network designed in this paper selects three-layer structure, inputs the closing price of \( T - 4 \) day, \( T - 3 \) day, \( T - 2 \) day, \( T - 1 \) day and \( t \) day, outputs the closing price of \( T + 1 \) day, and the learning rate is 0.5. After the following conditions are met, the learning is finished:

Step 10 Goal \( \leq 1e^{-5} \) or epochs \( \leq 1000 \).
Step 11 Select 100 iterations, and use GA-BP network training as shown in the figure below. Figure 6 is GA-BP network training. Figure 7 is training iteration diagram.

By comparing the output value with the actual value, it can be found that the neural network test sample in this paper achieves good prediction effect. Figure 8 is GA-BP fitting curve of test sample.

Compared with other algorithms, it is difficult to evaluate through the test indicators in econometric technology. Therefore, this paper summarizes the most widely used centralized evaluation indicators based on the research results at home and abroad.

Mean absolute error:
According to the above table, it can be found that compared with the general neural network, the real coded genetic algorithm optimization model not only has high accuracy and good effect, but also can prevent the occurrence of local minima, and can improve the convergence speed, which has high practical value. Table 1 is comparison results of three algorithms.

\[
\text{MAE} = \frac{\sum |e_j|}{N} \quad (4)
\]

Squared difference:
\[
\text{SSE} = \sum (e_j)^2 \quad (5)
\]

Mean square deviation:
\[
\text{MSE} = \frac{\sum (e_j)^2}{N} \quad (6)
\]
5 Conclusion

With the development of intelligent technology, neural network has become a frontier interdisciplinary research field, which is conducive to improving the design level and comprehensive performance of neural network system. Based on the existing research results, this paper improves the application efficiency of the algorithm by optimizing the specific links of the neural network. It systematically explains the technical terms and relevant indicators of the stock market investment, explains the common prediction methods, and systematically discusses the research hotspot and difficulty. This paper systematically discusses the related theory and application process, and analyzes the similarities and differences between genetic algorithm and neural network technology, which can avoid the problem of limited minimum. Choose the stock market of our country as the research sample, and then get the feasibility of short-term forecasting stock market, which lays a theoretical foundation for this study. Scientific and reasonable selection of input parameters can not only reflect the scale and quality of information in the stock market, but also avoid the problem of learning and training difficulty or even non-convergence due to information overlapping. As the stock market changes with time, the network training samples also need to be adjusted accordingly, otherwise it will be difficult to guarantee the accuracy of prediction. Therefore, the preprocessing of the original data is an essential part. Through the empirical study, it is found that the neural network model can make up for the existing problems of the common algorithm through the optimization of genetic algorithm.

Authors’ contributions

WZ designed research, performed research, analyzed data, and wrote the paper. All authors analyzed the data and were involved in writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Social Science Foundation of China (Grant Number: 18CGL015).

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests

The authors declare that they have no competing interests.

Author details

1 Department of Economy and Management, Zaozhuang University, Zaozhuang 277100, China. 2 School of Business Administration, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China. 3 College of Business Administration, Hunan University of Technology and Business, Changsha 410205, China. 4 School of Economics and Management, East China Jiaotong University, Nanchang 330000, China. 5 School of Management, Shanghai University of Engineering Science, Shanghai 201620, China.

Received: 10 August 2020   Accepted: 21 October 2020

Published online: 02 November 2020

References

1. S. Asadi, E. Hadavandi, F. Mehanmazir, M.M. Nakhostin, Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. Knowl. Based Syst. 35(11), 245–258 (2012)
2. P.C. Chang, C.H. Liu, J.L. Lin, C.Y. Fan, C.S. Ng. A neural network with a case based dynamic window for stock trading prediction. Expert Syst. Appl. 36(3), 6889–6898 (2009).

3. M. Jasemi, A.M. Kimiagari, A. Memariani. A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick. Expert Syst. Appl. 38(4), 3884–3890 (2011).

4. M. Shaverdi, S. Fallahi, V. Bashiri. Prediction of stock price of Iranian petrochemical industry using GMDH-type neural network and genetic algorithm. Appl. Math. Sci. 6(7), 319–332 (2012).

5. R.J. Kuo, C.H. Chen, Y.C. Hwang. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. Fuzzy Sets Syst. 118(1), 21–45 (2001).

6. A.H. Moghaddam, M.H. Moghaddam, M. Esfandaryari. Stock market index prediction using artificial neural network. J. Econ. Finance Adm. Sci. 21(41), 89–93 (2016).

7. R.L. Naik, D. Ramesh, B. Manjula, A. Govardhan. Prediction of stock market index using genetic algorithm. Comput. Eng. Intell. Syst. 3(7), 162–171 (2012).

8. Z.H. Khan, T.S. Alin, M.A. Hussain. Price prediction of share market using artificial neural network (ANN). Int. J. Comput. Appl. 22(2), 42–47 (2011).

9. Y. Chen, B. Yang. Flexible neural trees ensemble for stock index modeling. Neurocomputing 70(6), 697–703 (2007).

10. J.J. Wang, J.Z. Wang, Z.G. Zhang, S.P. Guo. Stock index forecasting based on a hybrid model. Omega 40(6), 758–766 (2012).

11. S. Wang, X. Wang, F. Meng, R. Yang, Y. Zhao. Investor behaviour monitoring based on deep learning. Behav. Inf. Technol. 2020. https://doi.org/10.1080/0144929X.2020.1717627

12. C.F. Tsai, Y.C. Hsiao. Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches. Decis. Support Syst. 50(1), 258–269 (2010).

13. W. Sun, Y. Zhao. Big data analytics for venture capital application: towards innovation performance improvement. Int. J. Inf. Manag. 50(2), 557–565 (2020).

14. L. Sun, Y. Zhao, W. Sun. Study on supply chain strategy based on cost income model and multi-access edge computing under the background of the Internet of Things. Neural Comput. Appl. 2019. https://doi.org/10.1007/s00521-019-04125-9

15. F. Wang, L. Ding, H. Yu et al. Big data analytics on enterprise credit risk evaluation of e-Business platform. IseB (2019). https://doi.org/10.1007/s10257-019-00414-x

16. R. Yang, L. Yu, Y. Zhao, H. Yu, G. Xu, Y. Wu, Z. Liu. Big data analytics for financial Market volatility forecast based on support vector machine. Int. J. Inf. Manag. 50(2), 452–462 (2020).

17. Z. Liu, S. Chen, B. Hu, M. Zhou, Y. Zhao. Research on staged pricing model and simulation of intelligent urban transportation. IEEE Access 7(9), 141404–141413 (2019).

18. M. Xia, H. Li, Y. Zhao. Blockchain financial investment based on deep learning network algorithm. J. Comput. Appl. Math. 372(7), 112723 (2020).

19. S. Cai, X. Wang, Y. Zhao. Revenue model of supply chain by internet of things technology. IEEE Access 7(1), 4091–4100 (2019).

20. V. Cho, MISMS—a comprehensive decision support system for stock market investment. Knowl. Based Syst. 23(6), 626–633 (2010).

21. F.C. Yuan, C.H. Lee. Using least square support vector regression with genetic algorithm to forecast beta systematic risk. J. Comput. Sci. 11(11), 26–33 (2015).

22. Y. Sun, C. Xu, G. Li, W. Xu, J. Kong, D. Jiang, B. Tao, D. Chen. Intelligent human computer interaction based on non redundant EMG signal. Alex. Eng. J. 59(3), 1149–1157 (2020).

23. F. Long, N. Xiong, A.V. Vasilakos, L.T. Yang, F. Sun. A sustainable heuristic qos routing algorithm for pervasive multi-layered satellite wireless networks. Wirel. Netw. 16(6), 1657–1673 (2010).

24. J. Li, N. Xiong, J.H. Park, C. Liu, M.A. Shihua, S.E. Cho. Intelligent model design of cluster supply chain with horizontal cooperation. J. Intell. Manuf. 23(4), 917–931 (2012).

25. W. Wei, B. Zhou, D. Polap, M. Wolsnak. A regional adaptive variational PDE model for computed tomography image reconstruction. Pattern Recogn. 93(8), 64–81 (2019).

26. Z. Zhang, D. Chen, Y. Zhang, X. Chen, M. Zhang, C. Wu. Deep learning for autonomous ship-oriented small ship detection. Saf. Sci. 130(10), 104812 (2020).

27. W. Guo, N. Xiong, A.V. Vasilakos, G. Chen, C. Yu. Distributed k–connected fault–tolerant topology control algorithms with PSO in future autonomic sensor systems. Int. J. Sens. Netw. 12(1), 53–62 (2012).

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.