https://doi.org/10.1051/itmconf/20224501068

Research on stock index forecast based on the forgetting-genetic network model

Bin Wei*, *Mingshu* Zhang, and *Longfei* Liu

College of Cryptographic Engineering, Engineering University of Armed Police Force, Xi'an Shaanxi, China

Abstract. The stock market plays an important role in national economic development. Predicting its trend, to ensure market stability, efficiency, and safety, is one of the hottest topics of current research. Some algorithms have been proposed to use in this study, however, the drawback in forecasting accuracy and applicability of the model and other issues limit its application in practice. In this paper, we designed a forgotten-genetic network model to predict the stock index. Finally, a comparison between existing algorithms to verify the validity and applicability of the proposed method in this study.

1 Introduction

With the rapid development of China's economy and its gradual integration with the global economy, the operating conditions and underlying laws of China's financial system have an important impact on the country's overall economic strength, and even directly affect the future development trend of the global economy. As one of the indispensable components, stocks play an important role. How to predict and analyse their trends to ensure the stability, efficiency and safety of the market is one of the hot issues to be solved urgently.

At present, many scholars at home and abroad have conducted analysis and research on this problem, put forward some prediction and analysis methods, and achieved good results. Such as MU [1] proposed a hybrid model (GARCH) for a comprehensive analysis of stock price volatility. However, these algorithms have some shortcomings such as prediction accuracy that cannot meet actual needs, only focusing on a certain point in time, ignoring the impact of the period on the results, and lack of analysis of the index fluctuation range, so that their use in a wider range is affected.

In addition, existing studies have shown that network-based methods can obtain relatively good results. Therefore, this paper designs a self-organizing and self-learning network model based on a genetic algorithm to predict the index. The algorithm is based on "the forgetting model" adopts the principle of the least weighted error to select the best network structure, and formulates short-term forecasts through systematic analysis of factors such as trends, cycles and irregular component changes. Finally, experiments (including the comparison of the convergence speed of the nonlinear sampling sequence

* Corresponding author: weibin82@126.com

and the actual prediction of the Shanghai Composite Index) verify the effectiveness of the algorithm in this paper and its applicability in the field of research.

2 Forgetting-genetic network model

The neural network has received general attention and extensive research and application in various research fields. However, currently commonly used neural networks all have the problems of sensitivity to initial values, slow convergence speed, and local optimization, while genetic algorithm (GA) proposed by Professor Holland [1], the basic framework of the algorithm is composed of code, fitness function, selection operator, crossover operator, mutation operator and other parts [2]. Therefore, this paper combines the advantages of the above two algorithms organically and uses the characteristics of a genetic algorithm to design a self-learning network model with a forgetting strategy.

2.1 Coding

Although the binary encoding method has universal advantages, its accuracy is closely related to the number of bits of the encoding. For the network, the binary encoding method has a recognition "blind spot", thereby reducing its efficiency. Real number encoding is the most direct description of the problem domain. Therefore, it does not require complex decoding, and at the same time, it can also have relatively high expression accuracy. Based on these advantages, we choose to use real number encoding in this article.

2.2 Chromosome coding

The chromosome in the article is divided into 4 parts, and its structure is $N_{hid} \mid F_{hid} \mid W \mid B$.

The structure of each part and the initialization method are as follows:

1) Number of intermediate nodes N_{hid}: according to formula (1) get $[N_{\min}, N_{\max}]$, Popsize individuals, are uniformly and randomly generated in the interval $[N_{\min}, N_{\max}]$.

2) Activation function F_{hid}: the candidate activation functions *F*= myTansig, Tansig, Logsig, myLogsig, and purelin are numbered 1 to 5 respectively.

3) Connection weight *W*: according to the intermediate nodes N_{hid}, the corresponding connection weights between the first two layers and the next two layers are randomly generated according to the vacancy probability and normal respectively P_{n1} P_{n2}.

4) Threshold *B*: normally generate the threshold of each layer node randomly.

2.3 The forgetting model

In addition, considering that the most recent samples have a more important influence on subsequent forecasts, that is to say, their forecast weights are different according to the distance relationship, which we call the "forgetting model". Therefore, in the process of constructing the predictive network model, this paper does not directly select the network structure with the smallest training error (that is, the individual with the highest fitness value), but design and use the "forgetting weighted confirmation" according to the "forgetting model". The step is to select the network with the best generalization performance from the results of a genetic optimization.

In summary, the main steps of the forgetting-genetic network algorithm are as follows:

1) Divide the sample into three parts according to the time sequence number: training sample, verification sample, and prediction area.

2) The population decoding obtains the network structure and further calculates the network output and the training error of each chromosome on the training samples.

3) Each generation takes the optimal network structure and enters the verification sample to obtain the corresponding verification error.

4) After the max loop generation genetic evolution, a set of error statistics can be obtained. Since the verification sample is closer to the prediction area, according to the "forgetting model", it is obvious that a network with a smaller verification error is more likely to obtain better prediction results. Therefore, this paper uses a weighting method to select the best prediction network structure and extracts the training error through regular sliding and the verification error, to prevent the network from training "overfitting".

3 Experiment and result analysis

3.1 Sample selection

In this paper, the comprehensive closing price of the Shanghai Stock Exchange Index is selected as the forecast object, with a total of 215 trading day samples from 2020.01.01 to 2020.11.25. In addition, because single-step forecasts are more contingent and do not have practical research and analysis value, this paper mainly focuses on short-term (3-day) multistep forecasts.

3.2 Experimental results

1) Enter 15 days, forecast 3 days

Table 1 shows the accuracy of the experiment. It can be seen that the algorithm in this paper has a relatively high short-term predictive value. Judging from the statistical results of the accuracy of each grouping, the prediction effect of the fourth group and the third group is relatively the worst. The last group has the highest accuracy rate. This is because the fourth group has been rising continuously, and the rebound has formed, so the nearest group 5 is 3 days away. The predicted value appears relatively simple.

Table 1. 15-3 Forecast accuracy rate statistics.

Number	Group1 Accuracy	Group2 Accuracy	Group3 Accuracy	Group4 Accuracy	Group5 Accuracy
1	0.8633	0.9641	0.8433	0.6737	0.9675
2	0.8223	0.8492	0.7122	0.6248	0.9731
3	0.7578	0.9923	0.8935	0.6936	0.9323
4	0.8401	0.7031	0.7043	0.7333	0.9843
5	0.8163	0.8553	0.7632	0.6556	0.9189
average	0.8120	0.8728	0.7833	0.6762	0.9552

2) Enter 7 days, forecast 3 days

The intermediate node obtained by the algorithm are 23, and the prediction experiment is repeated 5 times. The results are shown in Table 2. It can be seen from the table that the prediction effect of the bottom rebound area is still worse than the prediction of the decline and rise at both ends. The average accuracy of the 5 experiments has decreased significantly compared with Table 2, which indicates that the network structure at this time is difficult to predict the basics. The prediction accuracy of the network with the input of 15 days is higher than that of the network with the input of 7 days, and it is the most obvious in group 1, and it is also higher than 10% in group 4 (that is, the rebound zone). This shows that increasing the length of the training window to a certain extent helps to improve the accuracy.

Table 2. 7-3 Forecast accuracy rate statistics.

Number	Group1 Accuracy	Group2 Accuracy	Group3 Accuracy	Group4 Accuracy	Group5 Accuracy
1	0.5952	0.6623	0.6705	0.5473	0.8893
2	0.8636	0.8616	0.6016	0.6391	0.9330
3	0.5396	0.8573	0.4031	0.6009	0.6584
4	0.8644	0.7169	0.6119	0.5079	0.7030
5	0.5719	0.8831	0.6319	0.5492	0.8014
average	0.6869	0.7962	0.5838	0.5689	0.7970

Finally, to further verifies the effectiveness of the method in this paper, we have selected several network-based methods (GARCH, BPPM, EMD-RBF) to compare with them, and the results are shown in Table 4. From Table 4, we can see that several network algorithms performed better in this experiment. However, compared with other methods of all tests, this algorithm achieved the best results. Although the advantages in some tests are not particularly obvious, we can still conclude that the method designed in this paper has certain advantages compared with other algorithms and is more suitable for this research.

Table 3. Comparison of experimental results.

Method	Group1 Accuracy	Group2 Accuracy	Group3 Accuracy	Group4 Accuracy	Group5 Accuracy
GARCH	0.8015	0.8709	0.7501	0.6633	0.9527
BPPM	0.7798	0.7053	0.7735	0.6701	0.9390
EMD-RBF	0.7625	0.7083	0.7194	0.6305	0.9404
Our method	0.8120	0.8728	0.7833	0.6762	0.9552

4 Summary

Aiming at the problem of stock index prediction, based on the analysis of the deficiencies of existing methods, this paper designs a prediction method based on the forget-genetic network model based on the characteristics of this research, the number of nodes in the middle layer, activation function, initial threshold, and weight, etc. The elements of network structure are all self-organized learning by the forgetting-genetic algorithm. In the network structure, the algorithm designs a more reasonable coding structure that fully considers the role of verification errors, and uses weighted errors as the basis for the selection of the final network structure. Finally, the performance of the method in this paper is proved through the experiment of piece wise nonlinear function sequence and the method of stock index prediction.

References

1. MU Zeping, l.S., GARCH-neural network model for forecasting the volatility of bid-ask spread of the Chinese stock market. Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition), 2015. 27(1): p. 130-136.
2. Lilin, L.M.L.Y.J., Hybird multipopulation cellular genetic algorithm and its performance. Transactions of Nanjing University of Aeronautics and Astronautics, 2014. 31(4): p. 405-412.

3. Chi W L. Forecasting stock index based on BP neural network algorithm[C]. International Conference on Mechanical, Electronic, Control and Automation Engineering, 2018, 149(2):751-755.
4. Liu Z J, Yu M J, Fang Y. Correlation analysis of financial indicators and stock price fluctuation via BP neural network[C]. International Conference on Social Science, Economics and Management Research, 2019:89-96.
5. Qiu M, Song Y, Akagi F. Predicting the return of the Japanese stock market with artificial neural network [J]. Industrial Engineering and Management Science, 2015:7-12.