2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Improving the Genetic-Algorithm-Optimized Wavelet Neural Network for Stock Market Prediction

Yu Fang, Kamaladdin Fataliyev, Lipo Wang*, Xiuju Fu, Yaoli Wang

***Abstract*—This paper improves stock market prediction based on genetic algorithms (GA) and wavelet neural networks (WNN) and reports significantly better accuracies compared to existing approaches to stock market prediction, including the hierarchical GA (HGA) WNN. Specifically, we added information such as trading volume as inputs and we used the Morlet wavelet function instead of Morlet-Gaussian wavelet function in our prediction model. We also employed a smaller number of hidden nodes in WNN compared to other research work. The prediction system is tested using Shenzhen Composite Index data.**

***Index Terms*—Stock Market Prediction; Wavelet theory; Wavelet Neural Networks; Genetic Algorithms.**

I. INTRODUCTION

Stock markets are highly volatile and have always been a challenge to both economists and researchers [1][2]. In recent years, intelligent methods like wavelet neural networks (WNNs) [3][4] have been employed in this field [5][6], since wavelets can decompose economic time series into their time scale components [7][8][9][10][11] and it is a successful strategy to unravel the relationship between economic variables [7]. However, WNNs based on back propagation (BP) has two prominent vulnerabilities: slow convergence and local minima. For solving these problems, global search algorithms, such as genetic algorithms, have been used to optimize WNNs (e.g., [12]). Zhou and Wei have used WNNs optimized with hierarchical genetic algorithms (HGA) for stock market modeling and forecasting [13].

In this paper, we aim at improving the GA-optimized WNN approach to stock market prediction. We have proposed new optimization objectives and varying network parameters, such as wavelet types, hidden layer sizes, GA generations, GA population sizes, and GA topologies for improving WNN's performance in stock market prediction.

II. A REVIEW ON WAVELET NEURAL NETWORKS

A. Wavelet Theory

The Fourier theory enables a signal to be expressed as the sum of sines and cosines. However, a severe disadvantage is that no time resolution but only frequency resolution is in Fourier transform. One is not able to identify when the signal is present. Wavelet theory was proposed to overcome this problem.

Yu Fang, Kamaladdin Fataliyev, and Lipo Wang (corresponding author) are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (email: elpwang@ntu.edu.sg)

Xiuju Fu is with Institute of High Performance Computing, Singapore (email: fuxj@ihpc.a-star.edu.sg).

Yaoli Wang is with College of Information Engineering, Taiyuan University of Technology, Taiyuan, China.

A wavelet is described as a function with a zero average:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{1}$$

A wavelet is usually created from a mother wavelet $\psi(t)$ by scaling and translation. The fundamental of wavelet transform is to use wavelets to represent any function, so that there will be information extracted about both time and frequency domains.

There are two major categories of wavelet transforms: Discrete Wavelet Transforms (DWT) and Continuous Wavelet Transforms (CWT). Similar to Fourier transforms, CWT decomposes the signal into infinite number of scales and translations, providing a large amount of information.

B. Wavelet Neural Networks

By taking advantage of both the scaling properties of wavelets and the learning ability of neural networks, wavelet neural network (WNN) was first proposed by Zhang and Benveniste in 1992 [14]. It is a special feed-forward network supported by the wavelet theory. The advantages of the WNN include:

- In contrast to the random initialization commonly used in neural networks, WNNs can be initialized with regular pyramid type wavelet lattice, which is derived from the multi-resolution representation of wavelet analysis [14][15]. This method is intuitively more reasonable than random initialization.
- Taking observation data into account, pyramid-type initialization of the WNN can be even more efficient.
- With support from the wavelet theory, the parameters in the WNN have explicit meanings.

The WNN contains hidden neurons and normally has only one hidden layer. The differences from the traditional neural networks exist in WNN's hidden layer, where neurons use wavelets as their activation functions. These neurons are known as wavelons, with their respective translations and scales. It is commonly seen that the wavelon takes multiple inputs, in which case it is called multi-dimensional wavelons. The output of this wavelon then becomes the tensor product of the 1-dimensional wavelons.

In real-world applications, different designs might be adopted: bias terms on the output layer, feedbacks, direct connections between input and output layers, etc. Besides these variations, based on how scales and translations on wavelets are determined, WNNs may be divided into two categories: adaptive WNNs and fixed grid WNNs [16]:

1) Adaptive WNN: this type of WNN uses wavelets from the continuous wavelet transforms. The dilation and translation for each wavelet are adjustable parameters in the network, just like network weights.

2) Fixed grid WNN: This type of WNN use wavelets from discrete wavelet transforms and the scales and translations are all fixed to predetermined levels, following the dyadic rule. Geva [8] cascades wavelet-based signal processing techniques and artificial neural networks to perform time-series prediction. In that research, Discrete Wavelet Transform is applied to decompose the time-series signal into separate scales of wavelets. The resulted components are then put into separate ANNs to perform prediction in respective scales and after which, the prediction results are re-combined to form the prediction of original time series. This approach involves separate initialization and training on multiple networks. As a result, higher degree of complexity and possibly low computational efficiency might be expected in this model.

WNN has been a powerful tool for signal approximation. Many applications can be found in different domains. In recent decades, WNN had been used for predicting stock market trends. Lahmari [17] has tested various learning algorithms with adaptive WNNs on S&P 500 index price prediction and ranked their performances. It will be served as a valuable comparison resource in the later section.

III. A REVIEW ON GA WNNS

As the problems of slow convergence and local minima are presented in WNN, researchers are working to overcome the drawbacks of WNN by many approaches. One study which worked towards solving the challenges by integrating WNN with genetic algorithms was carried out by Ornes and Sklansky [18]. They had proposed such a model to predict S&P 100 index movement direction. The inputs of the system are: past index values, total daily transaction volume and the Lehman bond index. Their system performed wavelet transformation first and then used genetic algorithms as a feature selection tool before applying to the neural network. In comparison, Jiang and Huang [12] proposed a system to optimize weight of a WNN using adaptive annealing genetic algorithms. The study also calculated the selection, crossover and mutation probability, improved the convergence and stability of genetic algorithms to find out the optimized global solution.

Zhou and Wei [13] used hierarchical genetic algorithm (HGA) to determine the WNN structure, wavelet parameters including weights etc., all at the same time. They used this HGA-WNN to predict Shenzhen Composite Index and Wanke Stock Price. The inputs to the network are index values on the last five days and predicting target is the index value on the next day. This approach adopted both control genes and parameter genes to form a hierarchical structure. It treated topology and weights etc. as a chromosome. It involved bi-objective optimization. As a result, higher degree of complexity and possibly lower computational efficiency are expected in this model. In contrast, based on a widely accepted guiding principle, a guideline [16] is "in the situation with no prior knowledge, the simplest network, namely the smallest network complying with the given sample is the best choice." This motivated us to simplify the network structure and parameters.

IV. GA WNN STOCK MARKET PREDICTION MODEL

A. Network Structure Design

In this *study*, we use a 3-layer WNN to obtain the predicted value of next day stock opening price. Adaptive WNN is employed.

The inputs of the system are the historical values of opening, highest, lowest and closing prices, the adjusted closing price for dividends and splits and the trading volume. The final system is tested on Shenzhen Composite Index. To make the training more efficient, the input data is normalized and fitted to the range of [-1, 1]. GA is used to find the best parameter and weight values. After training is completed, the best individual with optimal fitness will place its parameter values into the system. By trial-and-error, the GA parameters are set: population size is 40, generation is 50, cross-over probability is 0.7, and stopping criteria is determined when maximum generation reached.

B. Selection of Wavelets

There are a number of wavelets available, such as Haar wavelets, Daubechies wavelets, Coiflet wavelets, Symlet wavelets, Morlet wavelets, etc. In this paper, the Morlet wavelet was chosen as the mother wavelet for its continuous support, explicit expression, differentiability and its capability to hold high order structural changes.

C. Initialization

For a WNN, its initial weight and parameter (scale factor and translation factor) value is usually generated by random inputs. However, those random values could make the network produce results far away from good solution or even stuck at local optimum. Thus the network might not achieve desirable output due to poor training effort. For better initializations, GA is adopted in this study for the calculation of initial weights and parameters for the training of the network.

D. Training Method

It has been shown by Randall and Martinez [19] that batch training is much slower than on-line training especially on large training sets, as on-line training can follow the error surface closely and allow using larger learning rate. Hence, this paper uses on-line training as the training method.

V. APPLICATION IN STOCK MARKET PREDICTION

A. Measure of Performance

It is necessary to clearly define how system performance is measured. In this study, Mean Square Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to measure the performance. These parameters are calculated as follows:

$$\text{MSE} = \frac{1}{N}\sum_{N}(Y_{\Pr edict} - Y_{Actual})^2 \qquad (2)$$

$$\text{MAE} = \frac{1}{N}\sum_{N} |Y_{\Pr edict} - Y_{Actual}| \qquad (3)$$

$$\text{MAPE} = \frac{1}{N}\sum_{N} \left|\frac{Y_{\Pr edict} - Y_{Actual}}{Y_{Actual}}\right| \qquad (4)$$

In addition, a direction prediction could assist investors to decide whether to buy, sell or hold in the stock. Thus we introduce a new performance measure: detecting direction of movement success rate, referred to as S_{rate}, which is defined as the ratio between the number of correct directions predicted and the total number of predicted directions.

B. Number of Hidden Nodes and Learning Criteria

In this study, in order to help make comparisons with other research, Shenzhen Composite Index between September 11, 2006 to September 18, 2007 is used. The data was downloaded from Yahoo Finance. It consists of 238 data samples and is divided into 2 groups: 84% or 200 of them are used for training and the rest for testing. There is empirical evidence that generalization from the training set was better when the number of hidden nodes were relatively small. By error-and-trial, it is determined that 6 hidden nodes should be used in contrast to 8 hidden nodes for Wanke and 6 hidden nodes for Shenzhen Composite Index in [13]. In addition, according to Kolmogorov's theorem, BP's hidden nodes should be m=2n+1 where n is the number of input nodes. In this case, hidden nodes should be 9. It can be shown that GA WNN is far more efficient in terms of hidden nodes than BP. Using error-and-trial, the learning rate of 0.001 is chosen.

C. Number of Prediction Steps

Some investors enjoy the excitement of short-term speculation and some value accumulative capital gain in longer term. As a result, it is preferable for the prediction system to predict accurately in flexible term period. How far in the future the constructed system can predict stock index value is tested. The results show that next day prediction yields the lowest MSE, highest fitting accuracy and success rate in direction detection. And MSE does increase when predicting for longer terms compared with single-step prediction. It is possibly because: in order to forecast for longer period, some small movements in shorter period is ignored by the system.

D. Results

Fig.1 shows the plot of the results. In the plot, red curve represents the prediction output and the blue curve represents the actual value. The vertical axis is the stock price and horizontal axis represents business days.

Table I summarizes the comparison of main performance indicators of forecast errors between the optimized system and the HGA system [13]. It can be seen that, GA WNN improves forecast accuracy outstandingly by reducing MAPE to only 12% of the one in HGA WNN. And GA WNN achieves this with much fewer iterations than HGA in [13].

Moreover, as S_{rate} is a new measure proposed in this paper, it is not adopted by the other research. Nevertheless, since the system constructed has predicted the direction of price movement correctly 83.78% of time, it has value for helping investors to make decisions.

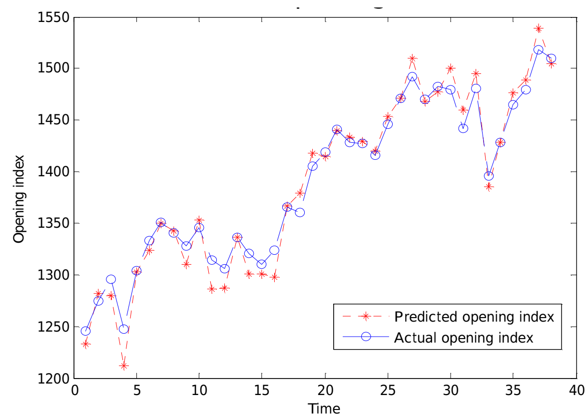

Fig. 1. Prediction of Shenzhen Composite Index

TABLE I. COMPARISON OF DIFFERENT PERFORMANCE INDICATORS

	GA WNN	HGA WNN
Number of iterations	50	500
MAPE(%)	0.79	6.63
Maximum Forecasting Error (%)	2.8	13.84
Minimum Forecasting Error (%)	0.08	0.54

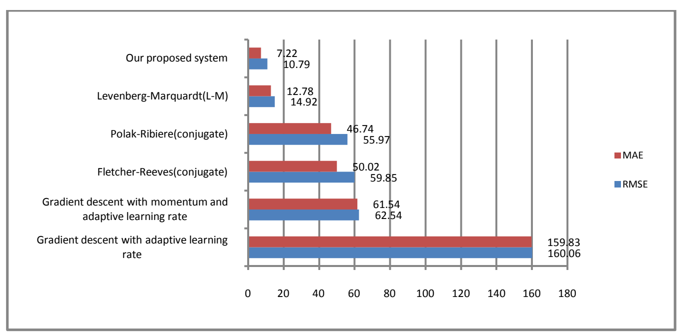

Fig. 2. Performance comparisons of various approaches in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Our proposed approach shows the best results.

The comparison of Wanke (stock code 000002) forecast error indicator between the Optimized System and the HGA system in Zhou and Wei's research [13] is summarized in Table II. It can be seen that, GA WNN improved forecast accuracy significantly by reducing MAPE to 25% of that in HGA WNN. And GA WNN achieved this with much fewer iterations than HGA in [13].

It is to be noted that our model includes not only more information related to historical price, but also the trading volume, which may contribute to the improvement of the performance of the prediction system too.

Furthermore, in order to compare the effect of genetic algorithms integrated with WNN, another comparison is done with Lahmiri's research [17][20]. For this comparison, data of S&P 500 prices from October 2003 to January 2008 on a daily basis is retrieved. The neural networks are trained with 80% of the sample data and tested with the residual 20%. And the system can find the optimal solutions at around 25 epochs, which is very efficient. Figure 2 shows the comparison results with some learning algorithms in [20].

TABLE II. COMPARISON OF DIFFERENT PERFORMANCE INDICATORS FOR WANKE FORECAST ERROR

	GA WNN	HGA WNN
Number of iterations	50	1000
MAPE(%)	1.66	6.59
Maximum Forecasting Error (%)	5.51	14.64
Minimum Forecasting Error (%)	0.001	0.01

After comparisons with different techniques in network training, we can see that GA achieves superior performance than quite a few other algorithms and it has good approximation ability and forecasting capabilities.

VI. CONCLUSION

In this paper, the application of WNN and GA to stock market prediction is studied. We construct the prediction system, discuss the parameters of the system, and test it on Shenzhen Composite Index. We compared our results with [13], [20] and [17]. In our paper, we chose the normal topology of genetic algorithms with the maximum number of generations only 1/10 and 1/20 of [13] and population size is only 2/5 of [13]. Zhou and Wei set objective function purely based on AIC Criterion. In order to make full use of the genetic algorithm, we chose to iterate the optimization process until both weight and error are optimized. In other words, the fitness function is an optimization loop instead of a simple equation. In addition, we used Morlet wavelet function instead of Morlet-Gaussian wavelet function in our system. Results show that, our optimized system demonstrates outstanding performance over other systems and achieves very good prediction accuracy in terms of both value and direction.

Compared to these previous approaches, a smaller number of hidden nodes were included in our system. The forecast results show that our optimized wavelet neural network has significant advantages over other wavelet neural networks, including the HGA wavelet neural network.

Most of the retail investors trade via brokers, who charge commissions. And investors are required to pay stamp tax in almost every country, which is levied as portion of the value of trade. As a result, short-term speculation might not be suitable for investors if the transaction costs are too high. Therefore, adjustment to improve the system longer-terms predictability can be investigated in the future.

REFERENCES

[1] S. Gupta and L. P. Wang, "Stock Forecasting with Feedforward Neural Networks and Gradual Data Sub-Sampling," *Aust. J. Intell. Inf. Process. Syst.*, vol. 11, no. 4, Apr. 2010.

[2] M. Zhu and L. P. Wang, "Intelligent trading using support vector regression and multilayer perceptrons optimized with genetic algorithms," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1–5.

[3] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Netw.*, vol. 3, no. 6, pp. 889–898, Nov. 1992.

[4] J. Zhang, G. G. Walter, Y. Miao, and W. N. W. Lee, "Wavelet neural networks for function learning," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1485–1497, Jun. 1995.

[5] A. Aussem, J. Campbell, and F. Murtagh, "Wavelet-based feature extraction and decomposition strategies for financial forecasting," *J. Comput. Intell. Finance*, vol. 6, no. 2, pp. 5–12, 1998.

[6] L. P. Wang and S. Gupta, "Neural Networks and Wavelet De-Noising for Stock Trading and Prediction," in *Time Series Analysis, Modeling and Applications*, W. Pedrycz and S.-M. Chen, Eds. Springer Berlin Heidelberg, 2013, pp. 229–247.

[7] J. B. Ramsey, "Wavelets in Economics and Finance: Past and Future," *Stud. Nonlinear Dyn. Econom.*, vol. 6, no. 3, Jan. 2002.

[8] A. B. Geva, "ScaleNet-multiscale neural-network architecture for time series prediction," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1471–1482, Nov. 1998.

[9] L. P. Wang and X. J. Fu, *Data Mining with Computational Intelligence*. Berlin/Heidelberg: Springer-Verlag, 2005.

[10] T. Zheng, K. Fataliyev, and L. P. Wang, "Wavelet neural networks for stock trading and prediction," SPIE Defense, Security, and Sensing, 29 April - 3 May 2013, Baltimore, USA, vol.8750, 0A (8750-9).

[11] G. Dong, K. Fataliyev, and L. P. Wang, "One-step and multi-step ahead stock prediction using backpropagation neural networks," in *Communications and Signal Processing (ICICS) 2013 9th International Conference on Information*, 2013, pp. 1–5.

[12] Ai-ping Jiang and Huang Feng-wen, "Methods for optimizing weights of wavelet neural network based on adaptive annealing genetic algorithm," in *16th International Conference on Industrial Engineering and Engineering Management, 2009. IE EM '09*, 2009, pp. 1744–1748.

[13] H. Zhou and Y. Wei, "Stocks market modeling and forecasting based on HGA and wavelet neural networks," in *2010 Sixth International Conference on Natural Computation (ICNC)*, 2010, vol. 2, pp. 620–625.

[14] Q. Zhang, *Wavelet Network□: The Radial Structure and an Efficient Initialization Procedure*. Linköping University, 1992.

[15] K. K. Teo, L. P. Wang, and Z. Lin, "Wavelet Packet Multi-layer Perceptron for Chaotic Time Series Prediction: Effects of Weight Initialization," in *Computational Science - ICCS 2001*, V. N. Alexandrov, J. J. Dongarra, B. A. Juliano, R. S. Renner, and C. J. K. Tan, Eds. Springer Berlin Heidelberg, 2001, pp. 310–317.

[16] N. Groenewold, “Share market efficiency: tests using daily data for Australia and New Zealand,” *Appl. Financ. Econ.*, vol. 7, no. 6, pp. 645–657, 1997.

[17] Salim Lahmiri, “A Comparative Study Of Backpropagation Algorithms In Financial Prediction,” *Int. J. Comput. Sci. Eng. Appl.*, vol. 1, no. 4, pp. 15–21, Aug. 2011.

[18] C. Ornes and J. Sklansky, “A neural network that explains as well as predicts financial market behavior,” in *Computational Intelligence for Financial Engineering (CIFEr), 1997., Proceedings of the IEEE/IAFE 1997*, 1997, pp. 43–49.

[19] D. R. Wilson and T. R. Martinez, “The general inefficiency of batch training for gradient descent learning,” *Neural Netw.*, vol. 16, no. 10, pp. 1429–1451, Dec. 2003.

[20] S. Lahmiri, “WAVELET TRANSFORM, NEURAL NETWORKS AND THE PREDICTION OF S&P PRICE INDEX: A COMPARATIVE STUDY OF BACK PROPAGATION NUMERICAL ALGORITHMS.,” *Bus. Intell. J. 19182325*, vol. 5, no. 2, 2012.