Stock market forecasting research based on Neural Network and Pattern Matching

LIN QianYu
School of Information Science and Technology, Xiamen University
Xiamen Fujian 361005, China
xmulqy@gmail.com

FENG ShaoRong*
School of Information Science and Technology, Xiamen University
Xiamen Fujian 361005, China
shaorong@xmu.edu.cn

***Abstract*—BP Neural Networks is one of the most popular tools in the analysis of stock data. Recent research activities in Pattern Matching indicate that Pattern Matching just simplify the complexity of stock trend prediction and provide a simple but effective way for the stock market prediction. This paper analysis the theory of BP Neural Networks and Pattern Matching, proposes a method for combining these two algorithms to establish a stock market forecasting system based on BP Neural Networks and Pattern Matching. This system overcomes the shortcomings of the local least in the Neural Networks forecasting system's objective function and Pattern Matching System's lack of stock changing probabilities, takes advantage of the unique strength in stock price forecasting of these two algorithms. Finally, test this system by analyzing and forecasting the Titan Oil's stock price. The experimental results show that this hybrid system achieves better forecasting accuracy than either of the models used separately. This hybrid system not only has a quick convergent rate and a precise forecast but also that it is easy to use and has much application value.**

Keywords-Stock; Forecasting; Back Propagation Neural Networks; Pattern Matching; Nonlinear

I. INTRODUCTION

Stock market attracts thousands of investors' hearts from the birth of it. The risk and profit of it has great charm and every investor wants to get some benefits from that, so the stock price forecasting is always a popular field of study in the area of financial data mining. People use various methods to predict market volatility, such as K-line diagram analysis method, Point Data Diagram, Moving Average Convergence Divergence, even coin tossing, fortune telling, and so on.

A large number of studies have shown that time-series Prediction Method based on multilayer Feed-forward Neural Networks is one of the best ways for stock forecasting by now. This is mainly because that Neural Networks has the ability for arbitrary nonlinear function approximation and information processing which other methods do not have. [1]But it also has some disadvantages which are mainly manifested in that ordinary BP algorithm has a slow rate of convergence and is easy to fall into local optimum. Thus, the creation and reliability of the model has been affected. Except that, Neural Networks usually uses last few nodes in the time series to predict the value of the next node, which doesn't make full use of the historical data's rules of changing. And Pattern Matching and Recognition can remedy this deficiency.[2] The prediction using Pattern Matching and Recognition method has a good-fitting because that the forecasting is done based on a long historical data whose change can reflect various influence factors which are included in the prediction results too, such as artificial factor, market discipline, emergency, and so on. But the stock is not only a simple repetition of the past, but also has its own rules for changing as time goes by. So we can combine the Artificial Neural Networks and Pattern Matching by getting the training data of the Neural Networks using Pattern Matching and Recognition system, training the weight values of the Neural Networks by time sequences which come closest to the test sequence. This way can spread the advantage and leave the disadvantage away, finally get a good forecasting result.

II. BP ARTIFICIAL NEURAL NETWORKS ALGORITHM

A. Artificial Neural Networks' overcome

Artificial Neural Networks is a complex network giant system widely interconnected by a large number of simple processing units which is similar to the neuron. It is an artificial construction of network which is able to achieve some kind of function based on human's understanding towards their brain's neural networks. [3]It is a theoretical mathematical model of brain neural networks and a kind of information processing system created by implementation and imitation of the brain neural networks' structure. It is actually a complex network composed by a lot of interconnected processing units which can do some complex logic operations and the implementation of non-linear relationship.

B. BP Algorithm

BP Neural Networks, which can also be called Multilayer Feed-forward Neural Networks, is composed by one input layer, one or more hidden layer and one output layer. It can be used to simulate nonlinear mapping model, solve some real world problems, such as classification, valuation, prediction, and so on. Three-layer Feed-forward Neural Network is a single hidden layer network most commonly used in the

This research is supported by the National Natural Science Foundation of China under Grant No. 50604012. Feng ShaoRong is the corresponding author.

978-0-7695-3997-3/10 $26.00 © 2010 IEEE
DOI 10.1109/ICEE.2010.490

theoretical study and practical application. [4] Figure 1 shows the structure.

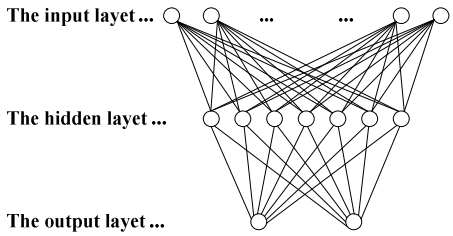

Figure1: the topology of three-layer Neural Network

The learning process of BP Algorithm can roughly be divided into two phrases: the forward propagation of information and the backward propagation of error. The input information from the input layer transmitted to the output layer through layer-by-layer processing of the hidden layer units during the process of forward propagation. Then, the backward propagation phrase is performed if the desired output can not be got in the output layer. Let error signal return along the original connection path by modifying neural weight of each layer to achieve minimal error. [5] A specific description of the algorithm is as follows:

1) forward propagation

a) output of the node in input layer: X_j

b) output of the node in hidden layer:

$$Y_i=f(\sum W_{ij}X_j+\theta_i)$$

In the above formula, W_{ij} is the connection weight between input layer and hidden layer, and θ_i is the neural threshold of hidden layer.

c) output of the node in output layer:

$$O_l=f(\sum T_{li}Y_i+\gamma_l)$$

In the above formula, T_{li} is the connection weight between hidden layer and output layer, andγ_l is the neural threshold of output layer.

2) backward propagation

Modify neural weights along the negative gradient direction of error function to enable the network convergence.

The error contributed by the output unit is:

$$\delta_{li}=[O_l'(t)- O_l(t)]\{O_l(t)[1- O_l(t)]\}$$

The error contributed by the hidden unit is:

$$\delta_{ij}=Y_i(t)[1-Y_i(t)](\sum\delta_l T_{li})$$

Then modify the connection weights between output layer and hidden layer by the following formula:

$$T_{li}(t+1)= T_{li}(t)+\eta\delta_{li}Y_i(t)$$

And modify the connection weights between hidden layer and input layer by the following formula:

$$W_{ij}(t+1) = W_{ij}(t) + \eta\,\delta_{ij}X_j(t)$$

In the above formulas, $O_l'(t)$ is the desired output, $O_l(t)$ is the actual output of the neural networks and η is step size whose value is related to the speed of the learning rate. $T_{li}(t+1)$, $W_{ij}(t+1)$ are both correction value of the weight currently while $T_{li}(t)$ and $W_{ij}(t)$ are both correction values of last learning cycle.

III. PATTERN MATCHING FORECASTING ALGORITHM

Pattern Matching Forecasting Algorithm is just choosing historical time-sequence whose data trend behavior comes closest to that of current sequence and predicting the future price trend based on the price trend of the closest historical time sequence's last node. [6]This algorithm mainly includes three steps: first, time-series selection; second, code the time series and find the closest time-sequence match; third, forecast the future stock price by the closest time-sequence data.

A. time-series selection

We use $X=\{x_1,x_2,\ldots,x_n\}$ as the assumed selected historical time-series. The goal is to predict the future value x_{n+1} based on the values x_{n-k+1}, x_{n-k+2}, …, x_n of historical observations in the time-series. Practice can tell us that the most appropriate value for n is between 250 and 400, a choice between 2 and 6 is best for k. In the following algorithm descriptions, we use 3 as assumed value of k to do the example descriptions. [7]

B. code the time-series and find the closest time-sequence match

Process the original time-series X as following:

$x_i'=1$in case of $x_{i+1}>x_i$; $x_i'=0$ in case of $x_{i+1}\leqslant x_i$ $x_i''=x_{i+1}-x_i$

So, we can get two new time-series: $X'=\{x_1', x_2', \ldots, x_{n-1}'\}$, $X''=\{x_1'', x_2'', \ldots, x_{n-1}''\}$. $X'$is the time-series of trend change in stock price while X'' is the time-series of stock price variation. x_{n-2}', $x_{n-1}'$and x_{n-2}'', x_{n-1}''can be get by corresponding process towards historical observation values x_{n-2}, x_{n-1}, x_n. Then find the matching data segment which comes closest to x_{n-2}', x_{n-1}'in pattern from series X'. There may be several data segments which satisfy this condition. The next step is to choose one best matching time-sequence by comparing the gap between any data segment of these and x_{n-2}, x_{n-1}, x_n.

Computational formula for the gap between data segment x_{n-2}, x_{n-1}, x_n and x_{j-2}, x_{j-1}, x_j is as following:

$$D(X_j, X_n)=| x_{j-2}''-x_{n-2}''|+| x_{j-1}''- x_{n-1}''|$$

Smaller value of $D(X_j, X_n)$ means closer vibrant amplitude and higher degree in pattern matching between two data segments of X_j and X_n; in the other hand, bigger value of $D(X_j, X_n)$ means smaller degree in pattern matching between these two data segments.

C. forecast the future stock price by the closest time-sequence data

Suppose that data segment x_{j-2}, x_{j-1}, x_j is the matching sequence which comes closest to x_{n-2}, x_{n-1}, x_n. Then the value of x_{n+1} can be predicted by the following formula。

$$x_{n+1} = x_n + \frac{|x_{n-2}''| + |x_{n-1}''|}{|x_{j-2}''| + |x_{j-1}''|} \times x_j''$$

IV. APPLICATIONS IN STOCK FORECASTING COMBING NEURAL NETWORKS AND PATTERN MATCHING

A. design of the Neural Networks' structure

The design of networks' structure includes making decision of the number of Neural Networks, the number of neurons in each layer and the topology design of the network.

1) the number of Neural Networks： The practice has shown that, the Neural Networks with structure of 4-layer is more accessible to fall into local minimum than the one with structure of 3-layer and the time for training the weight of network is also increased，so we use a 3-layer network with only one hidden layer in the application of stock forecasting. [8]

2) the number of input layer's nodes：daily stock turnover form a time series whose trend starts in waveform and is cyclical. We suppose that daily stock price has some kind of function relationship with the stock prices of the m days before. This m is called as analysis cycle, which is also the Neural Networks' input layer's node number. If the analysis cycle is selected appropriately has a direct impact on the forecasting result. [9] Usually we take 5 days, 10 days, 20days or 60 days as technical analysis cycle. By experiments, we finally choose 5 days.

3) the number of output layer's nodes：we set 1 as the number of neural networks' output nodes taking into account such factors as time, number of assignments and convenience of choosing，which means that the stock prices of previous m days' stock prices are just used to predict the price of next day. 。

4) the number of hidden layer's nodes： there are no theoretical guidance about how to determine the number of hidden nodes. Excessive network nodes increase the training time of network, weaken the network's generalization capacity and decline the predictive power. But the model can not be fully created if the networks nodes are not enough. [10] The number of hidden nodes is usually set between input layer's nodes and output layer's nodes.

5) topology design of the network： for BP Neural Networks, the activation function of nodes should choose those which is continuously differentiable. The Sigmoid function $f(x)=1/(1+e^{-x})$ usually is chosen and the graph of this function is shown in figure 2.

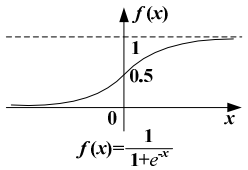

Figure2: transfer function Sigmoid

B. preparation of training sample sets

Pattern matching in the context of time-series forecasting refers to the process of matching current state of the time series with its past states. Consider the tagged time series $X=\{x_1,x_2,\ldots,x_n\}$. Suppose that we are at time n(x_n) trying to predict x_{n+1}. A pattern of size k $\rho=(x_{n-k+1}, x_{n-k+2}, \ldots, x_n)$ is first formulated from the last k tag values in the series. The size k of the structural primitive pattern used for matching has a direct effect on the prediction accuracy. Thus the pattern size k must be optimized for obtaining the best results. For this k is increased in every trial by one unit till it reaches a predefined maximum allowed for the experiment and the error measures are noted; the value of k that gives the least error is finally selected[11]. By experiments, we finally choose 5 as k.

The magnitude and direction of prediction depend on the match found. The success in correctly predicting series depends directly on the pattern matching algorithm. The aim of a pattern matching algorithm is to find the closest match of p in the historical data and use this for predicting x_{n+1}. But in this hybrid, we just use Pattern Matching and Recognition Algorithm to select ten sets of sequence which come closest in pattern matching to the time series $x_{n-4}, x_{n-3}, \ldots, x_n$ in historical time series $X=\{x_1, x_2, \ldots, x_n\}$. Then use these ten sequences as input vectors of the Neural Networks' training samples and the next value of each sequence in X as the teacher data of the corresponding sample. So we can get a training sample set whose size is ten.

C. stock forecasting using Neural Networks

After the structure of Neural Networks has been built, we can train the weights of the network using training sample sets chosen by Pattern Matching and Recognition Algorithm. The step size between input layer and hidden layer, hidden layer and output layer is both 0.7 and the training is over when the global error is less than 0.005[12].

The actual forecasting of the stock can be done after the end of training. Let time series $x_{n-4}, x_{n-3}, \ldots, x_n$ as the input of the trained neural networks, the forecasting value x_{n+1} can be got as the output.

V. ANALYSIS OF EXPERIMENTAL TEST RESULTS

First, analyze and forecast comprehensive index of Titan Oil's stock using three-layer Neural Networks. Use 165 consecutive trading days between August 29th, 2006 and May 27th, 2007 as training samples，100 consecutive trading days begin with May 28th, 2007 as test samples to de the predictive testing. Figure 3 shows the fitting curve of learning and forecasting. In the figure, solid line is the actual stock curve while dashed line is the curve of stock's forecasting value using Neural Networks.

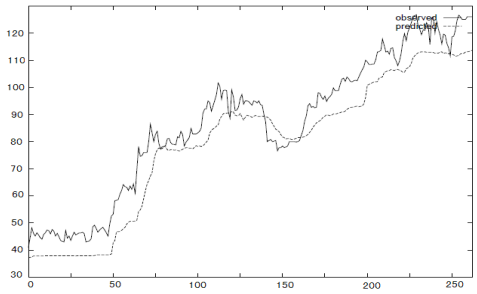

Figure3: the result of the Neural Networks Forecasting System

Figure 4 is the fitting curve of 100 trading days begin with May 28th, 2007 predicted using Pattern Matching and Recognition System. In this figure, solid line is the actual stock curve while dashed line is the curve of stock's forecasting value using Pattern Matching and Recognition system.

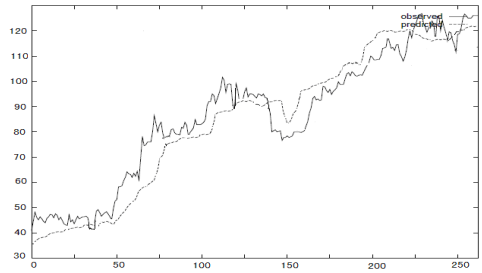

Figure4: the result of the Pattern Matching and Recognition System

Figure 5 is the fitting curve of 100 trading days begin with May 28th, 2007 predicted using hybrid forecasting system combining BP Neural Networks and Pattern Matching. In this figure, solid line is the actual stock curve while dashed line is the curve of stock's forecasting value using hybrid system.

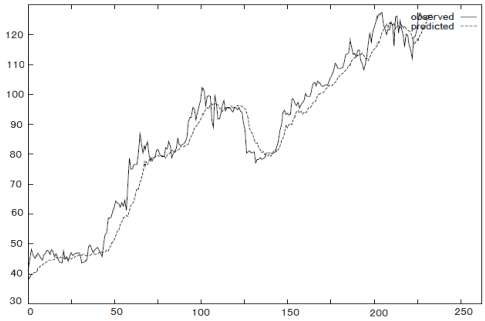

Figure5: the result of the hybrid forecasting system

These three methods predict future stock's prices both by the stock's historical data and both of them get a good forecasting result. But we can get a conclusion that the result got by hybrid forecasting system combining BP Neural Networks and Pattern Matching is more accurate than any other method by the three predictive maps.

Analysis of the reason shows that Neural Networks construct its model parameters based on the differentiation of the stock market environment. [13] The model is built by the training samples we provided, so the environment which the model established on this basis can be accustomed to is the stock market environment where training samples are. The samples chosen by Pattern Matching Algorithm have a high level in pattern matching with the test sample and their data internal connection is more similar than the others. [14] Train the Neural Networks using the selected samples, then get a model which comes closer to the stock market environment where the test samples are and a higher forecasting accuracy.

VI. CONCLUSION

This paper proposes a method for combining the BP Neural Networks with Pattern Matching to do stock forecasting. This method has a precise forecast and much application value. [15] Compared to the combining Genetic Algorithms with Neural Networks, this is a relatively new method. There are so many things worth our exploration and research although someone has taken a step ahead, such as Sameer Singh, Jonathan Fieldsend and so on. For example: the hybrid forecasting system has a longer running time than the original BP Neural Networks Prediction Model; add some units which can reflect external factors affecting the stock on the input layer. In a word, we need to take more in-depth studies in order to get a better accuracy in the stock prediction and research.

REFERENCES

[1] Contreras J, Rosario E, Nogales F J, COnejo A J. ARIMA Models to Predict Next-Day Electricity Prices[J]. IEEE Transactions on Power Systems, 2003.

[2] William Leigh, Ross Hightower, Naval Modani. Forecasting the New York Stock exchange composite index with past price and invest rate on condition of volume spike. Expert System with Applications, 2005.

[3] E. Gately, Neural Networks for Finacial Forecasting, John Wiley, 1996.

[4] Olson Dennis, Mossman Charles. Neural network forecasts of Canadian stock returns using accounting ratios. International Journal of Forecasting, 2003.

[5] Linlokken, G., "Syntactical pattern recognition in financial time-series," proceedings of the Norwegian Image Processing and Pattern Recognition Society Conference, Oslo, 1996.

[6] Rui Jiang, and K. Y. Szeto. Extraction of Investment Strategies based on Moving Averages: A Genetic Algorithm Approach. 2003.

[7] Yiwen Yang, Chaojn Yang. Short term forecasting of stock market based on R/S analysis and fuzzy neural networks. Systems Man and Cybernetics. 2003.

[8] Zhengxue Li, Wei Wu and Yuelong Tian, Convergence of an online gradient method for feedforward neural networks with stochastic inputs, Journal of Computational and Applied Mathematics. 2004.

[9] Zhang G P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model[J]. Neurocomputing. 2003.

[10] Singh, S. "A Long Memory Pattern Modeling and Recognition System for Financial Forecasting". Pattern Analysis and Applications, vol.2, issue 3, pp. 264-273, 1999.

[11] Chi Kin Chow, Tong Lee. Construction of multi-layer feed forward binary neural network by a genetic algorithm. Neural Networks, 2002, IJCNN'02. Proceedings of the 2002 International Joint Conference, Honolulu, HI, USA, 2002.

[12] Yam, J.Y.F., Chow, T.W.S. Feed forward networks training speed enhancement by optimal initialization of the synaptic coefficients. Neural Networks, IEEE Transactions, 2001.

[13] Alan M. Safer. The application of neural networks to predict abnormal stock returns using insider trading data. Applied Stochastic Models in Business and Industry. 2002.

[14] A.A. Freitas. Data Mining and Knowledge Discovery with Evolutionary Algorithms, Springer-Verlag, 2002.

[15] S. Tokianga and J. Lu, "On modeling of time series based on the Genetic Programming and its application to clustering(in Japanese)", Technical Report of the IEICE, SIP2004-61, pp.71-76, 2004.

.