

A Novel Prediction Method for Stock Index Applying Grey Theory and Neural Networks

Shen Yan

College of Science, Tianjin University of Technology, Tianjin, P.R. China, xyyacad@tom.com

Abstract This paper presents a better prediction model by the integration of neural network technique and grey theory for the stock index. In this paper, the grey theory applied include grey forecast model and grey relationship analysis. A GM(1, 1) grey forecast model was applied to predict the next day's stock index. Grey relationship analysis was used to filter the most important quantitative technical indices. To examine the influence of dimension of the model to prediction accuracy, seven different kinds of dimension 5, 6, 8, 10, 12, 14, and 15 were tested. The generated data were then regarded as new technical indices in grey relationship analysis and prediction of neural network. Finally, a Recurrent Neural Network was developed to train and predict the price trend of stock index. The conclusion shows our models can provide good prediction for this problem.

Keywords Stock Index; GM(1, 1); Grey Relationship Analysis; Recurrent Neural Network

1 Introduction

In past, the pursuits for predicting stock index mostly used time series like Auto Regression Integrated Moving Average (ARIMA) or the neural network models like Back Propagation Neural Network (BPN)[1, 2]. Generally, these models have two common features in the selection of input factors: (1) the emphasis of the quantitative indices. (2) the requirement of large amount of input data. Nevertheless, these quantitative indices are not really suitable for the financial market. Furthermore, the requirement of large amount of input data will lower down the convergent rate of neural network model. Such that, this research attempts to develop a better prediction model by the integration of neural network technique and grey theory for the Shanghai composite index.

In the applications of neural network techniques to the prediction of stock index, the input variables always choose the typical indices or the elected indices by the previous researches [3, 4]. After all, some people may doubt the representative of these variables in the prediction. Therefore, this research attempted to apply the gray relationship analysis in the preliminary selection of the quantifiable indices, i.e. the technical indices popularly used in the analysis of stock. Back-Propagation Neural Network (BPN) is the most popular neural network paradigm used in prediction. Researches in the prediction of stock index using BPN have been published considerable results. However, the conventional feedforward BPN may not perform so well in prediction of a time series data, because it takes no account of the influence of the anterior data during the prediction. That means the way of BPN learning is principally based on pattern collation to map the relationship

between input and output without taking account into the influence of the previous outputs. Recurrent Neural Network (RNN) has the capability of storing previous information for future prediction to overcome this drawback in the conventional BPN.

2 Prediction Method

In the development of the proposed hybrid prediction system for Shanghai composite index which principally integrates gray theory with artificial neural network technique, there are three main aspects included: gray forecasting of the Shanghai composite index future price, selection of prediction indices, and development of the artificial neural network. In the gray forecasting, the linear dynamic forecasting model GM(1, 1)^[5-9] with single variable is applied for forecasting the Shanghai composite index future price. The selection of technical analysis indices as the inputs of the artificial neural network for prediction is based on the evaluation of the relationship between these indices and price of Shanghai composite index. In this aspect, gray relationship analysis method is adopted for the selection of key indices, which have higher gray relationship values. Theoretically, the higher gray relationship value is, the greater similarity between the fluctuation curve of the Shanghai composite index and the moving curve of the evaluated technical index is. Finally, a recurrent neural network is created and tested for prediction during the development stage.

In designing the structure of the neural network, the technical indices selected by gray relationship analysis are regarded as the input variables; up and down of the Shanghai composite index future price are defined as the two output variables.

2.1 Gray Forecasting

In this research, the purpose of using gray forecasting is to predict the next day's price for Shanghai composite index. During the forecasting process, gray dynamic forecasting model should be operated in accordance with the principle of keeping the same dimension of data series. That means a new data is attached on the rear of the original data series and the first data in the original data series should be removed before the next forecasting. In this research, we attempt to explore the feasibility of gray forecasting model as a new type of technical index using in the prediction of the neural network. This new technical index is called "Gray Index". In addition, various degrees of dimension in gray forecasting model are chosen to examine the influence of dimension to the prediction of Shanghai composite index, various degrees of dimension should be defined.

The following example is designed to explain the manipulation of gray forecasting model under the principle mentioned above. Assume that Table 1 contains a set of prices of Shanghai composite index from January 7, 2008 to January 16, 2008. Now, let the GM(1, 1) gray forecasting model applied have 8 degrees of dimension.

Table 1: A set of historical data about data of Shanghai composite index.

Date	1/7	1/8	1/9	1/10	1/11	1/14	1/15	1/16
Index	5393.34	5386.53	5435.81	5456.54	5484.68	5497.90	5443.79	5290.61

Let the initial data series

$$x_0(k) = (x_0(1), x_0(2), \dots, x_0(8))$$

$= (5393.34, 5386.53, 5435.81, 5456.54, 5484.68, 5497.90, 5443.79, 5290.61)$

Step 1: Use the initial data series to perform the Accumulated Generating Operation (AGO).

$$x_1(k) = (x_1(1), x_1(2), \dots, x_1(8)) \\ = (5393.34, 10779.87, 16215.68, 21672.22, 27156.69, 32654.59, 38098.38, 43388.99)$$

Step2: Calculate

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix} = (B^T B)^{-1} B^T y_n, \\ B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix}$$

$$y_n = [x^{(0)}(2) \quad x^{(0)}(3) \quad \dots \quad x^{(0)}(n)]^T$$

$$\text{Thus, } \hat{a} = \begin{bmatrix} -0.0039956 \\ 5367.30481 \end{bmatrix} = \begin{bmatrix} a \\ u \end{bmatrix}$$

Use \hat{a} to produce the derivative equation:

$$\frac{dx_1}{dt} - 0.0039956x_1 = 5367.30481$$

Then, time function of the derivative equation is expressed as

$$\hat{x}_1(t+1) = (x_0(1) - \frac{u}{a})e^{-at} + \frac{u}{a}$$

Let $t = 8$, then $x_1(9) = 48558.31$ and $x_0(9) = x_1(9) - x_1(8) = 5169.32$

According to the procedure of gray forecasting, the closed price of Shanghai composite index January 17, 2008 is predicted as 5169.32

2.2 Gray Relationship Analysis

Gray relationship analysis is developed to filter the most important quantitative technical indices for prediction. The general indices collected embrace Moving Average (MA), Relative Strength Index (RSI), Bias, KD.J Stochastic Index (SI), Psychological Line (PSY), Moving Average Convergence and Divergence (MACD), Williams' Oscillator (WMS%R), ABI, BTI, etc. The parameters used for each index can be defined differently. The detailed procedure of gray relationship analysis can be described as the following steps:

Step 1: Define the reference data series $Y_0 = (y_0(1), y_0(2), \dots, y_0(n))$, where n is the number of elements in the series. For example, assume the eight days' prices of Shanghai composite index from January 7, 2008 to January 16, 2008 are the elements of the reference data series. It can be expressed as $Y_0 = (y_0(1), y_0(2), \dots, y_0(8))$ shown in Table 2.

Step2: Let the number of indices involved is m . Establish all of the comparative series, $Y_i = y_i(k)$, where $i=1, 2, \dots, m$; $k=1, 2, \dots, n$, for the technical indices selected. For instance, the values shown in Table 2 are the raw data, collected from January 7, 2008 to January 16, 2008, of some technical indices defined as below.

Step3: Normalize the reference data series and all of the comparative data series as the following formula:

$$Z_i = (z_i(k)) = \frac{y_i(k)}{(\sum_{k=1}^n y_i(k))/n},$$

Table 2: The raw data of some technical indices about Shanghai composite index from January 7, 2008 to January 16, 2008.

	1/7	1/8	1/9	1/10	1/11	1/14	1/15	1/16	Avg
Y_0	5393.34	5386.53	5435.81	5456.54	5484.68	5497.90	5443.79	5290.61	5423.65
Y_1	5268.86	5297.34	5317.49	5343.03	5368.16	5387.06	5405.28	5407.06	5349.29
Y_2	38.79	25.29	24.74	20.11	19.11	46.46	43.19	39.73	32.18
Y_3	-2.35	-5.37	-3.92	-4.54	-3.68	1.47	1.33	0.45	-1.62
Y_4	35.62	23.84	20.04	18.68	16.75	26.47	32.34	36.88	26.33
Y_5	594.51	133.64	462.92	472.09	310.34	129.17	-1410.3	-3397.8	-338.18
Y_6	83.33	75.00	75.00	75.00	83.33	75.00	66.67	58.33	64.58
Y_7	2.57	24.92	13.78	8.39	5.16	7.75	24.62	24.48	13.96
Y_8	5182.42	5187.34	5193.06	5199.59	5206.81	5214.54	5222.28	5229.26	5204.44

where $k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m$

Step4: Calculate all of the difference data series, $|Y_0(k) - Y_i(k)|$ where $k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m$, which are the absolute values of deducting the reference data series from the comparative series. Find out the maximum value, Δ_{\min} among the maximum group and the minimum value, Δ_{\max} , among the minimum group. Δ_{\min} and Δ_{\max} can be obtained by the following formula:

$$\Delta_{\min} = \min_i \min_k \Delta O_{0,i}(k), \Delta_{\max} = \max_i \max_k \Delta O_{0,i}(k)$$

where

$$\Delta O_{0,i}(k) = |z_0(k) - z_i(k)|, k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m$$

Step 5: Calculate relationship coefficient for each element in all of the comparative series using the formula below

$$L_{0,i}(k) = \frac{\Delta_{\min} + \rho * \Delta_{\max}}{\Delta O_{0,i}(k) + \rho * \Delta_{\max}},$$

where $k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m$.

In the formula, the coefficient is defined to be 0.5 in this research.

Step6: Calculate degrees of relationship between the reference series and the comparative series using the formula below

$$r_{o,i} = \frac{1}{n} \sum_{k=1}^n L_{0,i}(k),$$

where $k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m$.

The results can be obtained as Table 3

Table 3: Degrees of relationship between the reference series and the comparative series.

$r_{0,1}$	$r_{0,2}$	$r_{0,3}$	$r_{0,4}$	$r_{0,5}$	$r_{0,6}$	$r_{0,7}$	$r_{0,8}$
0.958	0.783	0.431	0.502	0.369	0.328	0.561	0.843

Step7: Order the technical indices according to their respective degrees of relationship from the biggest to the smallest. For this example, their order is

$$r_{0,1} > r_{0,8} > r_{0,2} > r_{0,7} > r_{0,4} > r_{0,3} > r_{0,5} > r_{0,6}$$

Based on the order, every time a different number of the indices located near the front end of the order are selected as the input of the neural network for training and testing. According to the performances of the networks in convergence and prediction accuracy, the influence of the number of selected indices can be exploited and the best network with the highest prediction accuracy can be determined.

2.3 Design of Recurrent Neural Network

As mentioned in Section 1, RNNs can map the relationship involving space and time. Therefore, a RNN can be the best choice to be developed for training and predicting the price movement of Shanghai composite index. In the proposed network structure, price movement of Shanghai composite index is defined as the output and the values gained from previous processing by gray relationship analysis are used as the input. Besides, this network has only one hidden layer. The output of the network consists of two neurons. One of the two neurons represents the upward trend of the price of Shanghai composite index; the other represents the downward trend of the price. If the output values of the two neurons are both near 0.5, the probability of upward and downward could be equal. During the training stage, the number of input neurons, the number of hidden neurons, and learning parameters, e.g. momentum and learning rate, in the network are tested to find out the optimal solution. The number of input neurons depends on how many technical indices are involved. The feasibility of the neural network models is examined by their prediction accuracy.

3 Case Study

Up to date, the number of the published technical indices has been about several decades. Generally, these indices were created by using Statistics to analyze the fluctuation of historical data of monetary merchandise such as stock price, exchange rate and price of futures. With the change of time, every index can produce a series of concrete values, which may form a signal of opportunity to suggest the investor for buying or selling. In this research, the quantifiable indices applied in gray relationship analysis and artificial neural network are the technical indices listed in Table 4. The total number of the indices is 29. In addition, seven forecasted values by the GM(1, 1) forecasting model are also considered as new types of technical indices which will be utilized in the gray relationship analysis and the neural network. Actually, these seven values are generated by the GM(1, 1) model with seven different kinds of degrees of dimension 5,6,8, 10,12, 14, and 15.

The number of historical records of Shanghai composite index, collected from March 2006 to February 2007, was totally 259 during the twelve months. In the process of gray relationship analysis, the degree of relationship between price of Shanghai composite index and each technical index in a single month was first calculated. Based on their degrees of relationship, these technical indices were arranged in order and were given an ordinal number to each index. That means the index with the highest degree of relationship was assigned the number of 1. Finally, compute the average of the ordinal numbers obtained in the twelve months for each index. According to their averages, these indices were arranged again and listed as Table 5. From the result shown in the table, there are five gray technical indices located in the top ten indices. This means the application of the gray technical indices in the prediction of monetary merchandise like stockindex future may be a good idea in this domain.

According to the order obtained in Table 5, the quantifiable technical indices were selected to examine their prediction capability. In the experiments, five different groups of the indices, which respectively contain 10, 15, 20, 25, and 29 indices, were organized to test their feasibility for prediction. In this test, three hundred records from March

Table 4: The technical indices applied for the gray relationship analysis.

Index	Parameter	Index	Parameter
MA	5	PSY	12
	10		13
	20	ARBR	26
	30	DI	1
W%R	10	MACD	EMA
ROC	12		DIF
RSI	6	Gray Index	12E-26E
	12		5
	24		6
BIAS	6		8
	12		10
	24		12
BTI	10		14
	6		15

Table 5: The order of 29 technical indices generated the gray relationship analysis using the historical data of Shanghai composite index.

Order	1	2	3	4	5	6	7	8
Index	DI	5GRAY	6GRAY	8GRAY	9EMA	12EMA	10GRAY	26EMA
Order	9	10	11	12	13	14	15	16
Index	12GRAY	5MA	12ROC	30MA	14GRAY	15GRAY	20MA	10MA
Order	17	18	19	20	21	22	23	24
Index	26ARBR	24RSI	12RSI	10BIT	6RSI	6BIT	12PSY	DIF
Order	25	26	27	28	29			
Index	13PSY	10W%R	24BIAS	12BIAS	6BIAS			

11, 2006 to May 10, 2007 were used. The parameters for the RNN were defined as the following. Learning rate and momentum were both given a value of 0.8. The number of neurons in the hidden layer was determined by taking the roundness and the numbers near the roundness of root mean square of the number of the input neurons multiplied by the number of the output neurons. The training time was determined to be 1000 epochs.

In this research, the programming shell Borland Delphi 6.0 was adopted for the development of the RNN model. The results of the experiments including average of mean square of error (MSE) and root mean square of error (RMS) are presented in Table 6. From the results shown in Table 6, it has been proven that the application of gray relationship analysis to screen the technical indices instead of pouring all of the indices into the prediction of artificial neural network can improve the prediction accuracy.

As described above, the input indices selected as the input of the RNN consist of a number of Gray Indices with higher degrees of relationship. Therefore, it is necessary to examine the feasibility of applying the Gray Indices in the prediction of the RNN. Thus, two different groups of technical indices, Group 1 containing the Gray Indices

Table 6: The training results of the different network structures with a various number of technical indices using the historical data of Shanghai composite index.

Group	1	2	3	4	5
No.of indices	10	15	20	25	29
Network structurwe	10-7-2	15-11-2	20-14-2	25-17-2	29-19-2
MSE	0.1159	0.1267	0.1163	0.1207	0.1845
RMS	0.0308	0.0316	0.0412	0.0261	0.0253

and Group 2 without containing any of the Gray Indices, are organized and compared in the experiments. From the results shown in Table 7, it was found that the RNN with Gray Indices had better performance in prediction than the RNN without Gray Indices. Therefore, the assumption that gray forecasting model can be adopted as a new type of technical index in the prediction of artificial neural network for monetary merchandise had been proven to be true.

Table 7: Comparison of the RNN with Gray Indices and the RNN without Gray Indices.

Group no.	1	2
Network structurwe	10-8-2	10-8-2
MSE	0.0537	0.0874
RMS	0.0219	0.0357

4 Conclusion

In this paper, we present a better prediction model by the integration of neural network technique and grey theory for the stock index. First, A GM(1, 1) grey forecast model was applied to predict the next day's stock index and Grey relationship analysis was used to filter the most important quantitative technical indices. Further, seven different kinds of dimension 5, 6, 8, 10, 12, 14, and 15 were tested in order to examine the influence of dimension of the model to prediction accuracy. The generated data were then regarded as new technical indices in grey relationship analysis and prediction of neural network. Finally, a Recurrent Neural Network was developed to train and predict the price trend of stock index. The conclusion shows our models can provide good prediction for the stock index.

References

- [1] Jun Gao. Theory and Simulation of Artificial Neural Network. Beijing: China Machine Press, 2003.
- [2] V. Kasparian, C. Batur, H. Zhang, et al. Davidon least squares-based learning algorithm for training feedforward neural network. *Neural Networks* 1994; 7(4): 661-670.
- [3] N. Baba, and M. Kozaki. An Intelligent Forecasting System of Stock Price Using Neural Networks. *IJCNN*, 1992, 371-377.
- [4] A. Zapraniš, G. Francis. Stock performance modeling using neural networks: a comparative study with regression models. *Neural Networks*, 1994; 7(2): 375-388.

-
- [5] A. Antonios, H. Phil. Futures Trading, and Spot Price Volatility: Evidence for FTSE-100 Stock Index Futures Contract Using GARCH. *Journal of Banking & Finance*. 1995, 19:117-129.
 - [6] Qin Zhao, Jun Gao, Tianfu Wu, et al. The grey theory and the preliminary probe into information acquisition technology. In: *Proceedings of the International Conference on Information Acquisition*, 2004: 402-404.
 - [7] S.J. Huang, C.L. Huang. Control of an inverted pendulum using grey prediction model. *IEEE Transaction on Industrial Application*, 2000, 36(2): 452-458.
 - [8] Deng Julong. *The Course on Grey System Theory*. HUST Press, Wuhan, China, 1990.
 - [9] Julong Deng. *Basic method of grey system*. Wuhan: Huazhong College of Technology Press, 1988.