# The Application of SVM and GA-BP Algorithms in Stock Market Prediction

HONGXIN XUE
Departament of Mathematics
North University
Taiyuan Shanxi
CHINA
18334792279@163.com

YANPING BAI
Departament of Mathematics
North University
Taiyuan Shanxi
CHINA
baiyp666@163.com

*Abstract:* Neural network has been popular in time series prediction in financial areas, because of their advantages in handling nonlinear systems. This paper hybridizes genetic algorithm and artificial neural network method (GA-BP), and hybridizes principal component analysis and support vector machine (PCA-SVM) to predict the next opening price in stock markets. Principal component analysis method is applied to extract contribution rate to meet 95% of the principal component as the input variables with FAW Car and Minmetals Rare Earth to be modeled and predicted, and genetic algorithm is employed to determine the initial weight and threshold of the BP neural network. The experiment results demonstrate that the combination methods (PCA-SVM and GA-BP) perform better, and the GA-BP method can get higher prediction accuracy than other three prediction methods.

*Key–Words:* Stock prediction, Principal components analysis, Support vector machine, Artificial neural network, Genetic algorithm.

## 1 Introduction

The financial forecasting or stock market prediction is one of the hottest fields of research lately due to its commercial applications owing to the high stakes and the kinds of attractive benefits that it has to offer [1], stock market is a significant and difficult problem in the financial analysis field. Ever since the birth of stock market, people start to focus on and analyze the stock market, trying to analyze the development trend of stock market. The trend of stock price presents high nonlinearity, and the transaction price and turnover include substantial internal rules and features determining the changes of stock price, these rules are implied in the substantial historical data. Through the study of historical transaction data, the artificial neural network may find out and depict the rules and features between the parameters from the complicated data, as well as the fitting of function.

People are interested in predicting the stock trend, index or price. Therefore, many different methods and techniques have been presented, the artificial neural network (ANN) and support vector machine (SVM) methods has been frequently used as a typical model [2,3]. Wen Fenghua, Xiao Jihong, He Zhifang, Gong Xu used the singular spectrum analysis (SSA) and support vector machine (SVM) to make price predictions [4]. Akhter Mohiuddin Rather, Arun Agarwal,V.N. Sastry proposed a robust and novel hybrid model for prediction of stock returns. The proposed model is constituted of two linear models: autoregressive moving average model, exponential smoothing model and non-linear model: recurrent neural network [5]. Wensheng Dai, Jui-Yu Wu, Chi-Jie Lu used the nonlinear independent component analysis (NLICA) and neural networks (BPN) for the prediction of Asian stock market indexes [6]. Yang et al. proposed an early warning system of commercial bank risks using the ANN model [7].

However, all the networks have their own limitations, which may influence their prediction performance when applied in stock market such as there are numerous factors impacting the stock prediction, each factor has distinct impact on the prediction precision, and some are even unnecessary. Consequently, the input variables of the prediction model shall be selected. The initial weight and threshold of BP neural network is selected at random [8], and if the parameters are not selected properly, it may slow down the network convergence rate and fall into the local optimum. In this paper, the principal components are extracted from the input variables with principal component analysis method, and the opening price of the stock is predicted with SVM and GA-BP method. BP neural network has self-adaptation learning ability and strong nonlinear simulation ability, while the genetic algorithm has no self-adaptation ability, but it can conduct convergence for obtaining the globally optimal solution, with perfect robustness. Consequently, in this paper, the neural network and genetic algorithm is combined, which may not only give full play to the general mapping ability of neural network, but also endow rapid convergence, global optimization ability and strong learning ability to neural network. [9].

## 2 Support Vector Machine

Support vector machine (SVM) is a machine learning method, which is based on statistical learning theory created by Vapnik [10,11]. SVM adopted the structural risk minimization criterion, minimizing sample point error and narrowing model generalization ability, and it is a convex quadratic optimization problem so that the local optimal solution is certainly the global optimal one. Its main idea is to find optimal hyper plane as a decision surface, maximizing the isolation edge between positive and negative examples [12,13]. Give a training data set:

$$T = \{(x_1, y_1), ..., (x_l, y_l) \in (X \times Y)\}^l \quad (1)$$

Where $x_i$ is the feature vector, $x_i \in X = R^n$, $y_i \in Y = \{1, -1\}$,$(i = 1, 2, \cdots, l)$;
To solving the following optimization problems:

$$min\frac{1}{2} + C\sum_{i=1}^{l} \zeta^2 s.t. y_i \cdot (\langle \omega, x_i\rangle + b) \geq 1 - \zeta_i,$$
$$(i = 1, 2, \cdots, l). \quad (2)$$

Where $C$ is the penalty parameter, $\xi_i$ is the slack variable. By introducing Lagrange multiplier $\lambda_i$, the optimization problem can be reformulated as Eq. (3)

$$\begin{aligned} L(\omega, b, \zeta, \lambda) = \tfrac{1}{2}\langle \omega, \omega\rangle + \tfrac{C}{2}\textstyle\sum_{i=1}^{l} \zeta_i^2 \\ - \textstyle\sum_{i=1}^{l} \lambda_i[y_i(\langle \omega, x_i\rangle + b) - 1 + \zeta_i] \end{aligned} \quad (3)$$

In order to get the minimum solution, partially differentiate $\omega$, $b$, $\zeta$ and let them equal to zero:

$$\begin{aligned} &\tfrac{\partial L}{\partial \omega}(\omega, b, \zeta, \lambda) = \omega - \textstyle\sum_{i=1}^{l} y_i\lambda_i x_i = 0 \\ &\tfrac{\partial L}{\partial b}(\omega, b, \zeta, \lambda) = \textstyle\sum_{i=1}^{l} y_i\lambda_i = 0 \\ &\tfrac{\partial L}{\partial \zeta}(\omega, b, \zeta, \lambda) = C\zeta - \lambda = 0 \end{aligned} \quad (4)$$

The feature space was defined by Kernel function, the original problem is transformed into solving the optimization problem:

$$\begin{aligned} &max\ \ W(\lambda) = \textstyle\sum_{i=1}^{l} \lambda_i - \tfrac{1}{2}\sum_{i,j=1}^{l} y_i y_j \lambda_i \lambda_j K(x_i, x_j) \\ &s.t. \textstyle\sum_{i=1}^{l} y_i\lambda_i = 0, (\lambda_i \geq 0, i = 1, 2, \cdots, l) \\ &0 \leq \lambda_i \leq C \end{aligned} \quad (5)$$

Where $K(x_i, x_j)$ is the kernel functions, it can have different forms, such as linear kernel function, polynomial kernel function, radial basis function (RBF). RBF is used in this paper, RBF is defined as:

$$K(x_i, x_j) = exp(\frac{-|x_i - x_j|^2}{\sigma^2}) \quad (6)$$

Where $x_i$ and $x_j$ is the sample observations value,$\sigma^2$ is the kernel parameter.

In SVM algorithm, in order to improved prediction accuracy, so a grid search method using k-cross validation is used to find the best width parameter $\sigma$ and penalty parameter $C$ [14,15]. It can effectively avoid the happening of over-fitting and learning state, eventually find ideal prediction accuracy for test set.

## 3 Principal Components Analysis

Principal component analysis (PCA) was first introduced by Pearson in 1901, and later was developed by Hotelling in 1933 [16,17]. Since the number of variables are far too numerous, and there is a certain correlation between each other, making the observed data overlaping for reflected information in a certain degree. PCA is one of several multiple indicators into the multivariate statistical methods of a few principal components analysis, these principal components reflect most of the information of the original variables, usually expressed a linear combination of the original variables, in order to make the information contained in these principal component not overlapping each other , it need no relation between principal components.

The purpose of PCA is to try to reassemble the original variables have certain correlation into a new set of global variables independent of each other to replace the original variables [18,19]. The most common approach is to denote the first comprehensive variable selected as $F_1$, naturally want it as much as possible to reflect the information of the original variables, where "information" is measured by the variance. Therefore $F_1$ should be the largest variance, so called $F_1$ the first principal component.If the first principal component extracted from original data information is not enough, the second principal component $F_2$

should also be considered. In order to effectively reflect the original information, $F_1$ existing information will no longer need to appear in $F_2$,using mathematical language expression is required $Cov(F_1, F_2) = 0$,called $F_2$ is the second principal component, and so on.

## 4 BP neural network

The back propagation network is a multi-layer network that can generalize the W-H learning rules, and conduct weight training for the nonlinear differentiable function [20]. BP algorithm consists of two parts, including the forward transfer of information and back propagation of error. In the forward propagation process, the input information may be transmitted from the input layer to hidden layer to the output layer, and the state in each layer of nerve cell would only impact the state of the next layer of nerve cell. If the output layer fails to obtain the expected output, the error variation of the output layer shall be calculated, and then the back propagation shall be conducted, for transferring the error signal along the original connection channel back for modifying the weight of nerve cell till it reaches the expected objective [21].

The training process of BP neural network mainly includes the following steps:

Step 1: Network initialization. Determine the nodes $n$ in the network input layer according to the input and output, nodes of hidden layer $l1$, nodes of output layer $l2$, initialize the connection weight between the input layer and nerve cell in hidden layer $\omega1_{ij}$, $\omega2_{ki}$, initialize the bias of the hidden layer $b1$, bias of the output layer $b2$, and the learning rate and nerve cell activation function are given.

Step 2: Calculation of the output of hidden layer. According to the input variable $X$, the connection weight between the input layer and hidden layer $\omega1_{ij}$ as well as the bias $b1$ of the hidden layer, the output of hidden layer shall be calculated $y1$.

$$y1_i = f1(\sum_{j=1}^{n} \omega1_{ij}x_j + b1_j), (j = 1, 2, \cdots, l1) \quad (7)$$

Where $l1$ is the node of the hidden layer, $f1$ is the activation function of the hidden layer, and $S$ logarithm or tangent activation function and linear function are commonly applied, in this paper, the function selected is:

$$f(x) = \frac{1}{1 + exp^{-x}} \quad (8)$$

Step 3: Calculation of the output of output layer. According to the output $y1$ of the hidden layer, the connection weight $\omega2_{ki}$ and bias $b2$, the predicted output $y2$ is calculated.

$$y2_k = f2(\sum_{i=1}^{n} \omega2_{ki}y1_i + b2_k), (k = 1, 2, \cdots, l2) \quad (9)$$

Step 4: Error calculation. According to the network prediction output $y2$ , the network prediction error $E$ is calculated.

$$E(W, B) = \frac{1}{2}\sum_{k=1}^{l2}(t_k - y2_k)^2 \quad (10)$$

Step 5: Updating the weight and bias. Update the network weight variation $\Delta\omega1_{ij}$, $\Delta\omega2_{ki}$ and threshold value variation $\Delta b1$ and $\Delta b2$ according to the network prediction error.

Step 6: Judge if the iteration of the algorithm ends, if it does not end, return to step 2.

## 5 GA-BP algorithm

BP learning algorithm is a global approximation algorithm, with good generalization ability. But the initial weight and threshold value of BP neural network is selected at random, and if the parameters are not selected properly, it may slow down the network convergence rate and fall into the local optimum. The genetic algorithms was a parallel immediate search optimization method proposed by professor Holland from Michigan University in 1962, and it was generated by simulating the natural genetic mechanism and biological evolutionism, shortened as GA algorithm. It is a global optimization algorithm, which may overcome the defects of artificial neural network, and improve the network training speed and prediction precision [22]. Consequently, genetic algorithm is employed to determine the initial weight and threshold of the BP neural network, which may be favorable for the fast convergence and obtaining of globally optimal solution, and predict with the build prediction model.

The process of optimizing the BP network weight and threshold value with genetic algorithm:

Step1: Generate a group of distribution at random, encode each weight and threshold value with a certain coding plan, thus to construct the s-chains, which may include all weights and threshold value of the neural network. When the network structure is known, a neutral network with confirmed structure, weight and threshold value is constructed.

Step2: Fitness function calculation, the error function of BP network is taken as the fitness function, for calculating the fitness of each chromosome in the population.

Step3: Operator is selected with roulette method, namely, the chromosome of each generation of the population shall be selected based on the selection strategy of the fitness scale.

Step4: Interlace operation, two individuals shall be selected, and a new excellent individual can be produced through the exchange of chromosomes.

Step5: Mutation operation, selection an individual from the population at random, and select a point of the chromosome for mutation to produce more excellent individual.

Step6: Repeat the step (2)- (5), for evolving the initial weight constantly, till the training objective can be satisfied, or the iteration times reach the preset goal.

## 6 Simulation

In this paper, the stock data from the stock trading software of communication, and the experiment was conducted in the MATLAB (R2014b).The data for the stock market prediction experiments has been collected for FAW Car Co., Ltd. (for short, FAW Car. Stock code: 000800) and China Minmetals Rare Earth Co., Ltd. (for short, Minmetals Rare Earth. Stock code: SZ000831). The total number of samples for the stock indices is 134 trading days, from 10rd October 2014 to 27th April 2015. Each sample consists of 19 technical indicators as the shole features.Among the data, the prices of the first 120 days are for training and those of the latter 14 days are for testing. Models are used for predicting the opening price of the index one day.

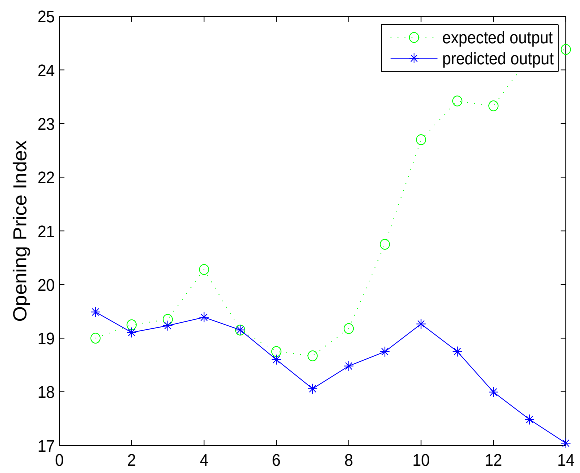
In this paper, a three-layer neural network model with a hidden layer is established on the basis of BP neural network and combining the genetic algorithm, and we used principal component analysis method to extract the principal components of the contribution rate to meet 95%. When taking ten principal components, the cumulative contribution rate of the Faw car and Minmetals Rare Earth reach 99.95% and 95.69%, respectively. So we select ten principal components as our input,thus to make predictions. The prediction made only with the SVM and BP, the PCA-SVM and GA-BP combination prediction are compared with their results in Table 1, Table 2, Figure 1 and Figure 2.

Table 1: The test results of FAW Car.

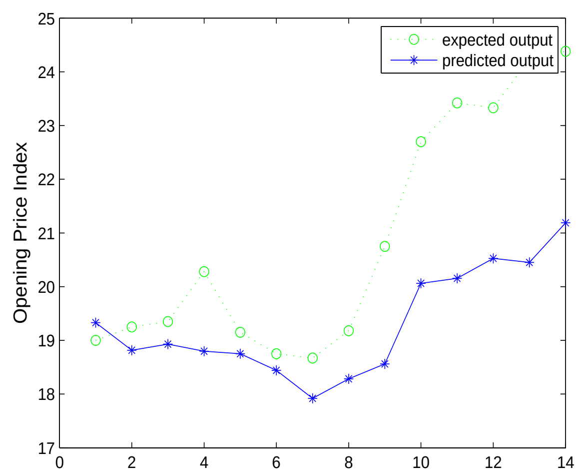
| Test Index | Time | MSE | $R^2$ |
|---|---|---|---|
| SVM | 7 | 0.17115 | 34.72% |
| PCA-SVM | 4 | 0.06039 | 84.68% |
| BP | 35 | 0.00101 | 97.55% |
| GA-BP | 43 | 0.00002 | 98.71% |

Table 2: The test results of Minmetals Rare Earth.

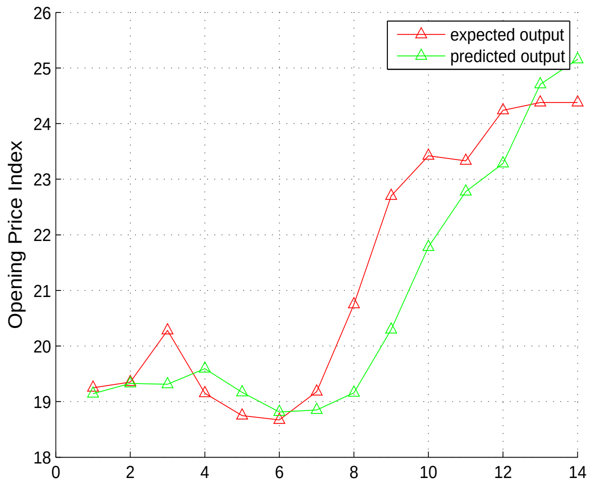
| Test Index | Time | MSE | $R^2$ |
|---|---|---|---|
| SVM | 6 | 0.06986 | 41.74% |
| PCA-SVM | 5 | 0.00739 | 81.10% |
| BP | 34 | 0.00123 | 97.22% |
| GA-BP | 41 | 0.00007 | 97.47% |



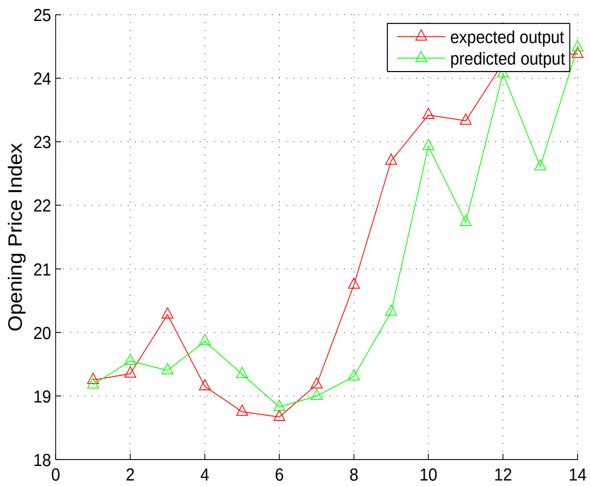

(a)




(b)


expected output
predicted output
Opening Price Index

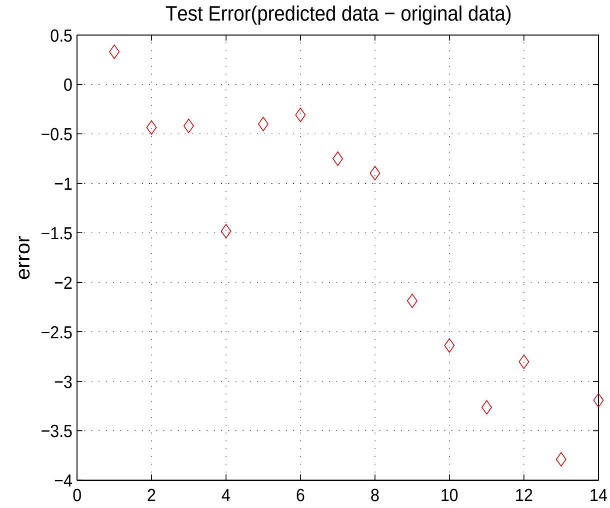
(c)


expected output
predicted output
Opening Price Index

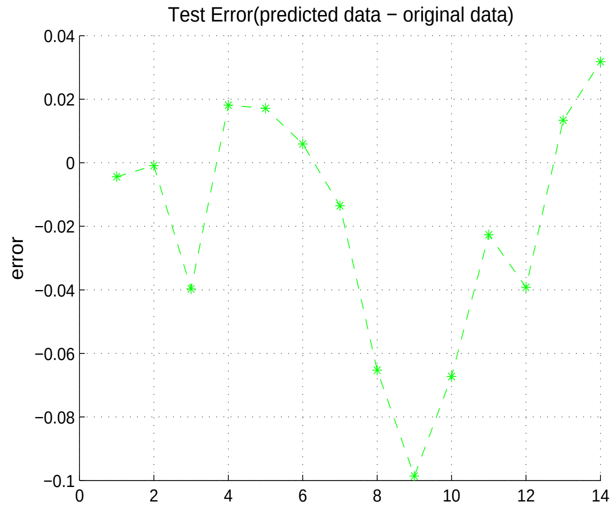
(d)


Test Error(predicted data − original data)
error

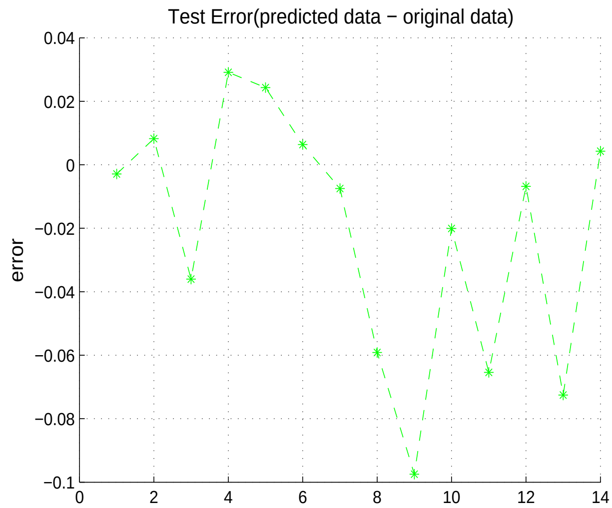
(e)


Test Error(predicted data − original data)
error

(f)

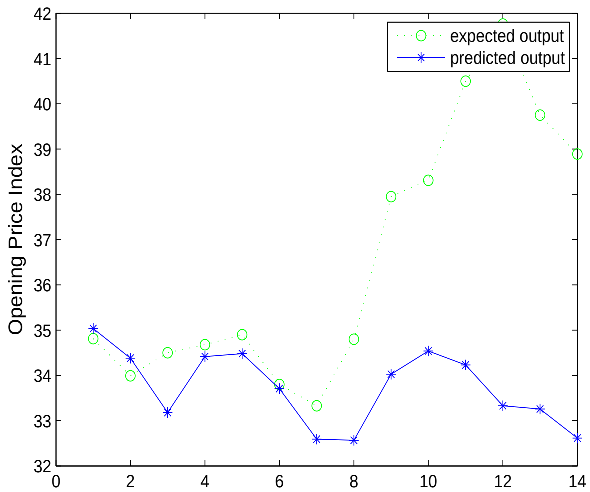

Test Error(predicted data − original data)
error

(g)

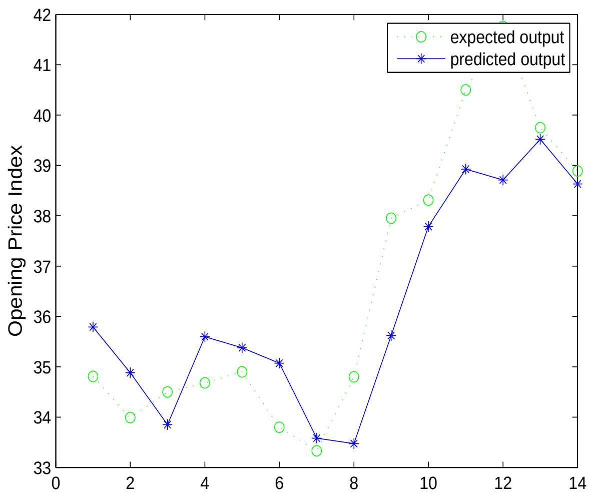

Test Error(predicted data − original data)
error

(h)

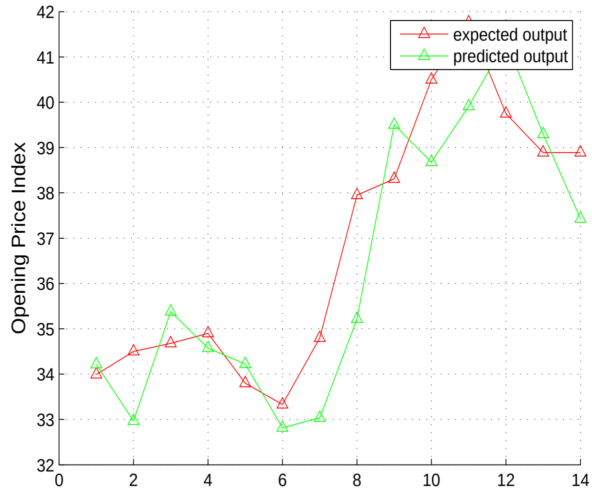
Figure 1: Test results of the FAW Car:(a) Test results of the Faw car in SVM model. (b) Test results of the Faw car in PCA-SVM model. (c) Test results of the Faw car in BP model. (d) Test results of the Faw car in GA-BP model. (e) Test error results of the Faw car in SVM model. (f) Test error results of the Faw car in PCA-SVM model. (g) Test error results of the Faw car in BP model. (h) Test results of the Faw car in GA-BP model.



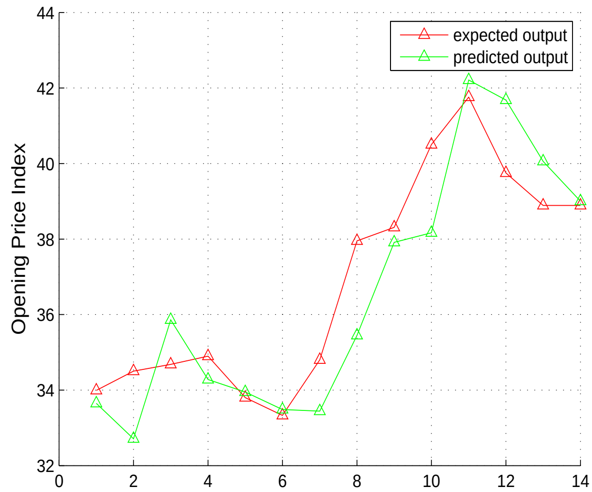

(a)



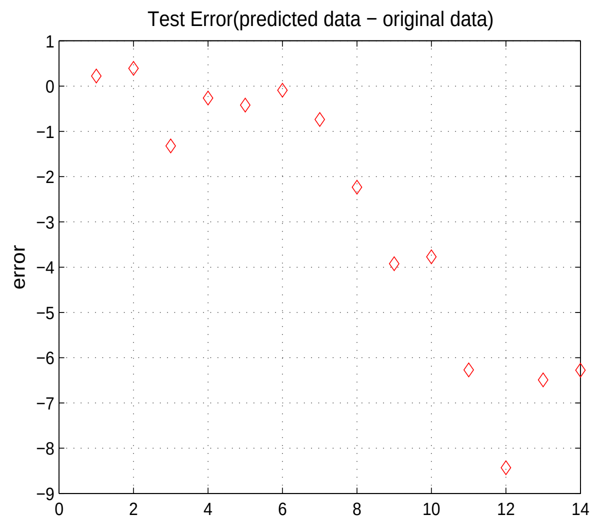

(b)



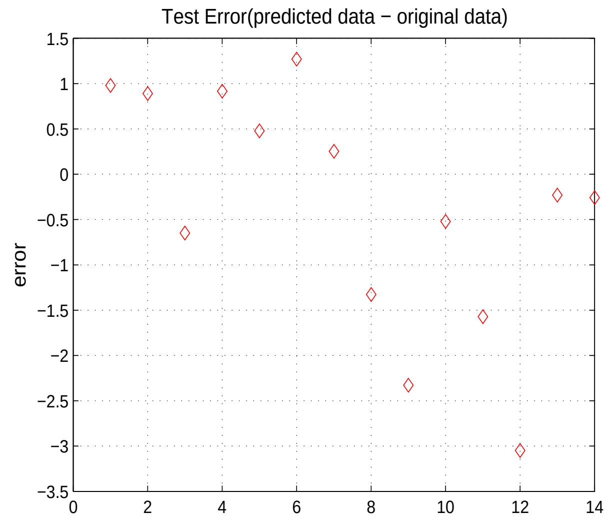

(c)



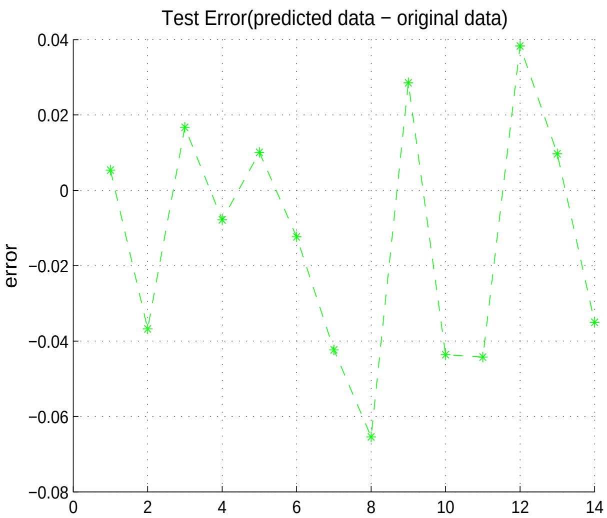

(d)



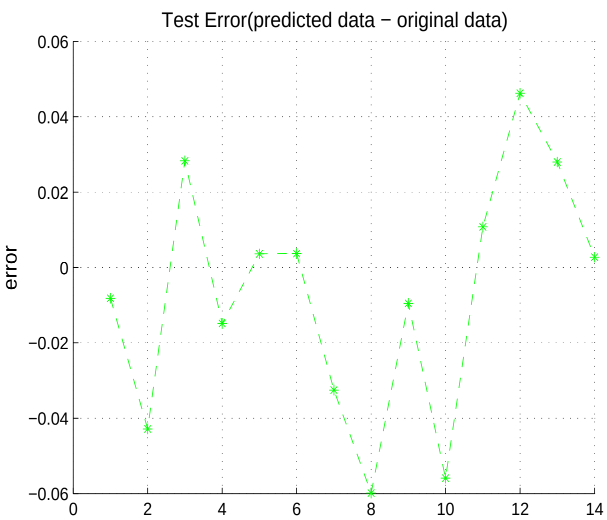

(e)




(f)




(g)




(h)

Figure 2: Test results of the Minmetals Rare Earth:(a) Test results of the Minmetals Rare Earth in SVM model. (b) Test results of the Minmetals Rare Earth in PCA-SVM model. (c) Test results of the Minmetals Rare Earth in BP model. (d) Test results of the Minmetals Rare Earth in GA-BP model. (e) Test error results of the Minmetals Rare Earth in SVM model. (f) Test error results of the Minmetals Rare Earth in PCA-SVM model. (g) Test error results of the Minmetals Rare Earth in BP model. (h) Test error results of the Minmetals Rare Earth in GA-BP model.

To measure the prediction performance, we extracted two major indexes, which respectively are the mean squared error (MSE) and squared correlation coefficient ($R^2$ ), and it is the better for more close to 1. The mathematical formula of MSE and $R^2$ can be expressed as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 \tag{11}$$

$$R^2 = \frac{\left(n\sum\limits_{i=1}^{n} f(x_i)y_i - \sum\limits_{i=1}^{n} f(x_i)\sum\limits_{i=1}^{n} y_i\right)^2}{\left(n\sum\limits_{i=1}^{n} f(x_i)^2) - \left(\sum\limits_{i=1}^{n} f(x_i)\right)^2\right)\left(n\sum\limits_{i=1}^{n} y_i^2 - (\sum\limits_{i=1}^{n} y_i)^2\right)} \tag{12}$$

From Table 1 and Table 2 we can seen that making combination predictions is more desirable than making mere SVM and BP predictions, and GA-BP combination predictions are better than PCA-SVM combination predictions, and run time of PCA-SVM is more shorter than others models, because the PCA-SVM method is using principal component analysis method to reduce the dimension of input variables. Figure 1 and Figure 2 shows that the final predictive value of GA-BP model provides a good fitting to the trend, and the GA-BP method can get higher prediction accuracy than other three prediction methods.

## 7 Conclusions

The stock market is a non-stationary time series, using conventional forecast method is often ineffective. This paper hybridizes genetic algorithm and artificial neural network method, and hybridizes principal component analysis and support vector machine to predict the next opening price in stock markets. The experimental results show GA-BP method can get higher prediction accuracy than other three prediction methods.

This paper only takes the technical indicators into consideration. In fact there are many factors affecting the stock price, so we still need to be improved and our work in this area is ongoing, and we strive to make the prediction practical as soon as possible.

**Acknowledgements:** The work is fully supported by the National Natural Science Foundation Grant no.61275120 of China.

*References:*

[1] Zhang, Xiangzhou, et al. "A causal feature selection algorithm for stock prediction modeling. " *Neurocomputing*, vol. 142, pp. 48–59, 2014.

[2] Wu, Shuen-De, et al. "Multi-scale analysis based ball bearing defect diagnostics using Mahalanobis distance and support vector machine."*Entropy*, vol. 15, pp. 416–433, 2013.

[3] Ni, Li-Ping, Zhi-Wei Ni, and Ya-Zhuo Gao. "Stock trend prediction based on fractal feature selection and support vector machine." *Expert Systems with Applications*, vol. 38, pp. 5569–5576, 2011.

[4] Fenghua, W. E. N., et al. "Stock Price Prediction based on SSA and SVM."*Procedia Computer Science*, vol. 31, pp. 625–631, 2014.

[5] Rather, Akhter Mohiuddin, Arun Agarwal, and V. N. Sastry. "Recurrent neural network and a hybrid model for prediction of stock returns." *Expert Systems with Applications*, vol. 42, pp. 3234–3241, 2015.

[6] Dai, Wensheng, Jui-Yu Wu, and Chi-Jie Lu. "Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes." *Expert Systems with Applications*, vol. 39, pp. 4444–4452, 2012.

[7] Yang, Baoan, et al. "An early warning system for loan risk assessment using artificial neural networks." *Knowledge-Based Systems*, vol. 14, pp. 303–306, 2001.

[8] Zhang, Yudong, and Lenan Wu. "Stock market prediction of SP 500 via combination of improved BCO approach and BP neural network." *Expert systems with applications*, vol. 36, pp. 8849–8854, 2009.

[9] Deb, Kalyanmoy, et al. "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II." *Lecture notes in computer science*, vol. 1917, pp. 849–858, 2000.

[10] Kara, Yakup, Melek Acar Boyacioglu, and ömer Kaan Baykan. "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange."*Expert systems with Applications*, vol. 38, pp. 5311–5319, 2011.

[11] Xuegong, Zhang. "Introduction to statistical learning theory and support vector machines. " *Acta Automatica Sinica*, vol. 26, pp. 32–42, 2000.

[12] Lin, Chih-Jen. "Formulations of support vector machines: a note from an optimization point of view." *Neural Computation*, vol. 13, pp. 307–317, 2001.

[13] Lee, Ming-Chi. "Using support vector machine with a hybrid feature selection method to the stock trend prediction." *Expert Systems with Applications*, vol. 36, pp. 10896–10904, 2009.

[14] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning*, vol. 20, pp. 273–297, 1995.

[15] Luo, Linkai, and Xi Chen. "Integrating piecewise linear representation and weighted support vector machine for stock trading signal prediction." *Applied Soft Computing*, vol. 13, pp. 806–816, 2013.

[16] Xiaochuan Wang, Feng Shi, Yang Li. "The analysis of 43 cases in MATLAB Neural Network." *Beijing: Beijing University of Aeronautics and Astronautics Press*, (Beijing, China), Aug. 2013.

[17] Xie Zhonghua. "The analysis of 40 cases in MATLAB Statistical Analysis and Application." *Beijing: Beijing University of Aeronautics and Astronautics Press*, (Beijing, China), 2013.

[18] Hansen, Toke Jansen, Trine Julie Abrahamsen, and Lars Kai Hansen. "Denoising by semi-supervised kernel PCA preimaging." *Pattern Recognition Letters*, vol. 49, pp. 114–120, 2014.

[19] Zhou, Changjun, et al. "Face recognition based on PCA and logistic regression analysis." *Optik-International Journal for Light and Electron Optics*, vol. 125, pp. 5916–5919, 2014.

[20] Li, Changdong, et al. "Application of back-propagation neural network on bank destruction forecasting for accumulative landslides in the three Gorges Reservoir Region, China." *Stochastic Environmental Research and Risk Assessment*, vol. 28, pp. 1465–1477, 2014.

[21] Chen, Guangying, et al. "The genetic algorithm based back propagation neural network for MMP prediction in CO 2-EOR process." *Fuel*, vol. 126, pp. 202–212, 2014.

[22] Yang, Yongqiang, Guijun Wang, and Yang Yang. "Parameters optimization of polygonal fuzzy neural networks based on GA-BP hybrid algorithm." *International Journal of Machine Learning and Cybernetics*, vol. 5, pp. 815–822, 2014.