2008 International Conference on Management Science & Engineering (15th)
September 10-12, 2008 Long Beach, USA

Stock Market Forecasting Model Based on A Hybrid ARMA and Support Vector Machines

ZHANG Da-yong[1,2], SONG Hong-wei[2], CHEN Pu[2]
1 Postdoctoral Research Station of Computer Science and Technology, Harbin Institute of Technology, P.R.China, 150001
2 School of Management, Harbin Institute of Technology, P.R.China, 150001

Abstract: Stock market forecasting has attracted a lot of research interests in previous literature. Traditionally, the autoregressive moving average (ARMA) model has been one of the most widely used linear models in time series forecasting. However, the ARMA model cannot easily capture the nonlinear patterns. And recent studies have shown that artificial neural networks (ANN) method achieved better performance than traditional statistical ones. ANN approaches have, however, suffered from difficulties with generalization, producing models that can overfit the data. Support vector machines (SVMs), a novel neural network technique, have been successfully applied in solving nonlinear regression estimation problems. Therefore, this investigation proposes a hybrid methodology that exploits the unique strength of the ARMA model and the SVMs model in the stock market forecasting problem in an attempt to provide a model with better explanatory power. Real data sets of stock market were used to examine the forecasting accuracy of the proposed model. The results of computational tests are very promising.

Keywords: BP neural network, financial time series, forecasting, support vector machine

1 Introduction

Stock market prediction is regarded as a challenging task because of high fluctuation and irregularity. [1] Thus, numerous models have been depicted to provide the investors with more precise predictions. In particular, artificial neural networks (ANN) method is most frequently used in previous literature [2~4] since the power of prediction is known to be better than the others. Due to the difficulty of interpreting the neural network models, however, most studies applying neural networks focused on prediction accuracy. Few efforts of using ANN models to provide better understanding of the bankruptcy prediction process have been reported in the literature[5]. In addition, ANN suffers from difficulties with generalization because of overfitting, and fully depends on researchers' experience or knowledge for preprocessing of selecting a large number of control parameters that include relevant input variables, hidden layer size, learning rate, and momentum [6~8].

Recently, the support vector machine (SVM) method, which was first suggested by Vapnik (1995) [9], has recently been used in a range of applications, including financial stock market prediction. The foundation of SVM has been developed by Vapnik, and it is gaining popularity due to many attractive features, and excellent generalization performance on a wide range of problems. The formulation embodies the structural risk minimization (SRM) principle, which has been shown to be superior to traditional empirical risk minimization (ERM) principle employed by conventional neural networks [10]. SRM minimizes an upper bound of generalization error, as opposed to ERM that minimizes the error on training data. In addition, the solution of SVM may be global optimum while other neural network models tend to fall into a local optimal solution. The SVM technique, in general, is widely regarded as the state of art classifier and previous researches indicated that SVM prediction approaches are superior to neural networks approaches[11~14]. Initially developed for solving classification problems, SVM techniques can be successfully applied in regression[15,16]. Unlike pattern recognition problems where the desired outputs are discrete values, the support vector regression (SVR) deals with 'real valued' functions. The SVR is derived from the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error. Previous researches reported the SVR model has successfully solved prediction problem in many domains. However, improving the prediction accuracy still remains a prime issue of concern in the area of prediction. Especially for stock market prediction, even a slight improvement on prediction accuracy could have a positive impact on the profit of investment. It was reported that a hybrid system in prediction and classification achieved a higher performance level against the traditional system.[17~21]

Zhang [22] combined the ARIMA and feedforward neural networks models in forecasting. This study

This work is supported by the National doctoral funds (20070213008) and humanities and social sciences ofthe National Education Ministry of China Grant (07JC630027).

presents a hybrid model of ARMA and SVMs to slove the stock price forecasting problem.

2 Hybrid model in forecasting

2.1 Autoregressive moving average models

ARMA models have been proposed by Box and Jenkins as a mix between autoregressive and moving average models for the description of time series. In an autoregressive model of order p (ARp), each individual value x_t is expressed as a finite sum of p previous values and white noise, z_t ;

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + z_t \tag{1}$$

The parameters ai can be estimated from the Yule–Walker equations which are a set of linear equations in terms of their autocorrelation coefficient. In a moving average model of order q (MAq), the current value x_t is expressed as a finite sum of q previous z_t :

$$x_t = \beta_0 z_t + \beta_1 z_{t-1} + \cdots \beta_q z_{t-q} \tag{2}$$

In this equation, z_i is the white noise residual of the easured and predicted value of x at time instance i .The model parameters β_i usually are determined by a set of nonlinear equations in terms of the auto correlations. The z's are usually scaled so that $\beta_0 = 1$. In the past, moving average models have particularly been used in the field of econometrics where economic indicators can be affected by a variety of 'random' events such as strikes or government decisions. An ARMA model with order (p,q) is a mixed ARp and MAq model and is given by:

$$\begin{aligned} x_t &= \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + z_t + \beta_1 z_{t-1} \\ &+ \cdots + \cdots + \beta_q z_{t-q} \end{aligned} \tag{3}$$

Using the backward shift operator B , the previous equation can be written as:

$$\phi(B) x_t = \theta(B) z_t \tag{4}$$

where $\phi(B)$ and $\theta(B)$ are polynomials of order p, q, respectively, such that:

$$\begin{aligned} &\phi(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p \quad \text{and} \quad \theta(B) \\ &= 1 - \beta_1 B + \cdots + \beta_q B^q \end{aligned} \tag{5}$$

The ARMA model is basically a data-oriented approach that is adapted from the structure of the data themselves. However, any significant nonlinear data set limit the ARMA. Therefore, the proposed hybrid model used the SVMs to deal with the nonlinear data pattern.

2.2 Theory of SVMs in stock market forecasting

Considering a set of training data $\{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$,where each $x_i \subset R^n$ denotes the input space of the sample and has a corresponding target value $y_i \subset R$ for i=1,…, l where l corresponds to the size of the training data[12]. The idea of the regression problem is to determine a function that can approximate future values accurately.

The generic SVR estimating function takes the form:

$$f(x) = (w \cdot \Phi(x)) + b \tag{6}$$

where $w \subset R^n$, $b \subset R$ and Φ denotes a non-linear transformation from R^n to high dimensional space. Our goal is to find the value of w and b such that values of x can be determined by minimizing the regression risk:

$$R_{reg}(f) = C \sum_{i=0}^{\ell} \Gamma(f(xi) - yi) + \frac{1}{2}\|w\|^2 \tag{7}$$

where $\Gamma(\cdot)$ is a cost function, C is a constant and vector w can be written in terms of data points as:

$$w = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \Phi(x_i) \tag{8}$$

By substituting equation (3) into equation (1), the generic equation can be rewritten as:

$$\begin{aligned} f(x) &= \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*)(\Phi(x_i) \cdot \Phi(x)) + b \\ &= \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) k(x_i, x) + b \end{aligned} \tag{9}$$

In equation (4) the dot product can be replaced with function $k(x_i, x)$, known as the kernel function. Kernel functions enable dot product to be performed in high-dimensional feature space using low dimensional space data input without knowing the transformation Φ . All kernel functions must satisfy Mercer's condition that corresponds to the inner product of some feature space. The radial basis function (RBF) is commonly used as the kernel for regression:

$$k(x_i, x) = \exp\left\{-\gamma |x - x_i|^2\right\} \tag{10}$$

Some common kernels are shown in Table 1. In our studies we have experimented with these three kernels.

Tab.1 Common kernel functions

Kernels	Functions
Linear	$x \cdot y$
Polynomial	$[(x * x_i) + 1]^d$
RBF	$\exp\left\{-\gamma |x - x_i|^2\right\}$

The ε -insensitive loss function is the most widely used cost function. The function is in the form:

$$\Gamma(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon, & \text{for} \quad |f(x) - y| \ge \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

By solving the quadratic optimization problem in(12), the regression risk in equation (7) and the ε -insensitive loss function (11)can be minimized:

$$\frac{1}{2}\sum_{i,j=1}^{\ell}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)k(x_i,x_j)-\sum_{i=1}^{\ell}\alpha_i^*(y_i-\varepsilon)-\alpha_i(y_i+\varepsilon)$$

subject to

$$\sum_{i=1}^{\ell}\alpha_i-\alpha_i^*=0,\quad \alpha_i,\alpha_i^*\in[0,C] \tag{12}$$

The Lagrange multipliers, α_i $and \alpha_i^*$, represent solutions to the above quadratic problem that act as forces pushing predictions towards target value y_i. Only the non-zero values of the Lagrange multipliers in equation (12) are useful in forecasting the regression line and are known as support vectors. For all points inside the ε -tube, the Lagrange multipliers equal to zero do not contribute to the regression function. Only if the requirement $|f(x)-y|\ge\varepsilon$ (See Figure 1) is fulfilled, Lagrange multipliers may be non-zero values and used as support vectors.

The constant C introduced in equation (13) determines penalties to estimation errors. A large C assigns higher penalties to errors so that the regression is trained to minimize error with lower generalization while a small C assigns fewer penalties to errors; this allows the minimization of margin with errors, thus higher generalization ability. If C goes to infinitely large, SVR would not allow the occurrence of any error and result in a complex model, whereas when C goes to zero, the result would tolerate a large amount of errors and the model would be less complex.

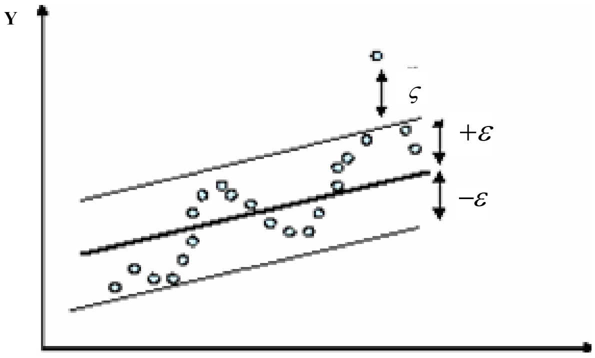

Fig.1 Support vector regression to fit a tube with radius to the data and positive slack variables ζ measuring the points lying outside of the tube.

Now, we have solved the value of w in terms of the Lagrange multipliers. For the variable b, it can be computed by applying Karush-Kuhn-Tucker (KKT) conditions which, in this case, implies that the product of the Lagrange multipliers and constrains has to equal zero:

$$\begin{aligned}&\alpha_i(\varepsilon+\zeta_i-y_i+(w,x_i)+b)=0\\&\alpha_i^*(\varepsilon+\zeta_i^*+y_i-(w,x_i)-b)=0\end{aligned} \tag{13}$$

$$\text{And}\quad \begin{aligned}&(C-\alpha_i)\zeta_i=0\\&(C-\alpha_i^*)\zeta_i^*=0\end{aligned} \tag{14}$$

where ζ_i and ζ_i^* are slack variables used to measure errors outside the ε -tube. Since $\alpha_i,\alpha_i^*=0$ and $\zeta_i^*=0$ for $\alpha_i^*\in(0,C)$, b can be computed as follows:

$$\begin{aligned}&b=y_i-(w,x_i)-\varepsilon\quad for\ \alpha_i\in(0,C)\\&b=y_i-(w,x_i)+\varepsilon\quad for\ \alpha_i^*\in(0,C)\end{aligned} \tag{15}$$

Putting it all together, we can use SVM and SVR without knowing the transformation.

2.3 The hybrid methodology

The behavior of foreign debt risk can not easily be captured. Therefore, a hybrid strategy that has both linear and nonlinear modeling abilities is a good alternative for forecasting foreign debt risk. Both the ARMA and the SVMs models have different capabilities to capture data characteristics in linear or nonlinear domains, so the hybrid model proposed in this study is composed of the ARIMA component and the SVMs component. Thus, the hybrid model can model linear and nonlinear patterns with improved overall forecasting performance. The hybrid model (Zt) can then be represented as follows:

$$z_t=Y_t+N_t \tag{16}$$

where Y_t is the linear part and N_t is the nonlinear part of the hybrid model. Both Y_t and N_t are estimated from the data set. $\tilde{Y}_t$ is the forecast value of the ARMA model at time t. Let ε_t represent the residual at time t as obtained from the ARMA model; then

$$\varepsilon_t=Z_t-\tilde{Y}_t \tag{17}$$

The residuals are modeled by the SVMs and can be represented as follows:

$$\varepsilon_t=f(\varepsilon_{t-1},\varepsilon_{t-2},\cdots,\varepsilon_{t-n})+\Delta_t \tag{18}$$

where f is a nonlinear function modeled by the SVMs and Δ_t is the random error. Therefore, the combined forecast is $\tilde{Z}_t=\tilde{Y}_t+\tilde{N}_t$

Notably, $\tilde{N}_t$ is the forecast value of (17).

3 Experiment and comparisons

3.1 Data description

In the experiments, two stock indices, S&P500 and Nikkei225, are daily and are obtained from Data stream. The entire data set covers the period from January 1 2000 to December 31 2004. The data sets are divided into two periods: the first period covers from January 1 2000 to December 31 2003 while the second period is from January 1 2004 to December 31 2004. The first period, which is assigned to in-sample estimation, is used to network learning and training. The second period is reserved for out-of-sample evaluation. For brevity, the original data are not listed in the paper, and detailed data can be obtained from the sources.

3.2 Performance criteria

The inspection of the predict result is the key of the forecasting performance of the model, because we can acquire the information of the characteristic of the different forecasting methods, and this is very useful for the people to choose and use the variety of the forecasting methods. The prediction performance is evaluated using the following statistical metrics, namely, the mean absolute error（MAE）, mean absolute percent error（MAPE）, mean squared error (MSE) and root mean square error(RMSE). The definitions of these criteria as follows:

(1)Mean absolute error（MAE）

$$MAE = \frac{1}{n}\sum_{i,j=1}^{n}\left|x_i - x_j\right| \quad i,j=1,2,\cdots,n \tag{19}$$

(2)Mean absolute percent error（MAPE）

$$MAPE = \frac{1}{n}\sum_{i,j=1}^{n}\left|\frac{x_i - x_j}{x_i}\right| \quad i,j=1,2,\cdots,n \tag{20}$$

(3)Mean squared error（MSE）

$$MSE = \frac{1}{n}\sum_{i,j=1}^{n}\left(x_i - x_j\right)^2 \quad i,j=1,2,\cdots,n \tag{21}$$

(4)Root-mean-square error（RMSE）

$$RMSE = \sqrt{\frac{1}{n}\sum_{i,j=1}^{n}\left(x_i - x_j\right)^2} \quad i,j=1,2,\cdots,n \tag{22}$$

The smaller the values of the four indexes, the higher are the prediction accuracy. Where: x_i represents the actual values; x_i represents the forecasting values; n represents the number of the sample.

3.3 The selection of the parameter

In this study, the ARMA model has three phases: model identification, parameter estimation, and diagnostic checking. After determining the suitable parameters of the ARMA model, it is important to examine how closely the proposed model fits a given time series. The autocorrelation function (ACF) was calculated to verify the parameters. Fig. 2 plots ACF of

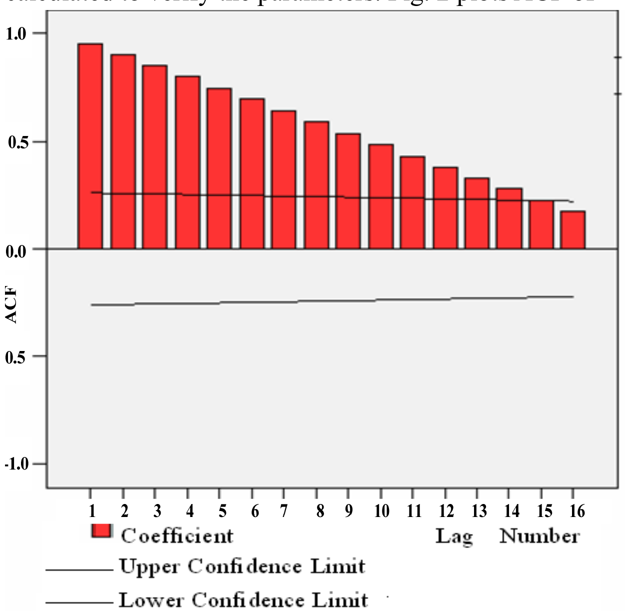

Fig.2 ACF of stock indices

stock indices. Fig.3 plots the PACF of stock indices. Fig.4 plots the estimated residual ACF and indicates that the residuals are not autocorrelated. PACF, the partial autocorrelation function, displayed in Fig. 5, is also used to check the residuals and indicates that the residuals are not correlated.

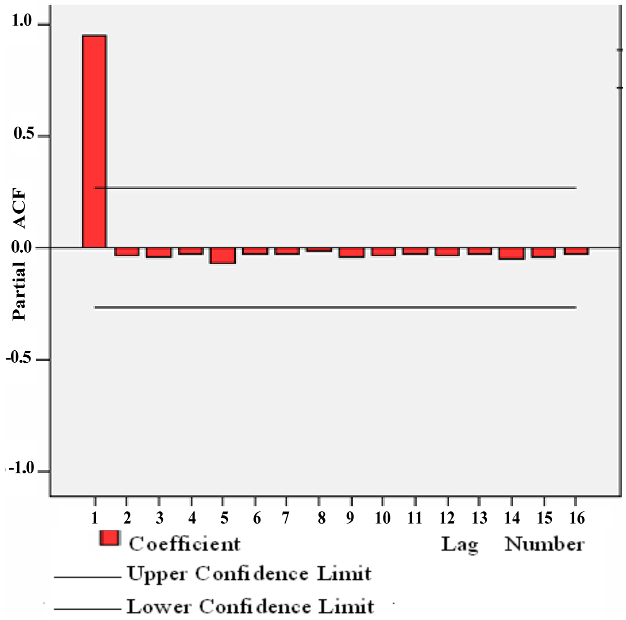

Fig.3 PACF of stock indices

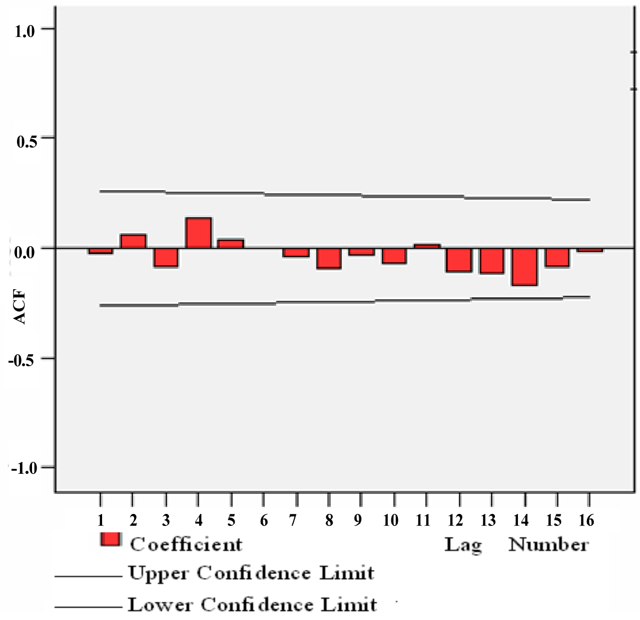

Fig.4 Estimated residual ACF

For the SVMs models, three parameters: δ and ε and C, were adjusted based on the validation sets. It is well known that SVM generalization performance (estimation accuracy) depends on a good setting of meta-parameters parameters C, and the kernel parameters. The problem of optimal parameter selection is further complicated by the fact that SVM model complexity (and hence its generalization performance) depends on all three parameters. Existing software implementations of SVM regression usually treat SVM meta-parameters as user-defined inputs. Selecting a particular kernel type and kernel function parameters is usually based on application-domain knowledge and also should reflect distribution of input (x) values of the training data.

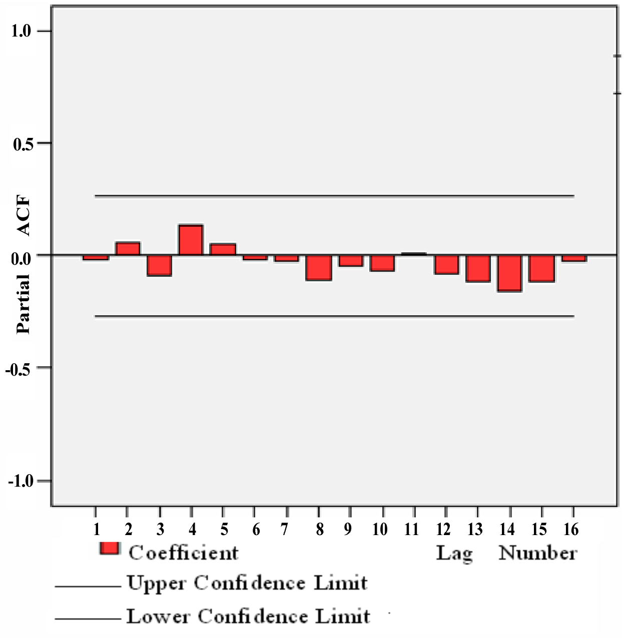

Fig.5 Estimated residual PACF

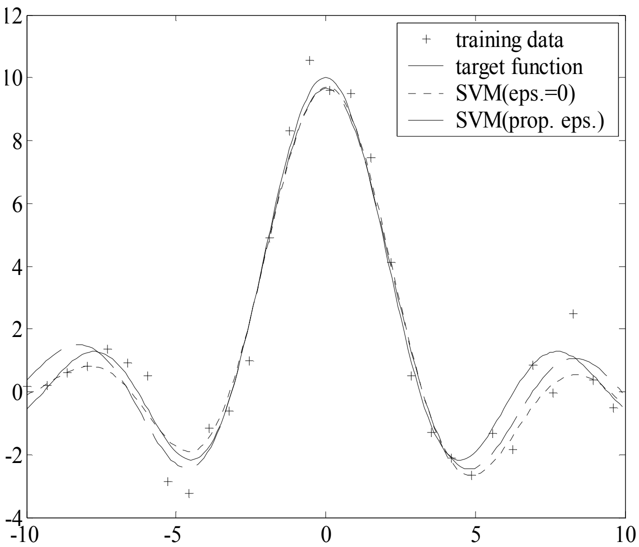

Fig. 6 SVM estimate using proposed parameter selection vs using ARMA

Parameter C determines the trade off between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in optimization formulation for example, if C is too large (infinity), then the objective is to minimize the empirical risk only, without regard to model complexity part in the optimization formulation.

Parameter ε controls the width of the ε -insensitive zone, used to fit the training data. The value of ε can affect the number of support vectors used to construct the regression function. The bigger ε, the fewer support vectors are selected. On the other hand, bigger ε-values results in more 'flat' estimates. Hence, both C and ε-values affect model complexity (but in a different way).

3.4 Experiment results

When the data are prepared, we begin to train hybrid model. In these experiments, we prepare 5 years' daily data. We use the first 4 years' daily data to train and validate the network, and use the last one years' data to test the prediction performance. For comparison, the standard three-layer BP neural network is used as benchmark model. This study varies the number of nodes in the hidden layer and stopping criteria for training. In this study, 5, 10, 20 hidden nodes for each stopping criteria because the BP network does not have a general rule for determining the optimal number of hidden nodes. The study uses 500, 1000, 2000 and 4000 learning epochs for the stopping criteria of BPNN. For standard BPNN model, the learning rate is set to 0.25. The hidden nodes use the sigmoid transfer function and the output node uses the linear transfer function. The study allows 5 input nodes in terms of the results of auto-regression testing.

Table 2 compares the forecasting results of different models. Those results indicate that the hybrid model outperforms the other two individual models (model 1 and 2) in terms of four indices, revealing that the hybrid model can capture all of the patterns in the data., and it can significantly reduce the overall forecasting errors. Furthermore, the results in Table 3 indicate the proposed SVR model is superior to backpropagation neural network model .

From the results of the empirical experiment, we can conclude that Hybrid shows better performance in stock market prediction problem avoiding overfitting and exhaustive parameter search.

Tab.2 Comparison of forecasting indices

	MAE	MAPE	MSE	RMSE
BNN model				
S&P 500	0.3495	0.2257	1.1494	0.4751
Nikkei 225	0.8726	1.1126	1.0794	1.0548
Hybrid model				
S&P 500	0.3466	0.2257	1.1433	0.4740
Nikkei 225	0.9120	1.4676	1.1359	1.2115

4 Conclusion

For more than half a century, the autoregressive moving average model has dominated many areas of time series forecasting. Recently, ANN has demonstrated the capability to capture the nonlinear data pattern. This study is motivated by evidence that different forecasting models can complement each other in approximating data sets, and proposed a hybrid model of the ARMA and the SVMs. The presented model is believed to greatly improve the prediction performance of the single ARIMA model or the single SVMs model in forecasting stock market. And then this exploratory research examines the potential of using an hybrid model to predict two main international stock indices, S&P 500 and Nikkei 225. Our empirical results suggest that the this model may provide better forecasts than the standard BPNN model. The comparative evaluation is based on a variety of statistics such as MAE and RMSE. For two stock indices included in our empirical investigation, the adaptive BPNN model outperforms the standard BPNN model in terms of MAE and RMSE. Furthermore, our experimental analyses reveal that the MAE and RMSE for two stock indices using the proposed method model are significantly better than those obtained using the standard BPNN model. Theoretically as well as empirically, hybridizing two dissimilar models reduces forecasting errors. However, future research should address some problems. This study demonstrated that a simple combination of the two best individual models does not necessarily produce the best results. Therefore, the structured selection of optimal parameters of the hybrid model is of great interest.

References

[1]Lean Yu, Shouyang Wang, and Kin Keung Lai. A novel adaptive learning algorithm for stock market prediction[C]//Lecture Notes in Computer Science , 3827, 2005: 443-452.

[2]Baba N, Kozaki M. An intelligent forecasting system of stock price using neural networks[C]// Proceedings of the international joint conference on neural networks, 1992:371-377.

[3]Cheung YM, Lai Hzh, Xu L. Application of adaptive RPCLCLP with trading system to foreign exchange investment[C]//Proceedings of the IEEE International Conference on neUral Networks, 1996:2033-2038.

[4]Saad EW, Prokhorov DV, Wunsch DC. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks[J]. IEEE Transactions on Neural Networks, 1998,9:1456-1470.

[5]Huang, Z., H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu. Credit rating analysis with support vector machine and neural networks: A market comparative study[J]. Decision Support Systems, 2004, 37: 543-558.

[6] Geman, S., E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma[J].Neural Computation, 1992, 4: 1-58.

[7]Lawrence, S., C.L. Giles, and A.-C. Tsoi. Lessons in neural network training: Overfitting may be harder than expected[C]//Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAl-97, AAAl Press, Mento Park, California, 1997:540-545.

[8]Sarle, W.S. Stopped training and other remedies for overfitting[C]// Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics, 1995:352-360.

[9]Vapnik, V. Statistical learning theory[M]. New York: Springer, 1998:121-126.

[10]Gunn, S.R., M. Brown and K.M. Bossley. Network performance assessment for neurofuzzy data modeling [R]. Lecture Notes in Computer Science, 1997,1280: 313-323.

[11]Chen, W. H., & Shih, J. Y. A study of Taiwan's issuer credit rating systems using support vector machines[J]. Expert Systems with Applications, 2006,30(3):427-435.

[12]Huang, C. L., Chen, M. C., &Wang, C. J. Credit card scoring with a data mining approach based on support vector machine[J]. Expert Systems with Applications, 2007, 33(4): 847-856.

[13] Huang, Z., Chen, H., Hsu, C. J., Chen, et al. Credit rating analysis with support vector machines and neural networks: A market comparative study[J]. Decision Support Systems, 2004,37(4): 543-558.

[14]Lee, Y. L. Application of support vector machines to corporate credit rating prediction[J]. Expert Systems with Applications, 2007, 33(1):67-74.

[15] Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. Support vector regression machines. Advances in Neural Information Processing Systems[M]. Cambridge MA: MIT Press, 1997:9.

[16]Vapnik, V., Golowich, S., & Smola, A. Support vector method for function approximation, regression estimation, and signal processing[M].Advances in Neural Information Processing Systems. Cambridge, MA: MIT press,1997:19-26.

[17]Brabazon, A., & Keenan, P. A hybrid genetic model for the prediction of corporate failure[J]. Computational Management Science,2004,1(3-4): 293-310.

[18]Chang, F. J., & Chang, Y. T. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir[J]. Advances in Water Resources, 2006, 29(1): 1-10.

[19]Kim, H. J., & Shin, K. S. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets[J]. Applied Soft Computing, 2007, 7(2):569-576.

[20]Lekakos, G., & Giaglis, G. A hybrid approach for improving predictive accuracy of collaborative filtering algorithms[J]. User Modeling and User Adapted Interaction, 2007, 17(5): 5-40.

[21]Min, S. H., Lee, J., & Han, I. Hybrid genetic algorithms and support vector machines for bankruptcy prediction[J]. Expert Systems with Applications, 2006, 31(3): 652-660.

[22]Zhang G. Peter Times series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003,50:159-175.