Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 1240–1248
Expert Systems with Applications
www.elsevier.com/locate/eswa

Dynamically exploring internal mechanism of stock market by fuzzy-based support vector machines with high dimension input space and genetic algorithm

Deng-Yiv Chiu *, Ping-Jie Chen

Department of Information Management, ChungHua University, No. 707, Section 2, WuFu Road, Hsin-Chu 300, Taiwan, ROC

Abstract

In the study, a new dynamic fuzzy model is proposed in combination with support vector machine (SVM) to explore stock market dynamism. The fuzzy model integrates various factors with influential degree as the input variables, and the genetic algorithm (GA) adjusts the influential degree of each input variable dynamically. SVM then serves to predict stock market dynamism in the next phase. In the meanwhile, the multiperiod experiment method is designed to simulate the volatility of stock market.

The input variables in the study include a total of 61 variables, including technical indicators in stock market, technical indicators in futures market, and the macroeconomic variables. To evaluate the performance of the new integrated model, we compare it with the traditional forecast methods and design different experiments to testify. In the experiment results, the model from the study does generate better accuracy rate than others.

Keywords: Fuzzy; Support vector machine; Genetic algorithm; Stock market dynamism

1. Introduction

Stock market is a complicated and volatile system due to too many possible influential factors. In the past studies, as a result, dynamism in the stock market was often considered as random movement. Nevertheless, according to the researches in the recent years, it is not entirely random. Instead, it is highly complicated and volatile (Black & Mcmillan, 2004). Many factors, including macroeconomic variables and stock market technical indicators, have been proven to have a certain level of forecast capability on stock market during a certain period of time (Lo, Mamaysky, & Wang, 2000). For example, the technical indicator that shows market volume or confidence proves to have forecast value on the futures transaction prices (Lo, 2004).

In the past decade, various methods have been widely applied in the stock market forecast such as linear and nonlinear mathematical models or multi-agent mechanism (Armano, Murru, & Roli, 2002; Kuo, 1998) to simulate the potential stock market transaction mechanism, such as artificial neural network (ANN) of multiple layers of threshold nonlinear function. Because of the advantages of arbitrary function approximation and needlessness of statistics assumption, ANN is widely applied in the simulation of potential market transaction mechanism (Matilla-Garcia & Arguellu, 2005; Oh & Kim, 2002). Also, to improve the forecast performance, some machine learning methods are applied. For example, genetic algorithm (GA) is used to reduce input feature dimension and select better model parameters (Davis, 1994) to increase the forecast accuracy rate.

Support vector machine (SVM) is a newly developed mathematical model with outstanding performances in handling high dimension entry space problems. Such a feature leads to a better performance of SVM in simulating potential market transaction mechanism than other methods.

* Corresponding author. Tel.: +886 1 3518 6524; fax: +886 1 3518 6543.
E-mail address: chiuden@chu.edu.tw (D.-Y. Chiu).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.11.022

An increasing number of researches adopt SVM to predict stock market dynamism and produce significant performances. Huang, Nakamori, and Wang (2004) adopt SVM to predict stock market dynamism with 9 macroeconomic factors. Yu, Wang, and Lai (2005) utilize SVM and GA to predict stock market dynamism with dynamically selecting 18 stock market technical indicators. However, few researches integrate the input variables into high dimension input space to fully develop the advantages of SVM. Meanwhile, most research experiments simply adopt two-period method, dividing the data into in-sample to obtain model parameters and out-of-sample to testify model accuracy rate.

Although the numerous related researches bring highly remarkable achievements, none of the models can continuously and successfully predict the dynamism of stock market (Schwert, 1989). The possible reasons are the high volatility of stock market and the selection of dynamic input variables. The volatility of the stock market makes factors influencing stock market change with time. An optimized forecast model is unable to guarantee to have the same forecast performance even after successful forecast of stock market dynamism during a certain period of time. For the selection of input variables, too few input variables will lead to inability to predict market mechanism due to insufficient factors and reduction of forecast accuracy rate. Too many input variables of forecast model will, however, bring too many noises and cause overfitting.

To improve the problems above, a new dynamic fuzzy model integrated with SVM is used in this study to explore the stock market dynamism. The fuzzy model, adjustable with time, is first used to consider influence factors with different features such as macroeconomic variables, technical indicators in both of stock market and futures market. GA locates the optimal parameters of fuzzy model for each influence factor changing with the time. Multiperiod experiment method divides data into many sections to train the forecast model with the earlier section data to predict the latter section data in order to simulate the stock market volatility. With the newly found influential degree of input variables, SVM handles high dimension input space without causing overfitting (Cristianini & Taylor, 2000) to explore the stock market movement dynamism in the next phase. Then, we compare it with other methods. The model from the study does generate better results than others.

In the literatures, SVM method was used to explore stock forecasting (Kim, 2003; Pai & Lin, 2005). GA-based feature selection methods were proposed (Armano, Marchesi, & Murr, 2005; Kim & Han, 2000; Kuo, Chen, & Hwang, 2001). Fuzzy method was adopted in some researches (Huang & Wang, 2006; Huang & Yu, 2005; Wang, 2003; Wong & Wang, 1991).

The architecture of the study is: Section 2 describes the SVM. Section 3 introduces genetic algorithm. Section 4 will provide detailed introduction of each part of the proposed integrated model. Section 5 testifies the performance of the proposed integrated architecture with stock market in Taiwan. Section 6 summarizes the conclusions and contribution of the study.

2. Support vector machine (SVM)

2.1. Linear separation

SVM defines the input variable supposition space with linear function and introduces the learning deviation to learn the mapping between input and output. As the linear fitting machine is operated in the feature space of the kernel function for learning, when the applied field has high dimension feature space feature, SVM shall effectively avoid overfitting problem and pose excellent learning performance (Cristianini & Taylor, 2000).

The main concept is to transfer the mapping of input space kernel to high dimension feature space before re-classification. To begin, SVM selects several support vectors from the training data to represent the entire data. In this study, the issue can be expressed as follows:

Provide the existing training data: $(x_1, y_1), \ldots, (x_p, y_p)$, where $x_i \in R^n, y_i \in \{1, -1\}$, p is the number of data and n is the dimension of stock market influence variables. When y equals to 1, the stock market goes up; when y equals to -1, the stock market goes down. In the linear analysis, an optimal hyperplane, $(w \cdot x) + b = 0$ can completely separate the sample into two conditions shown as below, where w is the weight vector and b is a bias

$$(w \cdot x) + b \geqslant 1 \rightarrow y_i = +1 \tag{1}$$

$$(w \cdot x) + b \leqslant -1 \rightarrow y_i = -1 \tag{2}$$

We have Eq. (3) below by multiplying Eqs. (1) and (2) by y_i

$$y_i(w \cdot x + b) \geqslant 1, \quad \forall i \tag{3}$$

For SVM, finding an optimal hyperplane has to solve the following optimization problem:

$$\min_{w,b} \frac{1}{2}\|W\|^2 \tag{4}$$

Constraint:

$$y_i(w \cdot x + b) \geqslant 1$$

In the linear separation, it is a typical quadratic programming problem. Lagrange formula can be used to find the solution, where α is a Lagrange multiplier

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{p} \alpha_i [y_i(w \cdot x_i + b) - 1] \tag{5}$$

By differentiating with respect to w and b, the followings are obtained.

$$\frac{\partial}{\partial w} L_p = 0, \quad w = \sum_{i=1}^{m} \alpha_i y_i x_i \tag{6}$$

$$\frac{\partial}{\partial b} L_p = 0, \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \tag{7}$$

In the linear analysis, by introducing Eqs. (6) and (7) to Eq. (5), the original problem can be considered as a dual problem. To find the optimal solution, the approach is

$$\max W(\alpha) = \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{8}$$

Constraint:

$$\sum_{i=1}^{p} \alpha_i y_i = 0, \alpha_i > 0, \quad i = 1, 2, \ldots, p$$

The classification formula applied to forecast the next day stock market dynamism can be obtained as below.

$$f(x) = \mathrm{Sign}\left(\sum_{i=1}^{p} y_i \alpha_i (x \cdot x_i) + y_i - w \cdot x_i\right) \tag{9}$$

2.2. Linear separation for nonseparable data

In many cases, data cannot be separated by linear separation. But, a linear separation hyperplane with minimum errors can be explored. The idea is to introduce a non-negative slack variables ξ_i, $i = 1, \ldots, m$, so that we can have follows:

$$(w \cdot x) + b \geqslant +1 - \xi_i \rightarrow y_i = +1 \tag{10}$$

$$(w \cdot x) + b \leqslant -1 + \xi_i \rightarrow y_i = -1 \tag{11}$$

In order to find such a hyperplane and minimize number of training errors, it has to solve the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{p} \xi_i \tag{12}$$

Constraint:

$$y_i(w \cdot x + b) \geqslant 1 - \xi, \xi \geqslant 0, \quad i = 1, \ldots, p$$

In the optimization linear separation for nonseparable data, Lagrange formula also can be used to find the solution

$$\max W(\alpha) = \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{13}$$

where α_i is a Lagrange multiplier.
Constraint:

$$\sum_{i=1}^{p} \alpha_i y_i = 0, 0 \leqslant \alpha_i \leqslant C, \quad i = 1, 2, \ldots, p$$

Here, the classification formula applied to forecast the next day stock market dynamism is the same as Eq. (9).

2.3. Non-linear separation

The non-linear separation question can be solved using a mapping function Φ, which called kernel function, can map input space of training data into a higher-dimensional feature space. The inner product is replaced by kernel function as Eq. (14)

$$(\Phi(x_i)\Phi(x_j)) := k(x_i, x_j) \tag{14}$$

In the non-linear separation, it is similar to quadratic programming problem. Lagrange formula can be used to find the solution as below

$$\max W(\alpha) = \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i \alpha_j y_i y_j k(x_i x_j) \tag{15}$$

Constraint:

$$\sum_{i=1}^{p} \alpha_i y_i = 0, 0 \leqslant \alpha_i \leqslant C, \quad i = 1, 2, \ldots, p$$

Any functions that meet Mercer's condition can be used as kernel functions. We adopt Radial kernel function below as kernel function of SVM (Cristianini & Taylor, 2000)

$$K(x_i, x_j) = \exp\left(-\frac{1}{10}\|x_i - x_j\|^2\right) \tag{16}$$

3. Genetic algorithm (GA)

GA is an efficient and better search method in the broad sense. With the simulation of biological evolution phenomenon, the parameter with higher fitness function value is left. Also, with mechanisms of crossover and mutation, issue of partial minimization during search is avoided and search time is shortened. Evolution process of genetic algorithm is shown as followings:

(1) *Initialization*: Each chromosome is created by randomly obtaining the diversity solutions.
(2) *Selection*: Select chromosome by evaluating the fitness value of each chromosome for searching near-optimization solution. The chromosomes with better fitness values are selected into the recombination pool using the roulette wheel or the tournament selection method as shown in Fig. 1.
(3) *Crossover*: Here, genes between two parent chromosomes are exchanged to obtain new offspring to attempt to get better solutions. Exchanging methods of genes crossover includes one point crossover, two point crossover, or homologue crossover between two chromosomes as shown in Fig. 2.

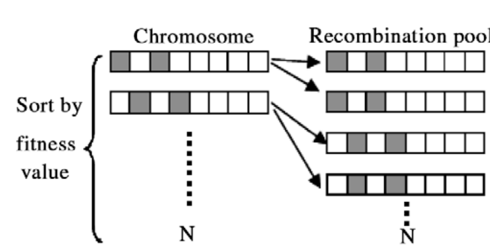

Fig. 1. Selection method in genetic algorithm.

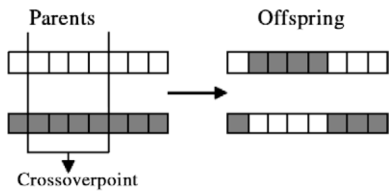

Fig. 2. Two point crossover method.

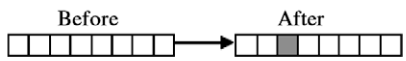

Fig. 3. Genetic mutation operation.

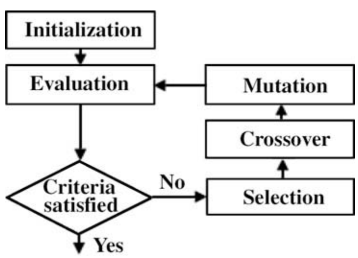

Fig. 4. Evolutionary cycle.

(4) *Mutation*: Using mutation to change genes code from 0 to 1 or vice verse can differ from population as a stochastic perturbation as shown in Fig. 3.
(5) *Evolutionary cycle*: Here, termination criteria is used to determine if the process should terminate or the process should go to step 2 repeatedly with next generation as shown in Fig. 4.

4. GA-based dynamic fuzzy model with SVM

The dynamic fuzzy model, model optimization with genetic algorithm, and system architecture of the proposed model methods are described as follows.

4.1. Dynamic fuzzy model

To increase the forecast accuracy rate, the input variables shall cover sufficient influence factors and have to precisely reflect the influential degree of each influence factor that changes with time. Due to the difference of each kind of factor, in many studies, relationships among different kinds of factors and stock market dynamism are discussed separately. To consider three kinds of factors that affect stock market, technical indicators in stock market, macroeconomic variables and technical indicators in futures market at the same time, and to adjust the influential degrees of input variables changing with time, we propose the dynamically adjustable fuzzy model to solve the issue as shown in Fig. 5. Each influential degree of input variable, ID, is determined by the adjusted scaled index, SI, and the confidence of the variable changing with time,

```
Step 1.Collect dataset of input X and output Y.
Step 2.Scale the dataset of input variable
    For t = 1 to total of dataset records
      For k = 1 to total of dataset attributions
        SI_k,t = 2 * ( X_k,t - min(X_k) ) / ( max(X_k) - min(X_k) ) -1;
Step 3.Build fuzzy model of input membership function
    For k = 1 to total of dataset attributions
      {Set a1, a, b, b1, range for variable k;
      t = t % range;
      If (t < a1) return 0;
      Else If ( t ≥ a1 && t < a )   return (t - a1) / (a - a1);
      Else If ( t ≥ a && t < b )   return 1;
      Else If ( t ≥ b && t ≤ b1 ) return (b1 - t) / (b1 - b);
      Else If ( t > b1)  return 0; }
Step 4.Perform fuzzy model for each input variable with time of
    influential degree
    For t=1 to total of dataset records
      For k =1 to total of dataset attributions
        ID_k,t = SI_k,t * μ_Ak (t);
```

Fig. 5. Fuzzy model process.

$\mu_{Ak}(t)$, as shown in Eq. (17), where t is moment, k is an input factor

$$\mathrm{ID}_{k,t} = \mathrm{SI}_{k,t} * \mu_{Ak}(t) \tag{17}$$

Due to the different field ranges of input variables, to increase the accuracy, we adopt linear transference to adjust the variable to the range of [−1, 1] as shown in Eq. (18)

$$\mathrm{SI}_{k,t} = 2 \cdot \frac{x_{k,t} - \min(x_k)}{\max(x_k) - \min(x_k)} - 1 \tag{18}$$

To reflect the changes of variables affecting the stock market with time, in the condition of considering the complexity of calculation and actual improvement of forecast accuracy, we compare the trapezoid, triangle and Gaussian membership functions and adopt the trapezoid membership function to simulate the changes. We adjust the membership function $\mu_{Ak}(t)$ of time, t, as shown in Eq. (19).

$$\mu_{Ak}(t) = \begin{cases} 0, & t < a_1 \\ \frac{t-a_1}{a-a_1}, & a_1 \leqslant t < a \\ 1, & a \leqslant t < b \\ \frac{b_1-t}{b_1-b}, & b \leqslant t \leqslant b_1 \\ 0, & t > b_1 \end{cases} \tag{19}$$

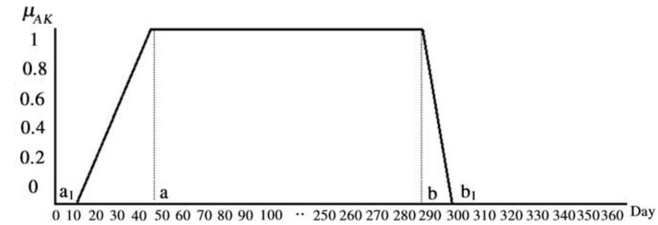

Fig. 6. Initial membership function in the fuzzy model for the influential degree of factor GDP from 2004/1 to 2004/12.

Each influence variable has its independent fuzzy model. The shape of each fuzzy model is determined by the four parameters, a_1, a, b, b_1 which will be dynamically adjusted in accordance with the fitness function from GA to properly express the influential degree of the variable.

For example, GDP (gross domestic product) is a macroeconomic variable. As it is published once a year, the maximum value of x axis in the initialization fuzzy model is 360. The initial a_1, a, b, b_1 are generated randomly and then adjusted in accordance with the fitness function value in GA as shown in Fig. 6.

After GA determines the parameters of each fuzzy model, SVM will serve to locate the relationships among influence variables (e.g. technical indicators in stock market, macroeconomic variables and technical indicators in futures market) and stock market dynamism.

4.2. Model optimization with genetic algorithm (GA)

Because of the stock market dynamism nature, influential degree of a factor changes with time and the model has to be dynamically adjusted. With GA, we can locate the approximate optimal solution and the better parameters of model in a certain period of time.

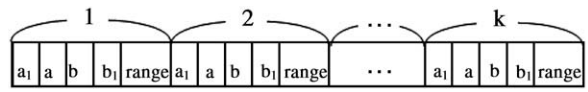

Fig. 7. Chromosome design.

The design of chromosome is shown as in Fig. 7. Each input variable includes five elements, a_1, a, b, b_1 and range. The elements a_1, a, b, b_1 determine the shape of the fuzzy model, which represents the changes of the influential degree of each variable. The range represents the published cycle of the variable. Due to the difference of each problem, GA is unable to obtain the same search results with fixed detailed setup. Taking time and accuracy rate into consideration, crossover in the experiment adopts two-point crossover with the probability of 0.8 and mutation rate of 0.02.

The forecast accuracy rate in the study is the criteria to evaluate the forecast model. Therefore, design of evaluation (fitness) function is the optimization of accuracy rate produced by SVM in the training period. GA locates the approximate optimal solution. Within 100 generations with change rate below 0.02, the evolution stops. The located parameters shall serve to establish the better model for a certain period of time.

4.3. System architecture of the proposed approach

The proposed integrated architecture is shown in Fig. 8. The explanation is as follows:

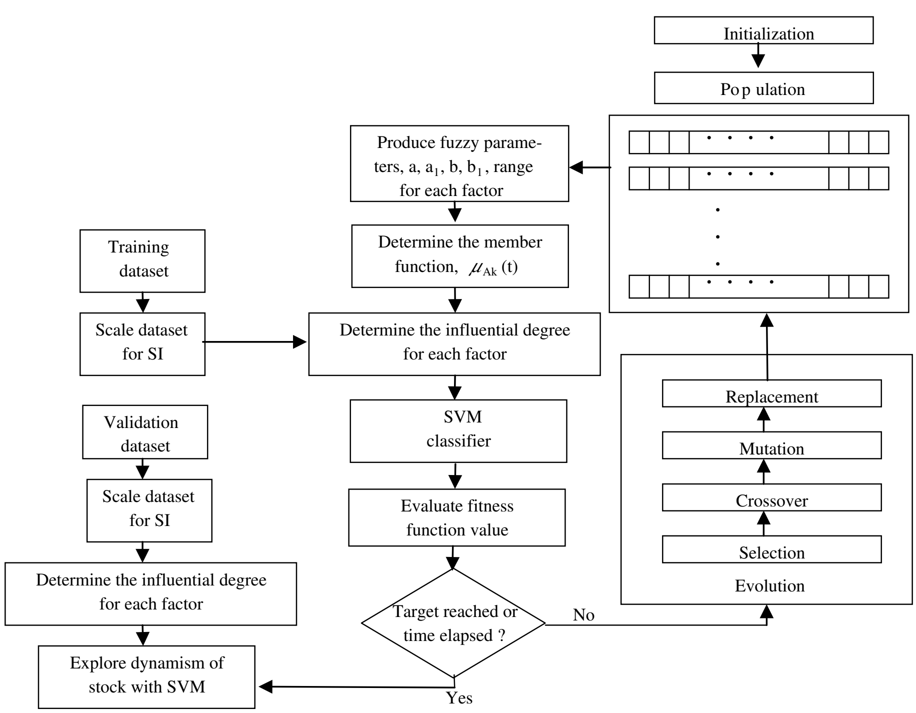

Fig. 8. The architecture of proposed model.

(1) The data is scaled by linear transference to range of [−1, 1].
(2) The parameters needed in fuzzy model for each factor are generated randomly first. Those parameters are a, a_1, b, b_1, and the range.
(3) Define the membership function for each factor using fuzzy model parameters.
(4) Determine the influential degree of each factor that can be used to evaluate fitness function value.
(5) SVM is applied to learning data with factor influential degrees to produce accuracy rate used as fitness function value.
(6) With the accuracy rate used as fitness function value, GA determines whether to conduct evolution or whether the target is reached.
(7) In the event of evolution, after selection, crossover, and mutation, parameters of fuzzy model for next generation are generated. Then, go to step 3.
(8) When the termination condition is reached, SVM is finally applied again to explore the stock market dynamism by employing the optimized parameters generated in previous steps.

5. Experiment

5.1. Experiment descriptions

We integrate fuzzy theory, GA and SVM to explore stock market dynamism, targeting the stock market in Taiwan. The input variables in the study include a total of 61 variables, including technical indicators in stock market, technical indicators in futures market and the macroeconomic variables in Taiwan (Dickinson, 2000; Min & Najand, 1999). The influence variables for both stock and futures market include On balance volume (OBV), Demand index (DI), Momentum (MTM), Relative strength index (RSI), Moving average convergence and divergence (MACD), Total amount per weighted stock price index (TAPI), Psychological line (PSY), Advance decline ratio (ADR), Williams (WMS), BIAS, Oscillator (OSC), Moving average (MA), K line (K), D line (D), Perform criteria (PC), Autoregressive (AR), Different (DIF), Consistency ratio (CR), Relative strength volume (RSV) and Exponential moving average (EMA).

Macroeconomic variables include Annual change in wholesale price index (WPC), Annual change in export price index (EPC), Annual change in industrial production index (IPC), Annual change in employees on payrolls (EMPC), Gross national production (GNP), Approved outward investment by industry (AOI), Gross domestic product (GDP), Import by key trading partners (IKT), Export by key trading partners (EKT), Long term interest rate (LT), Consumer price index (CPI), Government consumption (GC), MFGs' New Orders (MNO), Average monthly working hours (AMWH), Average monthly wages and salaries(AMWS), Bank clearings (BC), Manufacturing sales (MS), Quantum of domestic traffic (QDT), Monetary aggregate (M1B), Term architecture of interest rate (TS) and Short term interest rate (ST).

The original data of stock and futures market in Taiwan are retrieved from Taiwan Stock Exchange Corporation while macroeconomic variables are from Ministry of Economic Affairs, ROC. The historical data are for two years from January, 2003 to December, 2004 for a total of 714 pieces of data. Among which, 378 pieces go up while the rest go down.

As stock market is a complicated and volatile system, in order to express the changes of influential degree of each factor effectively, we apply two methods, multiperiod and two-period. In the multiperiod method, data of every fifty days serve as one set of training data to obtain parameters of the proposed integrated model. Forecast of the next day is made with such data as shown in Fig. 9. In the two-period method, all the data are divided into two parts. One part serves as training while another part for verification as shown in Fig. 10.

Each indicator has its independent fuzzy model. Based on the accuracy rate during past year, GA simulates the changes of influential degree. In the experiment results, most of the dynamic fuzzy models converge after 100 generations.

Take GDP as an example. From 2004/1 to 2004/12, the adjusted dynamic fuzzy model is shown in Fig. 11. In GDP, day 'a_1' represents the beginning of GDP influential degree while day 'a' refers to the climax of the influential degree. Day 'a_1' does not start from the beginning and it takes 30 days from $\mu_{\mathrm{Ak}} = 0$ to $\mu_{\mathrm{Ak}} = 1$. It means that the influential degree of factor GDP does not reflect immediately and it increases gradually. Meanwhile, it takes as long as 245 days from day 'a' to day 'b'. It means that the influential degree of GDP in the market can last a long period of time. Dynamic fuzzy models located by GA differ in different time periods. This is resulted from nature of stock market

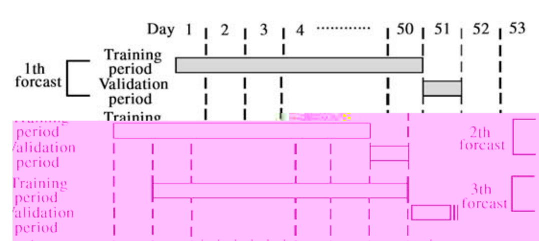

Fig. 9. Strategy for multiperiod stock market movement forecast.

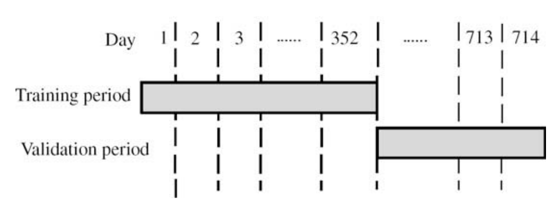

Fig. 10. Strategy for two-period stock market movement forecast.

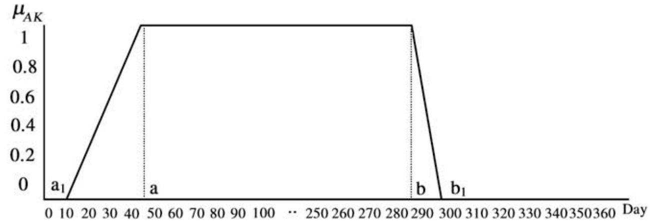

Fig. 11. Adjusted membership function for influential degree of factor GDP from 2004/1 to 2004/12.

changes leading to the same variable having different influential degree in the different periods.

5.2. Experiment results and comparisons

After termination condition in GA is reached, SVM is used again to explore the stock dynamism. In the experiment, Radial function is adopted as kernel function.

The results of performance comparisons among proposed and other forecast methods (buy_and_hold, discriminant analysis and artificial neural network) with different forecast models and input variables are shown in Table 1. Possible input variables are technical indicators in stock market, technical indicators in futures market, and macroeconomic variables. Input variables with mark (v) serve as input variables of the model. The buy_and_hold method poses the worst forecast performance at accuracy rate merely 50%. As the method is derived from random walk hypothesis, all the supposed information is reflected in the market transaction prices. In other words, dynamism is random and unpredictable.

The ANN with fuzzy model has average forecast accuracy rate as high as 59.25%. With GA, it will have average forecast accuracy as high as 63.25%. In the experiment of two-period with SVM, the average accuracy rate is 62.3%, and that in multiperiod experiment is 64.6%.

Table 1
Forecasting performance of different input variables with various forecast models

Method	Input variable			Accuracy	
	Technical indicators in stock market (20 variables)	Macroeconomic (21 variables)	Technical indicators in futures market (20 variables)	Hit ratio (%)	
				Multiperiod	Two-period
Proposed model	v	v	v	79	73
	v	v		77	70
		v	v	74	71
	v		v	70	69
			Average	75	70.75
SVM with fuzzy model	v	v	v	73	73
	v	v		72	70
		v	v	70	69
	v		v	69	69
			Average	71	70.25
SVM	v			67	65
		v		69	66
			v	58	56
			Average	64.6	62.3
ANN with fuzzy model and GA	v	v	v		63
	v	v			63
		v	v		64
	v		v		63
			Average		63.25
ANN with fuzzy model	v	v	v		61
	v	v			62
		v	v		58
	v		v		56
			Average		59.25
ANN	v				61
		v			62
			v		53
			Average		58.6
Discriminant analysis	v				53
		v			54
			v		51
			Average		52.6
Buy and hold			Average		50

With any influence factor as input variable, SVM poses average accuracy rate of 63.4%, outperforming 58.6% of ANN and 52.6% of discriminant analysis. This is because that SVM locates the learning deviation with generalization theories, instead of reducing training deviation in ANN. Thus, overfitting issue from too high variable dimension can be avoided. Such a feature makes SVM perform better in stock market dynamic exploration.

For SVM with fuzzy model in the study, with experiments of two-period and multiperiod methods, the average forecast accuracy rates are 70.25% and 71%, respectively. With combining GA into dynamically adjusted fuzzy model, the average forecast accuracy rates rise to 70.75% and 75%, respectively.

The forecast model in this study boasts the significant forecast performance. In the experiment results, when the study includes three kinds of input variables in proposed integrated model, the forecast accuracy rate is higher than that from single or two kinds of input variables.

This shows that more input variables help integrated forecast model reflect the relationship among stock market fluctuations. However, when ANN includes two or three kinds of input variables, the difference between its average accuracy rate (59.25%) and that with single input variable (58.6%) is merely 0.65%. This might be due to overfitting of ANN from too many noises. In this study, GA dynamically adjusts the influential degree of each variable to reflect changes of the market. Without GA, confidence of each variable $\mu_{\mathrm{Ak}}(t)$ is 1. That is, prior to availability of next new values, the influential degree of each variable remains unchanged.

For the proposed model with multiperiod method, GA improves the accuracy rate by 4% (71–75%). But it does not provide significant help in models in two-period method. The increase of average accuracy rate increase is only 0.5% (70.25–70.75%). This is resulted from two-period method being unable to precisely simulate market fluctuations. Therefore, GA and multiperiod method bring higher accuracy rate.

6. Conclusions

We propose a model integrating fuzzy theory, GA and SVM to explore movements in stock market in Taiwan. The new dynamic fuzzy model not only effectively simulates market volatility but also covers influence factors of different features. The integration of high dimension variables, with features of SVM, increases the forecast accuracy rate. The integrated forecast model in this study can serve as a valuable evaluation reference for researches on internal mechanism of stock market.

The main findings are as below.

(1) With GA, the performance is better.
(2) With SVM, performance of multiperiod is better than that of two-period method.
(3) SVM outperforms ANN and discriminant analysis.
(4) Fuzzy model does improve the accuracy.
(5) With SVM, experiment with three kinds of input variables outperforms those with one or two kinds of input variables.

In this study, our main purpose is to design a forecast model to integrate various factors to deal the dynamism of stock market.

In the future, the study can be extended as follows:

(1) Due to the complicate and dynamic nature of stock market, each factor may interact with others in every moment. In the future, factors interaction should be studied in more details.
(2) There are some techniques used to reduce the input dimensions to reduce the complexity of the model. This could be a possible attempt to improve the proposed model in the future.
(3) The parameters used in kernel function of SVM could also be optimized. It should also be a good approach to improve the overall performance.

References

Armano, G., Murru, A., & Roli, F. (2002). Stock market prediction by a mixture of genetic-neural experts. *International Journal of Pattern Recognition and Artificial Intelligence, 16*, 501–526.

Armano, G., Marchesi, M., & Murr, A. (2005). A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences, 170*, 3–33.

Black, A. J., & Mcmillan, D. G. (2004). Non-linear predictability of value and growth stocks and economic activity. *Journal of Business Finance and Accounting, 31*, 439–474.

Cristianini, N., & Taylor, J. S. (2000). *An introduction to support vector machines*. New York: Cambridge University.

Davis, L. (1994). Genetic algorithms and financial applications. In G. J. Deboeck (Ed.), *Trading on the edge* (pp. 133–147). New York: Wiley.

Dickinson, D. G. (2000). Stock market integration and macroeconomic fundamentals: An empirical analysis. *Applied Financial Economics, 10*, 261–276.

Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimization for support vector machine. *Expert Systems with Applications, 31*, 231–240.

Huang, W., Nakamori, Y., & Wang, S. Y. (2004). *Forecasting stock market movement direction with support vector machine. Computers and Operations Research* (Vol. 32). Elsevier SciOxford, pp. 2513–2522.

Huang, K., & Yu, H. K. (2005). A Type 2 fuzzy time series model for stock index forecasting. *Physica A: Statistical Mechanics and its Applications, 353*, 445–462.

Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications, 19*, 125–132.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*, 307–319.

Kuo, R. J., Chen, C. H., & Hwang, Y. C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems, 118*, 21–45.

Kuo, R. J. (1998). A decision support system for the stock market through integration of fuzzy neural networks and fuzzy delphi. *Applied Artificial Intelligence, 12*, 501–520.

Lo, A., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithm, statistical inference, and empirical implementation. *Journal of Finance, 55*, 1705–1765.

Lo, A. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, 30*, 15–44.

Matilla-Garcia, G., & Arguellu, C. (2005). A hybrid approach based on neural networks and genetic algorithm to the study of profitability in the spanish stock market. *Applied Economics Letter, 12*, 303–308.

Min, J. H., & Najand, M. (1999). A further investigation of the lead-lag relationship between the spot market and stock index future: Early evidence from Korea. *Journal of Futures Markets, 19*, 217–232.

Oh, K. J., & Kim, K. J. (2002). Analyzing stock market tick data using piecewise nonlinear model. *Expert Systems with Application, 22*, 249–255.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega, 33*, 497–505.

Schwert, G. W. (1989). Why does stock market volatility change over time. *Journal of Finance, 44*, 1115–1167.

Wang, Y. F. (2003). Mining stock price using fuzzy rough set system. *Expert Systems with Applications, 24*, 13–23.

Wong, F. S., & Wang, P. Z. (1991). A stock selection strategy using fuzzy neural networks. *Neurocomputing, 2*, 233–242.

Yu, L., Wang, S., & Lai, K. K. (2005). Mining stock market tendency using GA-based support vector machines. *Lecture Notes in Computer Science, 3828*, 336–345.