Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computers & Operations Research 32 (2005) 2513–2522

computers & operations research

www.elsevier.com/locate/dsw

# Forecasting stock market movement direction with support vector machine

Wei Huang[a,b], Yoshiteru Nakamori[a], Shou-Yang Wang[b,*,1]

[a]*School of Knowledge Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292, Japan*

[b]*Institute of Systems Science, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100080, China*

Accepted 29 March 2004

**Abstract**

Support vector machine (SVM) is a very specific type of learning algorithms characterized by the capacity control of the decision function, the use of the kernel functions and the sparsity of the solution. In this paper, we investigate the predictability of financial movement direction with SVM by forecasting the weekly movement direction of NIKKEI 225 index. To evaluate the forecasting ability of SVM, we compare its performance with those of Linear Discriminant Analysis, Quadratic Discriminant Analysis and Elman Backpropagation Neural Networks. The experiment results show that SVM outperforms the other classification methods. Further, we propose a combining model by integrating SVM with the other classification methods. The combining model performs best among all the forecasting methods.


*Keywords:* Support vector machine; Forecasting; Multivariate classification

## 1. Introduction

The financial market is a complex, evolutionary, and non-linear dynamical system [1]. The field of financial forecasting is characterized by data intensity, noise, non-stationary, unstructured nature, high degree of uncertainty, and hidden relationships [2]. Many factors interact in finance including political events, general economic conditions, and traders' expectations. Therefore, predicting finance market price movements is quite difficult. Increasingly, according to academic investigations, movements

* Corresponding author. Tel.: +86-10-62651381; fax: +86-10-62568364.
*E-mail address:* swang@mail.iss.ac.cn (S.-Y. Wang).
[1] This author is also with University of Tsukuba in Japan and Hunan University in China.

0305-0548/$ - see front matter 
doi:10.1016/j.cor.2004.03.016

in market prices are not random. Rather, they behave in a highly non-linear, dynamic manner. The standard random walk assumption of futures prices may merely be a veil of randomness that shrouds a noisy non-linear process [3–5].

Support vector machine (SVM) is a very specific type of learning algorithms characterized by the capacity control of the decision function, the use of the kernel functions and the sparsity of the solution [6–8]. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance. Another key property of SVM is that training SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of SVM is always unique and globally optimal, unlike neural networks training which requires nonlinear optimization with the danger of getting stuck at local minima.

Some applications of SVM to financial forecasting problems have been reported recently [9–13]. In most cases, the degree of accuracy and the acceptability of certain forecasts are measured by the estimates' deviations from the observed values. For the practitioners in financial market, forecasting methods based on minimizing forecast error may not be adequate to meet their objectives. In other words, trading driven by a certain forecast with a small forecast error may not be as profitable as trading guided by an accurate prediction of the direction of movement.

The main goal of this study is to explore the predictability of financial market movement direction with SVM. The remainder of this paper is organized as follows. In Section 2, we introduce the basic theory of SVM. Section 3 gives the experiment scheme. The experiment results are shown in Section 4. Some conclusions are drawn in Section 5.

## 2. Theory of SVM in classification

In this section, we present a basic theory of the support vector machine model. For a detailed introduction to the subject, please refer to [14,15]. Let $D$ be the smallest radius of the sphere that contains the data (example vectors). The points on either side of the separating hyperplane have distances to the hyperplane. The smallest distance is called the margin of separation. The hyperplane is called optimal separating hyperplane (OSH), if the margin is maximized. Let $q$ be the margin of the optimal hyperplane. The points that are distance $q$ away from the OSH are called the support vectors.

Consider the problem of separating the set of training vector belonging to two separate classes, $G = \{(x_i, y_i), i = 1, 2, \ldots, N\}$ with a hyperplane $w^{\mathrm{T}}\varphi(x) + b = 0$ ($x_i \in R^n$ is the $i$th input vector, $y_i \in \{-1, 1\}$ is known binary target), the original SVM classifier satisfies the following conditions:

$$w^{\mathrm{T}}\varphi(x_i) + b \geqslant 1 \quad \textit{if} \;\; y_i = 1, \tag{1}$$

$$w^{\mathrm{T}}\varphi(x_i) + b \leqslant -1 \quad \textit{if} \;\; y_i = -1, \tag{2}$$

or equivalently,

$$y_i[w^{\mathrm{T}}\varphi(x_i) + b] \geqslant 1 \quad i = 1, 2, \ldots, N, \tag{3}$$

where $\varphi : R^n \to R^m$ is the feature map mapping the input space to a usually high dimensional feature space where the data points become linearly separable.

The distance of a point $x_i$ from the hyperplane is

$$d(x_i, w, b) = \frac{|w^{\mathrm{T}}\varphi(x_i) + b|}{\|w\|^2}. \tag{4}$$

The margin is $2/\|w\|$ according to its definition. Hence, we can find the hyperplane that optimally separates the data by solving the optimization problem:

$$\min \phi(w) = \frac{1}{2}\|w\|^2 \tag{5}$$

under the constraints of Eq. (3).

The solution to the above optimization problem is given by the saddle point of the Lagrange function

$$L_{\mathrm{P1}} = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i [y_i(w^{\mathrm{T}}\varphi(x_i) + b) - 1] \tag{6}$$

under the constraints of Eq. (3), where $\alpha_i$ are the nonnegative Lagrange multipliers.

So far the discussion is restricted to the case where the training data is separable. To generalize the problem to the non-separable case, slack variable $\xi_i$ is introduced such that

$$y_i[w^{\mathrm{T}}\varphi(x_i) + b] \geqslant 1 - \xi_i \quad (\xi_i \geqslant 0 \;\; i = 1, 2, \ldots, N). \tag{7}$$

Thus, for an error to occur, the corresponding $\xi_i$ must exceed unity, so $\sum_{i=1}^{N} \xi_i$ is an upper bound on the number of training errors. Hence, a natural way to assign an extra cost for errors is to change the objective function from Eq. (5) to

$$\min \phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i \tag{8}$$

under the constraints of Eq. (7), where $C$ is a positive constant parameter used to control the tradeoff between the training error and the margin. In this paper, we choose $C = 50$ based on our experiment experiences. Similarly, solve the optimal problem by minimizing its Lagrange function

$$L_{\mathrm{P2}} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i[y_i(w^{\mathrm{T}}\varphi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^{N} \mu_i \xi_i \tag{9}$$

under the constraints of Eq. (7), where $\alpha_i, \mu_i$ are the non-negative Lagrange multipliers.

The Karush–Kuhn–Tucker (KKT) conditions [16] for the primal problem are

$$\frac{\partial L_{\mathrm{P2}}}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i) = 0, \tag{10}$$

$$\frac{\partial L_{\mathrm{P2}}}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0, \tag{11}$$

$$\frac{\partial L_{\mathrm{P2}}}{\partial \xi_i} = C - \alpha_i - \xi_i = 0, \tag{12}$$

$$y_i[w^{\mathrm{T}}\varphi(x_i) + b] \geqslant 1 - \xi_i, \tag{13}$$

$$\xi_i \geqslant 0, \tag{14}$$

$$\alpha_i \geqslant 0, \tag{15}$$

$$\mu \geqslant 0, \tag{16}$$

$$\alpha_i[y_i(w^{\mathrm{T}}\varphi(x_i) + b) - 1 + \xi_i] = 0, \tag{17}$$

$$\mu_i \xi_i = 0. \tag{18}$$

Hence,

$$w = \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i). \tag{19}$$

We can use the KKT complementarity conditions, Eqs. (17) and (18), to determine $b$. Note that Eq. (12) combined with Eq. (18) shows that $\xi_j = 0$ if $\alpha_j < C$. Thus we can simply take any training data for which $0 < \alpha_j < C$ to use Eq. (17) (with $\xi_j = 0$) to compute $b$.

$$b = y_j - w^{\mathrm{T}}\varphi(x_j). \tag{20}$$

It is numerically reasonable to take the mean value of all $b$ resulting from such computing. Hence,

$$b = \frac{1}{N_{\mathrm{s}}} \sum_{0<\alpha_j<C} [(y_j - w^{\mathrm{T}}\varphi(x_j)], \tag{21}$$

where $N_{\mathrm{s}}$ is the number of the support vectors.

For a new data $x$, the classification function is then given by

$$f(x) = \mathrm{Sign}(w^{\mathrm{T}}\varphi(x) + b). \tag{22}$$

Substituting Eqs. (19) and (21) into Eq. (22), we get the final classification function

$$f(x) = \mathrm{Sign}\left(\sum_{i=1}^{N} \alpha_i y_i \varphi(x_i)^{\mathrm{T}}\varphi(x) + \frac{1}{N_{\mathrm{s}}} \sum_{0<\alpha_j<C} \left(y_j - \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i)^{\mathrm{T}}\varphi(x_j)\right)\right). \tag{23}$$

If there is a kernel function such that $K(x_i, x_j) = \varphi(x_i)^{\mathrm{T}}\varphi(x_j)$, it is usually unnecessary to explicitly know what $\varphi(x)$ is, and we only need to work with a kernel function in the training algorithm. Therefore, the non-linear classification function is

$$f(x) = \mathrm{Sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + \frac{1}{N_{\mathrm{s}}} \sum_{0<\alpha_j<C} \left(y_j - \sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j)\right)\right). \tag{24}$$

Any function satisfying Mercer's condition [17] can be used as the kernel function. In this investigation, the radial kernel $K(s,t) = \exp(-\frac{1}{10}\|s - t\|^2)$ is used as the kernel function of the SVM because the radial kernel tends to give good performance under general smoothness assumptions. Consequently, it is especially useful if no additional knowledge of the data is available [18].

## 3. Experiment design

In our empirical analysis, we set out to examine the weekly changes of the NIKKEI 225 Index. The NIKKEI 225 Index is calculated and disseminated by Nihon Keizai Shinbum Inc. It measures the composite price performance of 225 highly capitalized stocks trading on the Tokyo Stock Exchange (TSE), representing a broad cross-section of Japanese industries. Trading in the index has gained unprecedented popularity in major financial markets around the world. Futures and options contracts on the NIKKEI 225 Index are currently traded on the Singapore International Monetary Exchange Ltd (SIMEX), the Osaka Securities Exchange and the Chicago Mercantile Exchange. The increasing diversity of financial instruments related to the NIKKEI 225 Index has broadened the dimension of global investment opportunity for both individual and institutional investors. There are two basic reasons for the success of these index trading vehicles. First, they provide an effective means for investors to hedge against potential market risks. Second, they create new profit making opportunities for market speculators and arbitrageurs. Therefore, it has profound implications and significance for researchers and practitioners alike to accurately forecast the movement direction of NIKKEI 225 Index.

### 3.1. Model inputs selection

Most of the previous researchers have employed multivariate input. Several studies have examined the cross-sectional relationship between stock index and macroeconomic variables. The potential macroeconomic input variables which are used by the forecasting models include term structure of interest rates (TS), short-term interest rate (ST), long-term interest rate (LT), consumer price index (CPI), industrial production (IP), government consumption (GC), private consumption (PC), gross national product (GNP) and gross domestic product (GDP) [19–27]. However, Japanese interest rate has dropped down to almost zero since 1990. Other macroeconomic variables weekly data are not available for our study.

Japanese consumption capacity is limited in the domestic market. The economy growth has a close relationship with Japanese export. The largest export target for Japan is the United States of America (USA), which is the leading economy in the world. Therefore, the economic condition of USA influences Japan economy, which is well represented by the NIKKEI 225 Index. As the NIKKEI 225 Index to Japan economy, the S& P 500 Index is a well-known indicator of the economic condition in USA. Hence, the S& P 500 Index is selected as model input. Another import factor that affects the Japanese export is the exchange rate of US Dollars against Japanese Yen (JPY), which is also selected as model input. The prediction model can be written as the following function:

$$Direction_t = F(S_{t-1}^{\mathrm{S\&P500}}, S_{t-1}^{\mathrm{JPY}}), \tag{25}$$

where $S_{t-1}^{\mathrm{S\&P500}}$ and $S_{t-1}^{\mathrm{JPY}}$ are first order difference natural logarithmic transformation to the raw S& P 500 index and JPY at time $t-1$, respectively. Such transformations implement an effective detrending of the original time series. $Direction_t$ is a categorical variable to indicate the movement direction

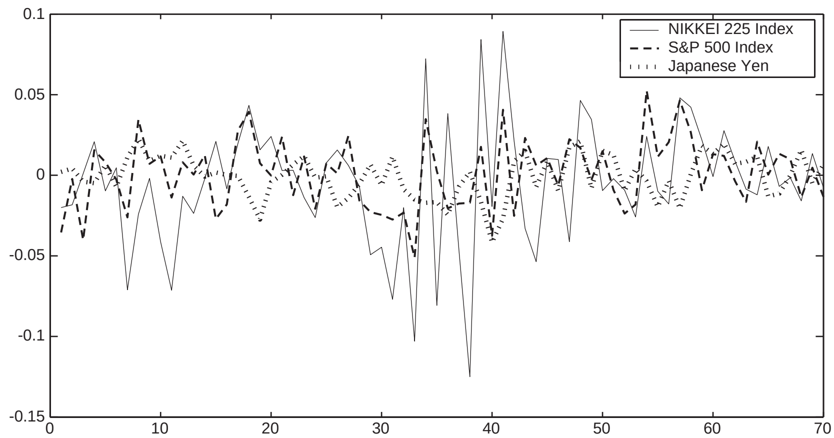


Fig. 1. First-order difference natural logarithmic weekly prices of NIKKEI 225 Index, S& P 500 Index and Japanese Yen (70 observations of the period from January 3, 1990 to May 8, 1991).

of NIKKEI 225 Index at time $t$. If NIKKEI 225 Index at time $t$ is larger than that at time $t-1$, $Direction_t$ is 1. Otherwise, $Direction_t$ is $-1$.

The above model inputs selection is only based on a macroeconomic analysis. As shown in Fig. 1, the behaviors of the NIKKEI 225 Index, the S& P 500 Index and Japanese Yen are very complex. It is impossible to give an explicit formula to describe the underlying relationship between them.

### 3.2. Data collection

We obtain the historical data from the finance section of Yahoo and the Pacific Exchange Rate Service provided by Professor Werner Antweiler, University of British Columbia, Vancouver, Canada, respectively. The whole data set covers the period from January 1, 1990 to December 31, 2002, a total of 676 pairs of observations. The data set is divided into two parts. The first part (640 pairs of observations) is used to determine the specifications of the models and parameters. The second part (36 pairs of observations) is reserved for out-of-sample evaluation and comparison of performances among various forecasting models.

### 3.3. Comparisons with other forecasting methods

To evaluate the forecasting ability of SVM, we use the random walk model (RW) as a benchmark for comparison. RW is a one-step-ahead forecasting method, since it uses the current actual value to predict the future value as follows:

$$\hat{y}_{t+1} = y_t, \tag{26}$$

where $y_t$ is the actual value in the current period $t$ and $\hat{y}_{t+1}$ is the predicted value in the next period.

We also compare the SVM's forecasting performance with that of linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and elman backpropagation neural networks (EBNN).

LDA can handle the case in which the within-class frequencies are unequal and its performance has been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set, thereby guaranteeing maximal separability. QDA is similar to LDA, only dropping the assumption of equal covariance matrices. Therefore, the boundary between two discrimination regions is allowed to be a quadratic surface (for example, ellipsoid, hyperboloid, etc.) in the maximum likelihood argument with normal distributions. Interested readers should refer to [28] or some other statistical books for a more detailed description. In this paper, we derive a linear discriminant function of the form:

$$L(S_{t-1}^{\text{S\&P500}}, S_{t-1}^{\text{JPY}}) = a_0 + a_1 S_{t-1}^{\text{S\&P500}} + a_2 S_{t-1}^{\text{JPY}} \tag{27}$$

and a quadratic discriminant function of the form:

$$Q(S_{t-1}^{\text{S\&P500}}, S_{t-1}^{\text{JPY}}) = a + \mathbf{P}(S_{t-1}^{\text{S\&P500}}, S_{t-1}^{\text{JPY}})^{\text{T}} + (S_{t-1}^{\text{S\&P500}}, S_{t-1}^{\text{JPY}})\mathbf{T}(S_{t-1}^{\text{S\&P500}}, S_{t-1}^{\text{JPY}})^{\text{T}}, \tag{28}$$

where $a_0, a_1, a_2, a, \mathbf{P}, \mathbf{T}$ are coefficients to be estimated.

Elman Backpropagation Neural Network is a partially recurrent neural network. The connections are mainly feed-forward but also include a set of carefully chosen feedback connections that let the network remember cues from the recent past. The input layer is divided into two parts: the true input units and the context units that hold a copy of the activations of the hidden units from the previous time step. Therefore, network activation produced by past inputs can cycle back and affect the processing of future inputs. For more details about Elman Backpropagation Neural Network, refer to [29,30].

### 3.4. A combining model

Given a task that requires expert knowledge to perform, $k$ experts may be better than one if their individual judgments are appropriately combined. Based on this idea, predictive performance can be improved by combining various methods. Therefore, we propose a combining model by integrating SVM with other classification methods as follows:

$$f_{\text{combine}} = \sum_{i=1}^{k} w_i f_i, \tag{29}$$

where $w_i$ is the weight assigned to classification method $i$, $\sum_{i=1}^{k} w_i = 1$. We would like to determine the weight scheme based on the information from the training phase. Under this strategy, the relative contribution of a forecasting method to the final combined score depends on the in-sample forecasting performance of the learned classifier in the training phase. Conceptually, a well-performed forecasting method should be given a larger weight than the others during the score combination. In the investigation, we adopt the weight scheme as follows:

$$w_i = \frac{a_i}{\sum_{i=1}^{k} a_i}, \tag{30}$$

where $a_i$ is the in-sample performance constructed by forecasting method $i$.

Table 1
Forecasting performance of different classification methods

| Classification method | Hit ratio (%) |
|---|---|
| RW | 50 |
| LDA | 55 |
| QDA | 69 |
| EBNN | 69 |
| SVM | 73 |
| Combining model | 75 |

Table 2
Covariances matrices of input variables when $Direction_t = -1$

| | $S_{t-1}^{\text{JPY}}$ | $S_{t-1}^{\text{S\&P500}}$ |
|---|---|---|
| $S_{t-1}^{\text{JPY}}$ | 0.00015167706 | 0.00002147347 |
| $S_{t-1}^{\text{S\&P500}}$ | 0.00002147347 | 0.00044862762 |

## 4. Experiment results

Each of the forecasting models described in the last section is estimated and validated by in-sample data. The model estimation selection process is then followed by an empirical evaluation which is based on the out-sample data. At this stage, the relative performance of the models is measured by hit ratio. Table 1 shows the experiment results.

RW performs worst, producing only 50% hit ratio. RW assumes not only that all historic information is summarized in the current value, but also that increments–positive or negative—are uncorrelated (random), and balanced, that is, with an expected value equal to zero. In other words, in the long run there are as many positive as negative fluctuations making long term predictions other than the trend impossible.

SVM has the highest forecasting accuracy among the individual forecasting methods. One reason that SVM performs better than the earlier classification methods is that SVM is designed to minimize the structural risk, whereas the previous techniques are usually based on minimization of empirical risk. In other words, SVM seeks to minimize an upper bound of the generalization error rather than minimizing training error. So SVM is usually less vulnerable to the over-fitting problem.

QDA out-performs LDA in term of hit ratio, because LDA assumes that all the classes have equal covariance matrices, which is not consistent with the properties of input variable belonging to different classes as shown in Tables 2 and 3. In fact, the two classes have different covariance matrices. Heteroscedastic models are more appropriate than homoscedastic models.

The integration of SVM and the other forecasting methods improves the forecasting performance. Different classification methods typically have access to different information and therefore produce different forecasting results. Given this, we can combine the individual forecaster's various information sets to produce a single superior information set from which a single superior forecast could be produced.

Table 3
Covariances matrices of input variables when $Direction_t = 1$

| | $S_{t-1}^{JPY}$ | $S_{t-1}^{S\&P500}$ |
|---|---|---|
| $S_{t-1}^{JPY}$ | 0.00018240800 | −0.00002932242 |
| $S_{t-1}^{S\&P500}$ | −0.00002932242 | 0.00044571885 |

## 5. Conclusions

In this paper, we study the use of support vector machines to predict financial movement direction. SVM is a promising type of tool for financial forecasting. As demonstrated in our empirical analysis, SVM is superior to the other individual classification methods in forecasting weekly movement direction of NIKKEI 225 Index. This is a clear message for financial forecasters and traders, which can lead to a capital gain. However, each method has its own strengths and weaknesses. Thus, we propose a combining model by integrating SVM with other classification methods. The weakness of one method can be balanced by the strengths of another by achieving a systematic effect. The combining model performs best among all the forecasting methods.

## Acknowledgements

This work is partially supported by Ministry of Education, Culture, Sports, Science and Technology of Japan; National Natural Science Foundation of China; Chinese Academy of Sciences; Key Laboratory of Management, Decision and Information Systems. The authors would also like to thank the editor and two referees for their very valuable comments and suggestions.

## References

[1] Abu-Mostafa YS, Atiya AF. Introduction to financial forecasting. Applied Intelligence 1996;6:205–13.
[2] Hall JW. Adaptive selection of US stocks with neural nets. In: Deboeck GJ, editor. Trading on the edge: neural, genetic, and fuzzy systems for chaotic financial markets. New York: Wiley; 1994. p. 45–65.
[3] Blank SC. Chaos in futures market? a nonlinear dynamical analysis. Journal of Futures Markets 1991;11:711–28.
[4] DeCoster GP, Labys WC, Mitchell DW. Evidence of chaos in commodity futures prices. Journal of Futures Markets 1992;12:291–305.
[5] Frank M, Stengos T. Measuring the strangeness of gold and silver rates of return. The Review of Economic Studies 1989;56:553–67.
[6] Vapnik VN. Statistical learning theory. New York: Wiley; 1998.
[7] Vapnik VN. An overview of statistical learning theory. IEEE Transactions of Neural Networks 1999;10:988–99.
[8] Cristianini N, Taylor JS. An introduction to support vector machines and other kernel-based learning methods. New York: Cambridge University Press; 2000.
[9] Cao LJ, Tay FEH. Financial forecasting using support vector machines. Neural Computing Applications 2001;10: 184–92.
[10] Tay FEH, Cao LJ. Application of support vector machines in financial time series forecasting. Omega 2001;29: 309–17.
[11] Tay FEH, Cao LJ. A comparative study of saliency analysis and genetic algorithm for feature selection in support vector machines. Intelligent Data Analysis 2001;5:191–209.

[12] Tay FEH, Cao LJ. Improved financial time series forecasting by combining support vector machines with self-organizing feature map. Intelligent Data Analysis 2001;5:339–54.
[13] Tay FEH, Cao LJ. Modified support vector machines in financial time series forecasting. Neurocomputing 2002;48:847–61.
[14] Burges C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 1998;2:121–67.
[15] Evgeniou T, Pontil M, Poggio T. Regularization networks and support vector machines. Advances in Computational Mathematics 2000;13:1–50.
[16] Fletcher R. Practical methods of optimization. New York: Wiley; 1987.
[17] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer; 1995.
[18] Smola AJ. Learning with kernels. PhD Dissertation, GMD, Birlinghoven, Germany, 1998.
[19] Ross S. The arbitrage theory of capital asset pricing. Journal of Economic Theory 1976;13:341–60.
[20] Fama E, Schwert W. Asset returns and inflation. Journal of Financial Economics 1977;5:115–46.
[21] Campbell J. Stock returns and the term structure. Journal of Financial Economics 1987;18:373–99.
[22] Chen N, Roll R, Ross S. Economic forces and the stock market. Journal of Business 1986;59:383–403.
[23] Fama E, French K. Dividend yields and expected stock returns. Journal of Financial Economics 1988;22:3–25.
[24] Fama E, French K. Permanent and temporary components of stock prices. Journal of Political Economy 1988;96: 246–73.
[25] Fama E, French K. The cross-section of expected stock returns. Journal of Finance 1992;47:427–65.
[26] Lakonishok J, Shleifer A, Vishny RW. Contrarian investment, extrapolation, and risk. Journal of Finance 1994;49:1541–78.
[27] Leung MT, Daouk H, Chen AS. Forecasting stock indices: a comparison of classification and level estimation models. International Journal of Forecasting 2000;16:173–90.
[28] Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis. New York: Prentice-Hall; 1995.
[29] Elman JL. Finding structure in time. Cognitive Science 1990;14:179–211.
[30] Kasabov N, Watts M. Spatial-temporal adaptation in evolving fuzzy neural networks for on-line adaptive phoneme recognition. Technical Report TR99/03, Department of Information Science, University of Otago, 1999.