Vol 14 No 5, May 2005
1009-1963/2005/14(05)/0922-08

Chinese Physics

© 2005 Chin. Phys. Soc.
and IOP Publishing Ltd

Chaotic time series prediction using mean-field theory for support vector machine*

Cui Wan-Zhao(崔万照), Zhu Chang-Chun(朱长纯),
Bao Wen-Xing(保文星), and Liu Jun-Hua(刘君华)

School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

(Received 11 November 2004; revised manuscript received 24 January 2005)

This paper presents a novel method for predicting chaotic time series which is based on the support vector machines approach, and it uses the mean-field theory for developing an easy and efficient learning procedure for the support vector machine. The proposed method approximates the distribution of the support vector machine parameters to a Gaussian process and uses the mean-field theory to estimate these parameters easily, and select the weights of the mixture of kernels used in the support vector machine estimation more accurately and faster than traditional quadratic programming-based algorithms. Finally, relationships between the embedding dimension and the predicting performance of this method are discussed, and the Mackey–Glass equation is applied to test this method. The stimulations show that the mean-field theory for support vector machine can predict chaotic time series accurately, and even if the embedding dimension is unknown, the predicted results are still satisfactory. This result implies that the mean-field theory for support vector machine is a good tool for studying chaotic time series.

Keywords: chaotic time series, support vector machine, mean-field theory
PACC: 0545

1. Introduction

Time series prediction is one of the fundamental issues in a variety of disciplines, ranging from economics through physics to engineering.[1−4] To make a prediction feasible, one has to use either first principles or real observations to discover certain laws governing the phenomena.[5,6] In the cases of lacking first principles, as encountered in many practical problems, deriving an evolution law directly from the past observations (data) seems to be a natural alternative. In fact, the idea has been extensively explored for many years under the title of time series prediction.[7] The prediction is made on the basis of a specified model, which is either linear, such as the ARMA model, or nonlinear, such as bilinear models. As effective alternatives to traditional regression models, various types of machine learning, including neural networks[8,9] and support vector machine (SVM)[10,11] have been recently used to model time series. These studies[10,11] have revealed that SVM is a potential technique for predicting the behaviour of time series because compared with the traditional machine learning approaches, such as the multiplayer perception,[12−14] along with the back-propagation, a SVM does not suffer from problems like the existence of many local minima and the choice of the number of free parameters,[15,16] and the SVM model is often determined by solving a convex quadratic programming problem in dual space. The mean-field theory (MFT), which originates from statistical physics of disordered system, provides efficient approximations which are able to cope with the complexity of probabilistic data models, and replaces the intractable task of computing high-dimensional sums and integrals by the tractable problem of solving a system of linear equations.[17] Recently, MFT has been introduced to SVM[18] with promising results. These actions provide an efficient tool for modelling the chaotic time series to predict future behaviours.

This paper will tackle the time series prediction problem using the SVM with the assumption that the parameters of the SVM are distributed according to a Gaussian process (GP) with a specific parameter

*Project supported by the National Natural Science Foundation of China (Grant Nos 60476037 and 60276037), and the Doctoral Foundation of the National Education Administration of China (Grant No 20020698014).

set estimated from the training data set. The MFT is used to approximate the quadratic programming (QP) problem which arises from the SVM approach. This approximation makes the learning process much efficient in both time and accuracy. Moreover, relationships between the embedding dimension and the mean-square-error are discussed using this method. Finally, the Mackey–Glass equation is used to test this method, and the results shows that this method can predict chaotic time series with reasonably good prediction accuracy; and even if the optimal embedding dimension is unknown, the predicted results are still satisfactory. This implies that the MFT for support vector machine is a good tool in studying chaotic time series.

2. Nonlinear prediction

Chaos occurs even in quite simple deterministic systems, which can model quite complex and erratic behaviours. In mathematical as well as physical systems, a tiny difference in input can result in overwhelming differences in output. This phenomenon is known as sensitive dependence on initial conditions. This means that any microscopic uncertainty in the exact initial state of the given system—whether due to noise, finite measurement precision or any type of perturbation—will result in rapid macroscopic errors in any effort to predict future behaviours.[19] This makes it difficult to make any perfect long-term prediction of a real chaotic system. However, chaotic systems are deterministic, and the knowledge of the past history provides information about what will happen in the future. Therefore, it should be possible to forecast short behaviour. The prediction of chaotic time series is made on the basis of the reconstruction phase space method with known chaotic time series data found by Takens.[20] After embedding the like random chaotic time series data to phase space, a topological attractor of chaotic time series can be observed. Chaotic time series display some stochastic behaviour in time domain, at the same time the determined behaviour of chaotic time series is shown in phase space structure. Therefore, chaotic time series in phase space can be analysed. Takens's embedding method[21] is described as follows: From the observed time series, the predicted time series model from $x(k+m)$ to $x(k+n)$ with the past k time series $\{x(k)\}$ we plan to construct can be described by

$$y(k)_{m,n} = F(x(k), x(k-T), \cdots, x(k-(D-1)T)), \quad (1)$$

where T is the time delay and D is the embedding dimension; Equation (1) indicates one point of a D-dimensional reconstructed phase space R^D. Therefore, a trajectory can be drawn in the D-dimensional reconstructed phase space by changing T; and in this paper, the case of T=1 is analysed. If the target system is a deterministic dynamical system and the observed time series is obtained through an observation system corresponding to continuous mapping from the state space of the dynamical system to the one-dimensional Euclidean space, the reconstructed trajectory is an embedding of the original trajectory when the D value is sufficiently large. If any attractor has appeared in the original dynamical system, another attractor, which retains the phase structure of the first attractor, will appear in the reconstructed state space. In order that such a reconstruction achieves embedding, it has been proven by Takens that $D = 2m + 1$ is sufficient, where m is the state space dimension of the original dynamical system. However, if the state space dimension of a dynamical system is unknown exactly or does not exist in practice, Taken's theorem will never be valid at least in principle. But even in this case, our method still models the past time series and predicts future time series based on the key idea of Taken's theorem.

3. Mean-field theory for SVM

In our method, the SVM is presented as a supervised regression tool, and considered as the maximum a posteriori (MAP) prediction with a Gaussian prior under the Bayesian framework (Bayes' theorem is used to relate the prior to the posterior distribution). The key idea is that, instead of defining prior distributions over parameters of learning machine, a Gaussian prior distribution is assumed over the function space on which the machine computes. In general, the supervised regression learning problem can be described as follows. Given a training data set of N points $\Theta = \{x_k, t_k\}_{k=1}^N$ with input data $x_k \in R^D$ and output data $t_k \in R$, and the goal is to infer the output t for a new input x. The traditional SVM Vapnik ε-loss function is defined as[15]

$$L_\varepsilon(t, y(x)) = \begin{cases} 0 & |t - y(x)| \le \varepsilon, \\ |t - y(x)| - \varepsilon & \text{otherwise}, \end{cases} \quad (2)$$

for a predefined nonnegative constant ε, which controls the noise tolerance. In order to construct a Bayesian framework under Vapnik's ε-insensitive loss

function L_ε, the probabilistic model is employed in which the probability of the output t at a given point x and the likelihood $P[t|y(x)]$ is assumed through a hidden function variable $y(x)$ by the following relationship:

$$P[t|y(x)] = \frac{C}{2(\varepsilon C+1)}\exp\{-CL_\varepsilon(t-y(x))\}. \quad (3)$$

Equation (3) can be interpreted as an additive noise model of the target t. Recently, this noise model has been investigated[22] and a probabilistic interpretation proposed. The standard support vector regression framework is a special case of four regularization networks with this particular noise model.[23] The prior probability distribution $P[y(x)]$ can be considered as a functional GP which is defined as a stochastic process specified by giving only the mean vector and covariance matrix for any finite subset of points. For any finite point set $X = [x_1, x_2, ..., x_N]$, the prior probability distribution $P[y(x)]$ is specified as a GP with a zero mean and a covariance function, and applying Bayes' rule results in the following posterior about the process function $y(X)$:

$$\begin{aligned} P[y(X)|\Theta] =& \frac{K(C,\varepsilon)^N}{\sqrt{\det(2\pi K_N^{-1})}} \\ & \times \exp\Big\{ -C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i)) \\ & -\frac{1}{2}y(X)^{\mathrm{T}}K_N^{-1}y(X)\Big\}/P[\Theta], \end{aligned} \quad (4)$$

where $K(C,\varepsilon) = C/2(\varepsilon C+1)$ and $K_N = [K(x_i, x_j)]$. A Gaussian kernel is used, which is

$$K(x_i, x_j) = \exp(-||x_i - x_j||^2/2a^2), \quad (5)$$

where a is the predefined hyper-parameter of the SVM kernel. And the normalization constant $P[\Theta]$ is obtained by

$$\begin{aligned} P[\Theta] =& \frac{K(C,\varepsilon)^N}{\sqrt{\det(2\pi K_N^{-1})}} \\ & \times \int \exp\Big\{ -C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i)) \\ & -\frac{1}{2}y(X)^{\mathrm{T}}K_N^{-1}y(X)\mathrm{d}y(X). \end{aligned} \quad (6)$$

After some mathematical manipulations, we can find that the MAP estimation of the posterior distribution $P[y(X)|\Theta]$ is the one which maximizes the numerator of Eq.(4). Equivalently, the MAP estimation of the prior distribution is the one which minimizes

$$\min_{y(X)} C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i)) + \frac{1}{2}y(X)^T K_N^{-1}y(X). \quad (7)$$

And Eq.(7) can be converted into a QP problem by introducing some slack variables and dual variables.[15] The calculated posterior average needed for deriving Bayes' algorithm is typically intractable, and approximation techniques are required. Recently, an advanced mean-field theory is introduced based on the ideas of statistical mechanics to cope with the Gaussian classification problem,[17] and this theory is equivalent to the so-called TAP mean-field theory.[24] From the posterior distribution defined in Eq.(4), the prediction of a new test input x is given by

$$\begin{aligned} \langle y(x)\rangle =& \int y(x)P[y(x)\Theta]\mathrm{d}y(x) = \int y(x)P[y(x), y(X)|\Theta]\mathrm{d}y(x)\mathrm{d}y(X) \\ =& \frac{K(C,\varepsilon)^N}{\sqrt{\det(2\pi K_{N+1}^{-1})}} \int y(x)\frac{\exp\Big\{ -C\sum\limits_{i=1}^{N} L_\varepsilon(t_i - y(x_i)) - \frac{1}{2}y(X,x)^T K_{N+1}^{-1}y(X,x)\Big\}}{P[\Theta]}\mathrm{d}y(x)\mathrm{d}y(X) \\ =& \frac{K(C,\varepsilon)^N}{P[\Theta]}\sum_{i=1}^{N} K(x,x_i)\int N(y(X)|0,K_N)\frac{\partial}{\partial y(x_i)}\exp\Big\{ -C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i))\Big\}\mathrm{d}y(X) \\ =& \sum_{i=1}^{N} K(x,x_i)w_i, \end{aligned} \quad (8)$$

where $y(X,x) = [y(x_1), y(x_2), \cdots, y(x_N), y(x)]^{\mathrm{T}}$; w_i is a constant defined as

$$w_i = \frac{K(C,\varepsilon)^N}{P[\Theta]}\int N(y(X)|0,K_N)\frac{\partial}{\partial y(x_i)}\exp\Big\{ -C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i))\Big\}\mathrm{d}y(X), \quad (9)$$

and $K_{N+1} = \begin{pmatrix} K_N & K_N(X)^{\mathrm{T}} \\ K_N(X) & K(x,x) \end{pmatrix}$ with $K_N(X)$ defined by $K_N(X) = [K(x_1,x), K(x_2,x), \cdots, K(x_N,x)]$; and x denotes x_{N+1}.

Denoting an average for this predicted distribution by

$$\langle \cdots \rangle_i = \int \cdots P[y(x_i)|\overline{\Theta}]\mathrm{d}y(x_i), \tag{10}$$

the coefficient in Eq.(9) is rewritten as

$$w_i = \frac{\left\langle K(C,\varepsilon)(\partial/\partial y(x_i))\exp\left\{ -C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i))\right\}\right\rangle_i}{\left\langle K(C,\varepsilon)\exp\left\{ -C\sum_{i=1}^{N} L_\varepsilon(t_i - y(x_i))\right\}\right\rangle_i}. \tag{11}$$

The magnitude of w_i can be interpreted as the normalized variant rate of the likelihood. In order to calculate such weights, a simple and direct method is used to apply some Gaussian approximation to the local predicted distribution as follows:

$$P[y(x_i)|\overline{\Theta_i}] \approx \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left\{ -\frac{(y(x_i) - \langle y(x_i)\rangle_i)^2}{2\sigma_i^2}\right\}, \tag{12}$$

with the variance defined as $\sigma_i^2 = \langle y(x_i)^2\rangle_i - \langle y(x_i)\rangle_i^2$. Substituting Eq.(12) into Eq.(11), we obtain the following expression

$$w_i \approx \frac{F_i}{G_i} = \frac{F(\langle y(x_i)\rangle_i, \sigma_i^2)}{G(\langle y(x_i)\rangle_i, \sigma_i^2)}, \tag{13}$$

where F_i and G_i are computed, respectively, by the following explicit formulae:

$$\begin{aligned} F(\langle y(x_i)\rangle_i, \sigma_i^2) =& \frac{C}{2}\exp\left\{ \frac{C}{2}(2(\langle y(x_i)\rangle_i - 2y_i + 2\varepsilon + C\sigma_i^2)\right\}\left[1 - \mathrm{erf}\left[\frac{-\langle y(x_i)\rangle_i + y_i + \varepsilon + C\sigma_i^2}{\sqrt{2\sigma_i^2}}\right]\right] \\ & - \frac{C}{2}\exp\left\{ \frac{C}{2}(-2(\langle y(x_i)\rangle_i + 2y_i + 2\varepsilon + C\sigma_i^2)\right\}\left[1 - \mathrm{erf}\left[\frac{\langle y(x_i)\rangle_i - y_i + \varepsilon + C\sigma_i^2}{\sqrt{2\sigma_i^2}}\right]\right], \end{aligned} \tag{14}$$

$$\begin{aligned} G(\langle y(x_i)\rangle_i, \sigma_i^2) =& \frac{1}{2}\mathrm{erf}\left[\frac{y_i - \langle y(x_i)\rangle_i + \varepsilon}{\sqrt{2\sigma_i^2}}\right] - \frac{1}{2}\mathrm{erf}\left[\frac{y_i - \langle y(x_i)\rangle_i - \varepsilon}{\sqrt{2\sigma_i^2}}\right] \\ & + \frac{1}{2}\exp\left\{ \frac{C}{2}(2(\langle y(x_i)\rangle_i - 2y_i + 2\varepsilon + C\sigma_i^2)\right\}\left[1 - \mathrm{erf}\left[\frac{y_i - \langle y(x_i)\rangle_i + \varepsilon}{\sqrt{2\sigma_i^2}}\right]\right], \\ & + \frac{1}{2}\exp\left\{ \frac{C}{2}(-2(\langle y(x_i)\rangle_i + 2y_i + 2\varepsilon + C\sigma_i^2)\right\}\left[1 - \mathrm{erf}\left[\frac{-y_i + \langle y(x_i)\rangle_i + \varepsilon}{\sqrt{2\sigma_i^2}}\right]\right], \end{aligned} \tag{15}$$

where $\mathrm{erf}(\cdot)$ is the error function defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\mathrm{d}t. \tag{16}$$

Equations(13)–(15) are the mean-field equations corresponding to the weight parameters. In order to evaluate the weight coefficients, one has to determine the local predicted average $\langle y(x_i)\rangle_i$ and variance σ_i^2 in the approximated Gaussian (12). Recently, an effective mean-field equation for both $\langle y(x_i)\rangle_i$ and σ_i^2 have been obtained by the TAP linear respondent approach.[17] Denote by $\langle y(x_i)\rangle_i$ the posterior average at x_i which is given by Eq.(8) as

$$\langle y(x_i)\rangle = \sum_{j=1}^{N} K(x_i, x_j)w_j, \tag{17}$$

then the formulas for $\langle y(x_i)\rangle_i$ and σ_i^2 can be explicitly expressed as

$$\langle y(x_i)\rangle_i \approx \langle y(x_i)\rangle - \sigma_i^2 w_i, \tag{18}$$

$$\sigma_i^2 \approx \frac{1}{\left[\left(\sum + K\right)^{-1}\right]_{ii}} - \sum\nolimits_i, \tag{19}$$

with

$$\sum = \mathrm{diag}\left(\sum\nolimits_1, \sum\nolimits_2, \ldots, \sum\nolimits_N\right)$$

and

$$\sum\nolimits_i = -\sigma_i^2 - \left(\frac{\partial w_i}{\partial \langle y(x_i)\rangle_i}\right)^{-1}.$$

An explicit expression for $\partial w_i/\partial \langle y(x_i)\rangle_i$ can be obtained from Eq.(13) as

$$\frac{\partial w_i}{\partial \langle y(x_i)\rangle_i} \approx C^2 - w_i^2 - \frac{w_i \langle y(x_i)\rangle_i + \sigma_i^2 C^2 \int_{y_i-\varepsilon}^{y_i+\varepsilon} P[y(x_i)|\overline{\Theta}]\mathrm{d}y(x_i)}{\sigma_i^2 G(\langle y(x_i)\rangle_i, \sigma_i^2)}. \tag{20}$$

Since $P[y(x_i)|\overline{\Theta_i}]$ is assumed to be Gaussian, the integral in Eq.(20) can be easily computed using the error function. The required information for prediction at a test input x^D in Eq.(8) is w_i and the local predictive posterior average $\langle y(x_i)\rangle_i$ and the local predicted posterior variance σ_i^2. The mean-field equation can be solved by an iteration method:

(1) Initialize: Setup the learning rate η, the embedding dimension D, the hyper-parameter of SVM kernel a and ε, and randomly set w_i, $i = 1, \ldots, N$;

(2) Calculate the Kernel matrix K and let $\sigma^2 = K_{ii}$;

(3) Calculate $\langle y(x_i)\rangle$, $\langle y(x_i)\rangle_i$, F_i and G_i, respectively, and adjust w_i by

$$w_i = w_i + \eta(F_i/G_i - w_i). \tag{21}$$

(4) For every M, e.g. M=20 iterations of w_i, update σ_i^2

$$\sigma_i^2 = \frac{1}{\left[\left(\sum + K\right)^{-1}\right]_{ii}} - \sum\nolimits_i, \tag{22}$$

where $\sum_i = -\sigma_i^2 - 1/Dw_i$,

$$Dw_i = C^2 - \sigma_i^2 - \frac{w_i < y(x_i) >_i + \sigma_i^2 C^2 IG_i}{\sigma_i^2 G_i},$$

$$IG_i = \frac{1}{2}\mathrm{erf}\left[\frac{y_i - \langle y(x_i)\rangle_i + \varepsilon}{\sqrt{2\sigma_i^2}}\right] - \frac{1}{2}\mathrm{erf}\left[\frac{y_i - \langle y(x_i)\rangle_i - \varepsilon}{\sqrt{2\sigma_i^2}}\right].$$

(5) If the local predictive posterior average $\langle y(x_i)\rangle_i$ falls below a given tolerance, the iteration procedure is completed. Otherwise, it will return to step (3).

4. Results and discussion

We have applied the method described above to the typical time series, and the prediction performance of our method is evaluated by the root-mean-square-error (R_{mse}):

$$R_{\mathrm{mse}} = \frac{1}{N}\sqrt{\sum_{k=1}^{N}(y(k) - t_k)^2}, \tag{23}$$

where t_k is the actual output and $y(k)$ is the predicted output. And in our stimulations, the values of tuning parameters a, C and ε are 0.5, 0.1 and 0.01, respectively.

The Mackey–Glass time series originally introduced as a model of blood cell regulation, have become quite common as an artificial forecasting benchmark, and are obtained by numerically integrating the delay differential equation[25]

$$\frac{\mathrm{d}x(k)}{\mathrm{d}k} = \frac{0.2x(k-\tau)}{1 + x^{10}(k-\tau)} - 0.1x(k), \tag{24}$$

with delay τ=17 and initial conditions $x(k) = 1.2, 0 < k \leq \tau$. The Mackey–Glass time series in time domain is shown in Fig.1, in which the time series are chaotic, exhibiting a cyclic but not periodic behaviour.

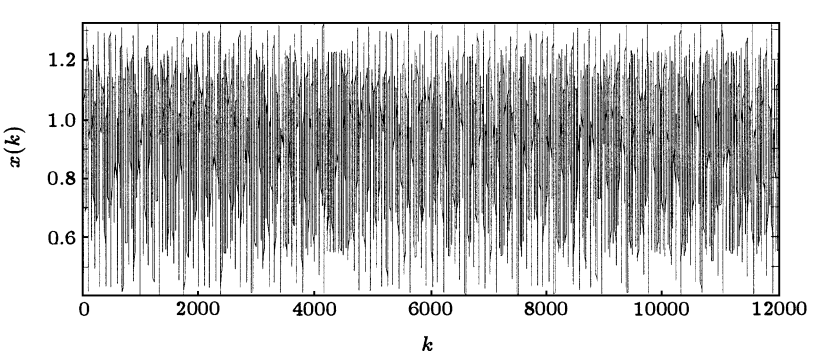


Fig.1. Time series generated by numerically integrating the Mackey–Glass equation.

We take 2000 time series from the numerical solution as our test time series. The first 1000 time series are used to fit a predictive model with our method, and the remaining 1000 time series are used to test one-step prediction, here we choose the embedding dimension D =4. Comparisons of predicted outputs and the actual Mackey–Glass time series are given in Fig.2 (a), and the $R_{\rm mse}$ is 0.0130. The relative errors between predicted and actual time series, which are defined as

$$\text{error}(k) = \frac{y(k) - t_k}{t_k}, \tag{25}$$

are shown in Fig.2(b). It is obvious that the MFT for the SVM is capable of capturing the underlying chaotic dynamics of the system based on some data points, and can predict future chaotic time series accurately.

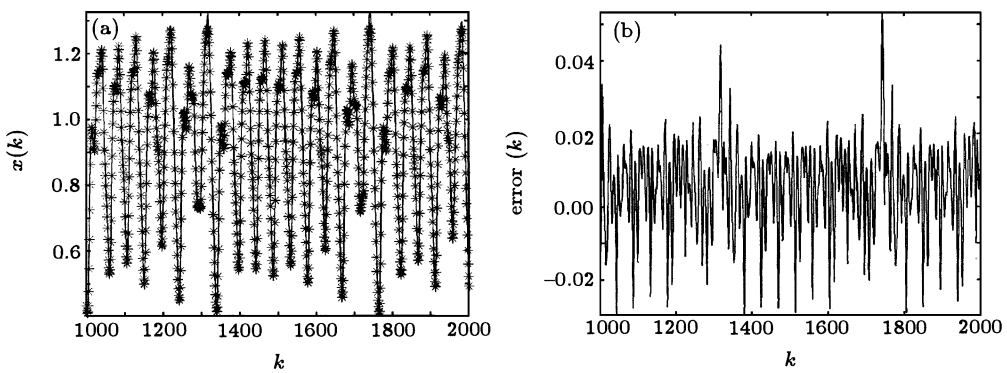


Fig.2. (a) Comparisons of predicted time series $y(k)$, k=1000–2000 with actual time series t_k, k=1000–2000 based on the past time series $x(k)$ k=1–1004 with D=4, and the $R_{\rm mse}$ is 0.0130. The star "*" and the solid line "—" represent the predicted time series and the actual time series, respectively. (b) Relative errors between the predicted time series and the actual time series based on the past time series $x(k)$, k=1–1004 with D=4.

In some cases, the future time series from $x(k+m)$ to $x(k+n)$, for example, from $x(5000)$ to $x(6000)$ is needed for the past k time series $\{x(k)\}$, such as from $x(1)$ to $x(1000)$. If the time series of the training sets do not follow the time series of testing sets, and even there are large time intervals between them, can we predict the future time series $x(k+m)$, ...,$x(k+n)$? In this case, Takens' theorem and extensions are still valid, and predictions of the future time series are possible in a relatively short time. Using a fitted predictive model based on the time series from $x(1)$ to $x(1004)$, the future time series from $x(5000)$ to $x(6000)$, and from $x(9000)$ to $x(10000)$ are predicted, respectively, here we still choose the embedding dimension D =4. Predicted results of MFT for SVM compared with the actual time series are given in Figs.3(a) and 4(a), respectively. Their $R_{\rm mse}$ are 0.0122 and 0.0126, respectively, and their relative errors are shown in Fig.3 (b) and Fig.4(b), respectively. As just mentioned, the modelling performance of predicting

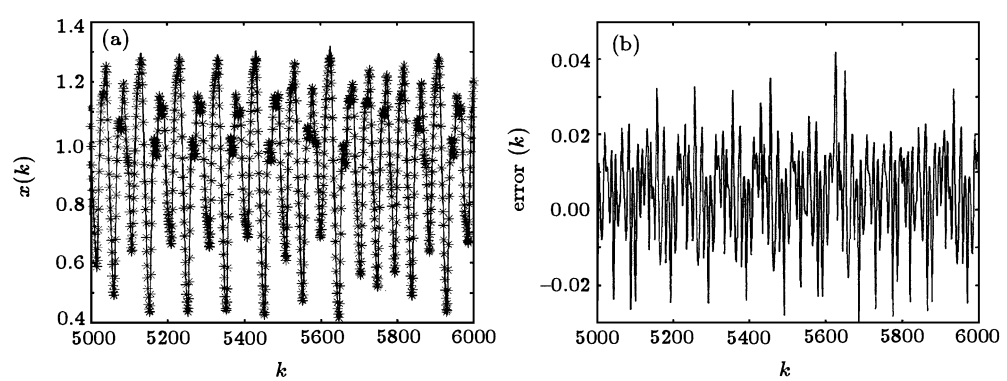


Fig.3. (a) Comparisons of predicted time series $y(k)$, k=5000–6000 with actual time series t_k, k=5000–6000 based on the past time series $x(k)$, k=1–1004 with D=4; and the $R_{\rm mse}$ is 0.0122. The star "*" and the solid line "—" represent the predicted time series, and the actual time series, respectively. (b) Relative errors between the predicted time series $y(k)$, k=5000–6000 with the actual time series t_k, k=5000–6000, based on the past time series $x(k)$, k=1–1004 with D=4.

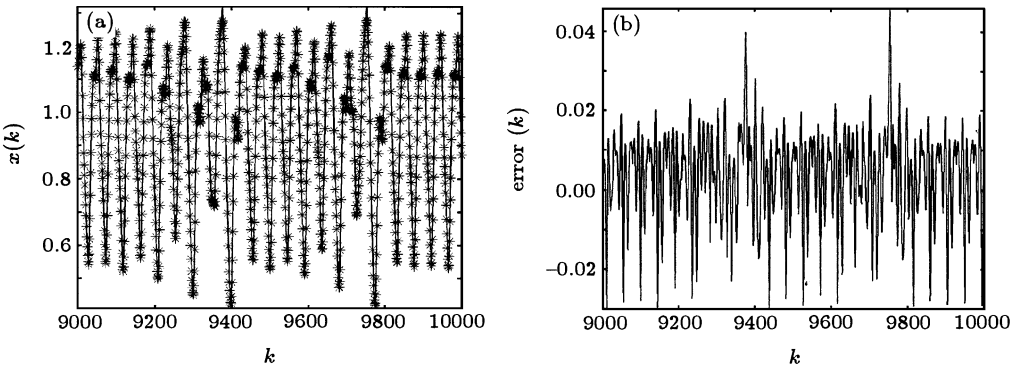


Fig.4. (a) Comparisons of predicted time series $y(k)$, k=9000–10000 with actual time series t_k, k=9000–10000, based on the past time series $x(k)$, k=1–1004 with D=4; and the $R_{\rm mse}$ is 0.0126. The star "*" and the solid line "—" represent the predicted time series and the actual time series, respectively. (b) Relative errors between predicted time series $y(k)$, k=9000–10000 with actual time series t_k, k=9000–10000, based on the past time series $x(k)$, k=1–1004 with D=4.

future chaotic time series is nearly unchangeable by using the MFT for SVM when there are large time intervals between the time series of the training sets and the predicted time series. These results show that the MFT for SVM can capture the underlying dynamics of chaotic time series accurately based on some data points, and can makes some long-term predictions of time series possible.

In many realistic systems, the exact embedding dimensions may be unknown exactly. Taken's theorem and expansions are valid only when the embedding dimension of chaotic time series is known, or can be obtained by the Grassberger–Procaccia method.[26] In such cases, can we predict the future behaviours based on the current behaviours? In fact, we can still model the chaotic time series on the basis of the key ideas of the reconstructed phase space. However, Taken's theorem does not show the relationships between the embedding dimension and the predicted performance. The future time series from $x(1000)$ to $x(2000)$, from $x(5000)$ to $x(6000)$, and from $x(9000)$ to $x(10000)$ are predicted with a fitted predictive model based on the time series from $x(1)$ to $x(1000+D-1)$ with a varying embedding dimension D from 1 to 20, respectively. The relationships between the predicting performance and the embedding dimension are shown in Figs.5–7. From Figs.5–7, it can be obtained that their $R_{\rm mse}$ are minimal when the embedding dimension D is 3, and they are 0.0108, 0.0109 and 0.0110, respectively; while their $R_{\rm mse}$ are maximal when their embedding dimensions D are 1, 13 and 17, and they are 0.0395, 0.0516 and 0.0413, respectively. In fact, the cases of $D > 12$ is few. But even so, the predicted performance is still satisfactory. These results show that the mean-field theory for support vector machine can predict the future time series accurately even when the optimal embedding dimension cannot be obtained exactly.

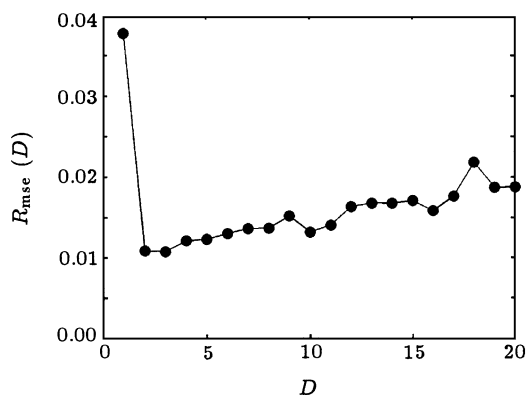


Fig.5. Relations between the $R_{\rm mse}$ and the embedding dimension for D=1–20. The predicted time series is from 2000 to 3000 based on the past time series from 1 to $(1000+D-1)$.

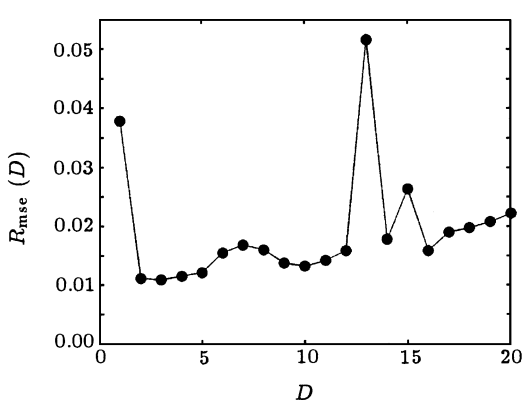


Fig.6. Relations between $R_{\rm mse}$ and the embedding dimension for D=1–20. The predicted time series is from 5000 to 6000 based on the past time series from 1 to $(1000+D-1)$.

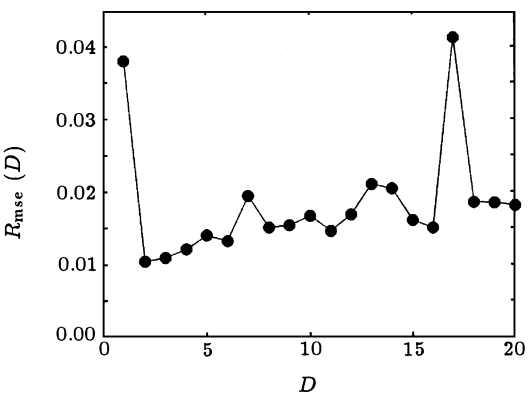


Fig.7. Relations between the $R_{\rm mse}$ and the embedding dimension for D=1–20. The predicted time series is from 9000 to 10000 based on the past time series from 1 to $(1000+D-1)$.

5. Conclusion

In this paper, a novel technique for the prediction of chaotic time series has been proposed. The method uses the MFT in the implementation of SVM as a time series modelling method. The MFT reduces the quadratic programming problem, which comes from the SVM formulation, to an iterated procedure. This reduction facilitates visible implementation of the SVM method. Relationships between the embedding dimension and the predicted performance (mean-square-error) are discussed using this method. Finally, the Mackey–Glass equation is applied to test this method, and the results shows that this method can predict chaotic time series accurately, and even if the optimal embedding dimension is unknown, the predicted results are still satisfactory. This implies that the MFT for SVM is a potential tool in studying chaotic time series.

References

[1] Gershenfeld N *et al* 1999 *Nature* **397** 392
[2] Nie J 1997 *Neurocomputing* **16** 63
[3] Gan J C *et al* 2003 *Acta Phys. Sin.* **52** 2995 (in Chinese)
[4] Zhang J S *et al* 2000 *Chin. Phys. Lett.* **17** 88
[5] Frame J D *et al* 1987 *Phys. Rev. Lett.* **59** 845
[6] Weigend A S *et al* 1991 *Back-propagation, weight-elimination and time-series prediction* (San Mateo, CA: Morgan Kaufmann)
[7] Tong H 1990 *Non-linear Time-Series: A Dynamical System Approach* (Oxford: Oxford University Press)
[8] Tan W *et al* 2003 *Acta Phys. Sin.* **52** 795 (in Chinese)
[9] Gao L Y 1995 *Physica* D **85** 225
[10] Cui W Z *et al* 2004 *Acta Phys. Sin.* **53** 3303 (in Chinese)
[11] Ye M Y *et al* 2004 *Chin. Phys.* **13** 454
[12] Hornik K *et al* 1989 *Neural Networks* **2** 359
[13] Cui W Z *et al* 2004 *Acta Phys. Sin.* **53** 1583 (in Chinese)
[14] Cui W Z *et al* 2005 *Thin Solid Films* **473** 224
[15] Vapnik V N 1995 *The Nature of Statistical Learning Theory* (New York: Springer)
[16] Cui W Z *et al* 2004 *Journal of Xi'an Jiaotong University* **38** 562 (in Chinese)
[17] Opper M *et al* 2000 *Neural Comput.* **12** 2655
[18] Gao J B *et al* 2003 *Neurocomputing* **50** 391
[19] Zhang J *et al* 2004 *Comput. Electr. Eng.* **30** 1
[20] Takens F 1981 *Detecting Strange Attractors in Turbulence, Lecture Notes in Mathematics* (Berlin: Springer) p366
[21] Kathleen T *et al* 1996 *Chaos, An Introduction to Dynamical Systems* (New York: Springer)
[22] Pontil M *et al* 2000 *On the Noise Model of Support Vector Machine Regression, Lecture Notes in Computer Science* (Berlin: Springer) p316
[23] Evgeniou T *et al* 2000 *Adv. Comput. Math.* **13** 1
[24] Thouless D *et al* 1977 *Philos. Mag.* **35** 593
[25] Mackey M C *et al* 1977 *Science* **197** 287
[26] Grassberger P *et al Phys. Rev. Lett.* **50** 346