http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 5 (3): 529-533, 2006
ISSN 1812-5638

Predict Chaotic Time Series Using Minimax Probability Machine Regression

[1,2]Liu Zunxiong, [2]Xie Xin and [1]Zhang Deyun
[1]School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]School of Information Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract: Minimax Probability Machine Regression (MPMR) is proposed for chaotic time series global prediction in this study. In MPMR, regression function maximizes the minimum probability that future predication will be within an |Å to the true regression function. Multi-step predictions up to 20 steps were done on Mackey-Glass chaotic time series with MPMR and Local Weighted Linear Regression (LWLR). The results demonstrate that MPMR have better prediction performance, compared with LWLR. Kernel function shape parameter and regression tube value will influence the MPMR-based system performance. In experiments, the cross validation methods are employed to choose the two parameters.

Key words: Chaotic time series, prediction, minimax probability machine regression, global model

INTRODUCTION

Chaotic time series prediction has been widely investigated in various fields ranging from physics, chemistry, biology, meteorology, hydrology to finance. The prediction of chaotic time series has been of great practical significance, which has drawn the attention of many researchers. In essence, chaotic time series is obtained by observing a certain state variable from a nonlinear dynamic system and the classical linear method are inadequate to predict chaotic time series. There are several techniques for modeling and forecasting nonlinear time series such as local linear mode (Farmer and Sidorowich, 1987; Jayawardena *et al*., 2002), chaotic attractor (Kengping and Tianlun, 2002), radial basis function (Casdagli, 1989) and feedforward network model (Tronci *et al*., 2003; Oliveira *et al*., 2000), which belong to the global and local model, respectively.

Minimax Probability Machine Regression(MPMR) is new regression method based on Minimax Probability Machine Classification(MPMC). MPMC is first put forward for linear classifcation that maximize the minimum probability of correct classification on future data (Lanckriet *et al*., 2002) and can be furthermore generalized as nonlinear classifier by utilizing Mercer kernels. MPMC can obtain distribution-free results for linear discriminants with the help of convex optimization and the worst-case bound on the probability of misclassification of future data can be estimated explicitly. In the same way MPMR maximizes the minimum probability that the regression model is within ε tube of the true regression function.

In this study MPMR is supposed to make multi-step global prediction for chaotic time series, achieving better prediction accuracy compared with other methods, such as Local Weighted Linear Regression (LWLR).

PHASE SPACE RECONSTRUCTION

The theory of PSR (Phase Space Reconstruction) is the basis for nonlinear dynamic system analysis with a observation series, should be implemented under the embedding theorem (Takens, 1981). A reconstructed phase space is an m-dimensional metric space into which a time series is embedded. Given a chaotic time series $\{x_i, i = 1, 2, L\ n\}$, if m (number of dimension of phase space) is large enough, the phase space is homeomorphic to the state space that generated the time series, i.e., the phase space contains the same information as the original state space. The time-delayed embedding of a time series maps a set of time series observations taken from {x(i)} onto **x**(i); where, **x**(i) is a vector or point in the phase space, defined as $\mathbf{x}(i) = (x(i), x(i-\tau), L\ x(i-(d-1)\tau))$, where, $i = (m-1)\tau+1$, L, n and d is embedding dimension, τ is delay time. In the reconstructed phase space there is a smooth map f: $\mathbf{R}^m \rightarrow \mathbf{R}$ such that:

$$x(i+\tau) = f\left(x(i), x(i-\tau), \cdots x(i-(m-1)\tau)\right) = f\left(\mathbf{x}(i)\right) \quad (1)$$

Where, $\mathbf{x}(i) = (x_{i1}, x_{i2}, L\ x_{im},)$ is state vector at time i (current state), $y_i = x_{i+\tau}$ is time series value at time $i+\tau$ (future time). Thus time series prediction can be carried out as long as an appropriate expression for f is found. The function can be estimated using local and global

Corresponding Author: Liu Zunxiong, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

model with different phase point sets. Local model estimates the mapping function using the local approximation method, being done with phase point sets neighboring to the prediction point in a piecewise manner. In global models the function approximation is estimated for the whole domain with all phase points, here MPMR is used to do it in global model.

MINIMAX PROBABILITY MACHINE REGRESSION

The approximation mapping function f should be established by learning from the training dataset with regression problem, so there is corresponding output estimation $\hat{y} = f(\mathbf{x})$ for a given input $\mathbf{X}$. The model, maximizing the minimum probability that the regression model is within ε tube of the true regression function, is better.

Minimax probability machine classification: Minimax Probability Machine Classification(MPMC) is first used to separate with linear decision boundary two class of points $\{\mathbf{u}\}_{i=1}^{nu}$ and $\{\mathbf{v}\}_{i=1}^{nv}$, which mean vectors and covariance matrices given by $\mathbf{u} \sim (\bar{u}, \Sigma_u)$ and $\mathbf{v} \sim (\bar{v}, \Sigma_v)$. Thus, the hyperplane $\mathbf{a}^T\mathbf{z} = b(\mathbf{a}, \mathbf{z} \in \mathbb{R}^n, b \in \mathbb{R})$ should be determined, which separates the two classes of points with maximal probability. It can be defined:

$$\max_{\alpha,a,b} \alpha \quad \text{s.t.} \quad \begin{matrix} 1-\alpha \geq \sup \Pr\{\mathbf{a}^T\mathbf{u} \leq b\} \\ 1-\alpha \geq \sup \Pr\{\mathbf{a}^T\mathbf{v} \geq b\} \end{matrix} \tag{2}$$

With the help of the theory (Popescu and Bertsimas, 2001), $\sup \Pr\{\mathbf{a}^T\mathbf{v} \geq b\} = 1/(1+d^2)$ and

$$d^2 = \inf_{\mathbf{a}^T\mathbf{v} \geq b} (\mathbf{v}-\bar{\mathbf{v}})^T \Sigma_v^{-1} (\mathbf{v}-\bar{\mathbf{v}}),$$

The above optimization problem can be converted to,

$$\min_{\mathbf{a}} \sqrt{\mathbf{a}^T \Sigma_u \mathbf{a}} + \sqrt{\mathbf{a}^T \Sigma_v \mathbf{a}} \text{ s.t. } \mathbf{a}^T(\bar{u}-\bar{v}) = 1$$

The solution to the problem is $\mathbf{a}_*$,

$$b_* = \mathbf{a}_*^T\bar{u} - \kappa_*\sqrt{\mathbf{a}^T \Sigma_u \mathbf{a}} = \mathbf{a}_*^T\bar{v} - \kappa_*\sqrt{\mathbf{a}^T \Sigma_v \mathbf{a}}$$

and

$$\kappa_*\left(\sqrt{\mathbf{a}_*^T \Sigma_u \mathbf{a}_*} + \sqrt{\mathbf{a}_*^T \Sigma_v \mathbf{a}_*}\right) = 1$$

Kernel method can be introduced to separate two class of points $\{\mathbf{u}\}_{i=1}^{nu}$ and $\{\mathbf{v}\}_{i=1}^{nv}$ unseparated with a hyperplane. The input space is mapped to a higher dimensional feature space with a mapping function $\varphi: \mathbb{R}^n \to \mathbb{R}^f$ and the data be mapped

$$\mathbf{u} \mapsto \varphi(\mathbf{u}) \sim (\overline{\varphi(\mathbf{u})}, \Sigma_{\varphi(u)})$$

and

$$\mathbf{v} \mapsto \varphi(\mathbf{v}) \sim (\overline{\varphi(\mathbf{v})}, \Sigma_{\varphi(v)})$$

The binary classifier has the form(c=-1 for the first class and c=+1 for the second):

$$c = \text{sign}[\sum_{i=1}^{nu+nv} \gamma_i K^c(z_i, z) + b_c] \tag{3}$$

Where,

$$K^c(z_i, z) = \varphi(z_i)\varphi(z),$$
$$z_i = \mathbf{u}_i, (i = 1, \text{L } nu)$$
$$z_i = \mathbf{v}_i, (i = nu+1, \text{L}, nu+nv)$$
$$\Upsilon = (\Upsilon_1, \ldots \Upsilon_{nu+nv})$$

The above problem can be solved with the following optimization:

$$\min_{\gamma}\left\{\left\|\frac{\tilde{\mathbf{K}}_u}{\sqrt{nu}}\gamma\right\|_2 + \left\|\frac{\tilde{\mathbf{K}}_v}{\sqrt{nv}}\gamma\right\|_2\right\} \quad \text{s.t. } \gamma^T(\tilde{k}_u - \tilde{k}_v) = 1 \tag{4}$$

Where,

$$\tilde{\mathbf{K}}_u = \mathbf{K}_u - \mathbf{1}_{nu}\tilde{k}_u, \ \tilde{\mathbf{K}}_v = \mathbf{K}_v - \mathbf{1}_{nv}\tilde{k}_v, \ \tilde{k}_u, \tilde{k}_v \in \mathbb{R}^{nu+nv}$$

Defined as:

$$[\tilde{k}_u]_i = \frac{1}{nu}\sum_{j=1}^{nu} K^c(\mathbf{u}_j, z_i)$$

and

$$[\tilde{k}_v]_i = \frac{1}{nv}\sum_{j=1}^{nv} K^c(\mathbf{v}_j, z_i);$$

$\mathbf{1}_k$ is a k dimensional column vector of ones; $\mathbf{K}_u$ contains the first nu rows of the Gram matrix $\mathbf{K}$, $\mathbf{K}_u$ contains the last nv rows of the Gram matrix $\mathbf{K}$ and $\mathbf{K}$ is a square matrix consisting of the elements $K_{ij} = K^c(z_i, z_j)$.

Given that γ solves the minimization problem in (4), b_c can be caculated using:

$$b_c = \gamma^T \tilde{k}_u - \kappa\sqrt{\frac{1}{Nu}\gamma^T \tilde{K}_u{}^T \tilde{K}_u \gamma} = \gamma^T \tilde{k}_v + \kappa\sqrt{\frac{1}{Nv}\gamma^T \tilde{K}_v{}^T \tilde{K}_v \gamma} \quad (5)$$

Where,

$$\kappa = \left(\sqrt{\frac{1}{nu}\gamma^T \tilde{K}_u{}^T \tilde{K}_u \gamma} + \sqrt{\frac{1}{nv}\gamma^T \tilde{K}_v{}^T \tilde{K}_v \gamma}\right)^{-1}$$

Here the maximum probability of incorrect classification is bounded by $1-\alpha$, $\alpha = \kappa^2/(1+\kappa^2)$

Minimax probability machine regression: MPMC-based regression, Minimax Probability Machine Regression (Strohmann and Grudic, 2003), requires that two (1+m) dimensional vectors $\mathbf{u}_i$ and $\mathbf{v}_i (i = 1, L\ N)$ are produced with every sample data $(\mathbf{x}_i, y_i)$. That is $\mathbf{u}_i = (y_i + \varepsilon, x_{i1}, x_{i2}, L\ x_{im})$ and $\mathbf{v}_i = (y_i - \varepsilon, x_{i1}, x_{i2}, L\ x_{im})$, therefore making two classes of points. Appling them to MPMC, γ is obtained by minimizing Eq. 4, resulting in regression surface for MPMR given by:

$$\sum_{i=1}^{2N} \gamma_i K^c(\boldsymbol{z}_i, \boldsymbol{z}) + b_c = 0 \quad (6)$$

Given the input $x = (x_{i1}, x_{i2}, L\ x_{im})$, the regression N+1,...., 2N model output $\hat{y}$ can be obtained by solving Eq. 6 about y, where $\mathbf{z} = (y, x_1, x_2, L\ x_m)$. With above produced dataset there exist $\boldsymbol{z}_i = \mathbf{u}_i$ for i = 1, L N and $\boldsymbol{z}_i = \mathbf{v}_{i-N}$ for i = N+1, L 2N. If $K^c(\boldsymbol{z}_i, \boldsymbol{z})$ is nonlinear, solving equation (6) for $\hat{y}$ is a nonlinear single variable optimization problem.

To allow Eq. 6 be solved analytically, nonlinear $K^c(\boldsymbol{z}_i, \boldsymbol{z})$ should be restricted as:

$$K^c(\boldsymbol{z}_i, \boldsymbol{z}) = y_i' \hat{y} + K(\boldsymbol{x}_i, \boldsymbol{x}),\ K(\boldsymbol{x}_i, \boldsymbol{x}) = \varphi(\boldsymbol{x}_i)\varphi(\boldsymbol{x})$$

Is Mercer kernel, where,

$$\mathbf{z} = (\hat{y}, x_1, x_2, \cdots x_m), \boldsymbol{z}_i = \mathbf{u}_i, y_i' = y_i + \varepsilon$$

for

$$i = 1, \cdots N,\ \boldsymbol{z}_i = \mathbf{v}_{i-N}, y_i' = y_i - \varepsilon$$

for

$$I = N+1, \ldots 2N$$

The analytical solution to Eq. 6 exists in the form of $\beta_i = -2\varepsilon(\gamma_i + \gamma_{i+N})$ and $b = -2\varepsilon b_c$, thus Eq. 4 optimization for γ be reduced to solving the following linear least square problem:

$$\min_t \left\| \tilde{\mathbf{K}}_u (\gamma_0 + \mathbf{F}\mathbf{t}) \right\|_2 \quad (7)$$

Where,

$$\gamma = \gamma_0 + \mathbf{F}\mathbf{t}$$

$$\gamma_0 = (\tilde{\mathbf{k}}_u - \tilde{\mathbf{k}}_v) / \left\| \tilde{\mathbf{k}}_u - \tilde{\mathbf{k}}_v \right\|^2,$$

$$\mathbf{t} \in \mathbb{R}^{2N-1}, \mathbf{F} \in \mathcal{R}^{2N \times (2N-1)}$$

Is an orthogonal matrix whose columns span the subspace of vectors orthogonal to $\tilde{\mathbf{k}}_u - \tilde{\mathbf{k}}_v$.

CHAOTIC TIME SERIES GENERATION

Considering the time series generated by Mackey-Glass equation, a time-delay ordinary differential equation displaying well-known chaotic behaviors. It is described in the following formula:

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1x(t) \quad (8)$$

This time series is chaotic, always used as benchmark dataset for nonlinear dynamic system modeling and forecasting in the neural network and fuzzy modeling research communities. Here the fourth-order Runge-Kutta method used to find the numerical solution to the above MG equation, the time series value at integer points $\{x_t\}_{t=1}^N$ obtained. Assuming that $x_0 = 1.2$, $\tau = 17$ and

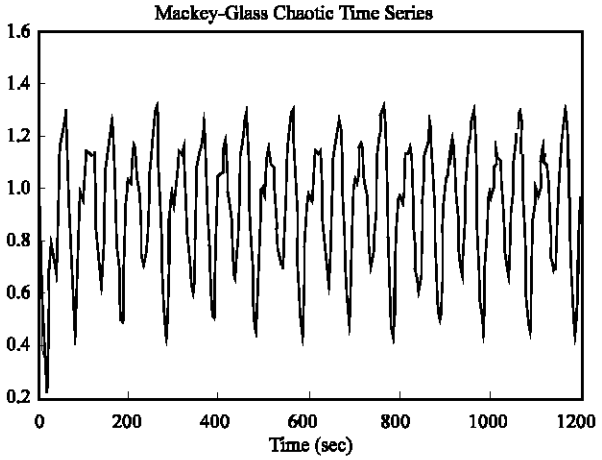

Fig. 1: The chaotic time series generated with Mackey-glass equation for prediction

x(t) = 0 for t<0. This time series shown in Fig. 1 make the experimental data for the simulation.

According to standard practice with the Mackey-Glass chaotic time series, The embedding dimension and time delay are set in m = 4 and $\tau = 6$ with the reconstructed phase space for the time series. For each t, there is a point in the form of $\mathbf{x}_t = (x_t, x_{t-6}, x_{t-12}, x_{t-18})$, as input training data for MPMR. The corresponding output x_{t+6}. In order to predict the future value, the mapping function $x_{t+6} = F(\mathbf{x}_t)$ is obtained with MPMR in global model here.

PREDICTION SIMULATION

For each t, ranging for 118 to 1117, 1000 input/output data pairs $\{\mathbf{x}_t, x_{t+6}\}_{t=118}^{1117}$ are formed, making the first 500 points for the training dataset and the last 500 points for the test dataset. Modeling the training set is done with MPMR, which kernel being set to RBF function. Moreover for the width θ of the RBF function and tube value ε selection, we perform cross-validation with (θ, ε) combination. In this experiment they are validated to be 2 and 3 respectively.

In the above approch, a single-step predictor is constructed with MPMR. However multi-step predictions can be accomplished from the single-step predictor by simply iterating it, which means that the previously predicted values will be used to predict latter step values. For the considered instance, after the prediction result at t+6, $\tilde{x}_{t+6}$ being made, the corresponding input x_{t+6} will be lined in $\mathbf{x}_{t+6} = (\tilde{x}_{t+6}, x_t, x_{t-6}, x_{t-12})$ to predict the value at t+12, $\tilde{x}_{t+12}$. The process can be iterated to obtain multi-step prediction results, 20-step prediction will be made with the MPMR-trained model in the simulation. Two error metrics are utilized to evaluate the prediction performance, one is MAPE defined as:

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|\frac{x_t - \tilde{x}_t}{x_t}\right| n}{n},$$

The other is RMSE calculated with

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n}\left(x_t - \tilde{x}_t\right)^2 / n};$$

Where, n is equal to 500, corresponding to the number of total prediction points. In Fig. 2 presents samples of the 18-step ahead prediction results on the 500-point test dataset and the true value in graphical form, the actual values and the predicted values are shown in solid line and dotted line respectively. It can be seen that the prediction values fit the true values well and the

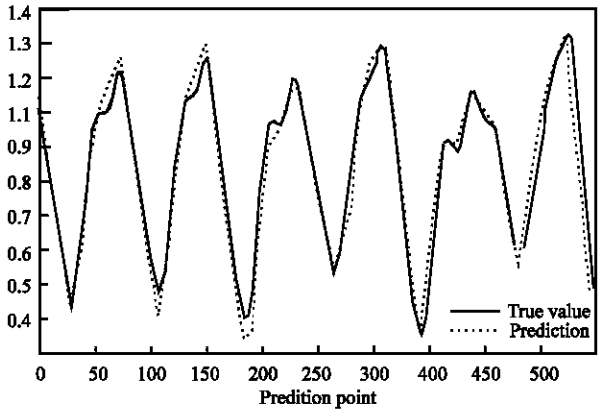

Fig. 2: The 18-step ahead prediction values on the test dataset and the actual values

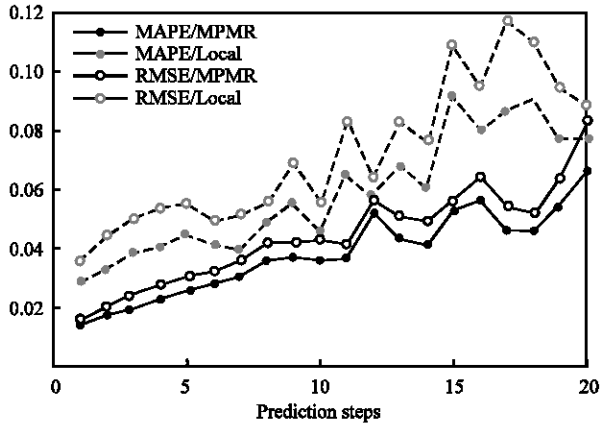

Fig. 3: The error metric curve MAPE and RMSE for the multi-step ahead prediction on the test dataset value

calculated MAPE and RMSE are 0.04528 and 0.051747. Multi-step ahead predictions from 1 to 20 step are experimented on the test dataset and the related error metrics MAPE and RMSE are calculated and shown in Fig. 3 (indicated by legend MPMR). The results tell that the prediction accuracy drops with the increase of prediction step, verifying short-term predictability for chaotic time series.

As comparison, Local wighted linear regression(LWLR) has been used to model the same chaotic time series, multi-steps ahead prediction from 1 to 20 steps being done on the test dataset, where the number of nearest neighbors is 8. Its error metrics relation to prediction steps is also shown in Fig. 3 (with legend Local). It's evident that the two error metrics for MPMR are better than LWLR. In Table 1 presents experiment error metric values MAPE and RMSE with multi-step ahead prediction of 1, 5, 10, 15 and 20 steps for MPMR and LWLR.

Table 1: Two error metrics with multi-step ahead prediction for MPMR and LWLR

Prediction steps		1	5	10	15	20
MAPE×10^{-2}	MPMR	1.3001	2.521	3.5396	5.2121	6.6162
	LOCAL	2.7544	4.4033	4.5105	9.0302	7.6375
RMSE×10^{-2}	MPMR	1.4589	2.9585	4.1873	5.608	8.299
	LOCAL	3.4959	5.4673	5.5184	10.804	8.7871

The calculation cost for MPMR optimization algorithm is proportional to the number of training samples because of kernel matrix calculation, so the training dataset can be reduced properly before learning process is executed. MPMR is also used with the first 200 points of 500-point training dataset and a good prediction accuracy on test dataset is achieved. If neural network model used, more training samples should be requested and the learning time for modeling will be relatively longer. That is to say, MPMR is superior to neural network model for chaotic time series prediction in the ratio of performance to time.

CONCLUSIONS AND FUTURE WORK

MPMR takes convex optimization to obtain distribution-free results and the worst-case bound on the probability that the regression model is within ε tube of the true regression function, can be directly estimated. In this study multi-steps ahead prediction on Mackey-Glass chaotic time series is done with MPMR, the good performance being present. It can be said MPMR be excellent predictor with chaotic time series, compared with other approaches. In the future MPMR predictor will be generalized into the engineering field, taking use of the advantage of local model and combined with it.

REFERENCES

Casdagli, M., 1989. Nonlinear prediction of chaotic time series. Phys., D 35: 335.

De Oliveira, K.A., A. Vannucci and E.C. da Silva, 2000. Using artifficial neural network to forecast chaotic time series. Phys., A 284.

Farmer, J.D. and J.J. Sidorowich, 1987. Predicting chaotic time series. Phys. Rev. Lett., 59: 845.

Jayawardena, A.W., W.K. Li and P. Xu, 2002. Neighbourhood selection for local modelling and prediction of hydrological time series [J]. J. Hydrol., 258: 40-57.

Kengping, L. and C. Tianlun, 2002. Phase Space Prediction Model Based on the Chaotic Attractor. Chin. Phys. Lett., 19: 904-907.

Lanckriet, G.R.G., L.E. Ghaoui, C. Bhattacharyya and M.I. Jordan, 2002. Minimax probability machine. In: Dietterich, T.G., S. Becker and Z. Ghahramani (Eds.). Advances in Neural Information Processing Systems 14, Cambridge, MA, 2002. MIT Press.

Popescu, I. and D. Bertsimas, 2001. Optimal inequalities in probability theory: A convex optimization approach. Technical Report TM62, INSEAD, Dept. Math. O.R., Cambridge, Mass.

Strohmann, T. and G.Z. Grudic, 2001. A formulation for minimax probability machine regression, Accepted for Publication at NIPS, 2002. Neural Information Processing Systems: Natural and Synthetic. Monday, Dec. 3--Saturday, Dec. 8, 2001 Vancouver, British Columbia, Canada.

Takens, F., 1981. Detecting strange attractors in fluid turbulence. Springer lecture notes in mathematics (Springer, Berlin), 898: 366-381.

Tronci, S., M. Giona and R. Baratti, 2003. Reconstruction of chaotic time series by neural models: A case study. Neurocomputing, 55: 581-591.