The Fourteenth International Symposium on Artificial Life and Robotics 2009 (AROB 14th '09),
B-Con Plaza, Beppu, Oita, Japan, February 5 - 7, 2009

ARRBFNs with SVR for prediction of chaotic time series with outliers

Yu-Yi Fu[1], Chia-Ju Wu[2], Chia-Nan Ko [3], Jin-Tsong Jeng[4] and Li-Chun Lai[5]

[1]Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology
Douliou, Yunlin 640, Taiwan
[2]Department of Electrical Engineering,National Yunlin University of Science and Technology
Douliou, Yunlin 640, Taiwan
[3]Department of Automation Engineering, Nan Kai University of Technology
Tsaotun, Nantou 542, Taiwan
[4]Department of Computer Science and Information Engineering, National Formosa University
Huwei, Yunlin 632, Taiwan
[5]Department of Electrical Engineering, Chung Chou Institute of Technology
Yuanlin, Changhua 510, Taiwan
(email:[1]t098@nkut.edu.tw;[2]WUCJ@yuntech.edu.tw;[3]t105@nkut.edu.tw;[4]tsong@nfu.edu.tw;[5]lclai@dragon.ccut.edu.tw)

Abstract: In this paper, the annealing robust radial basis function networks (ARRBFNs) which consist of a radial basis function networks and a support vector regression (SVR) and an annealing robust learning algorithm (ARLA) are proposed for the prediction of chaotic time series with outliers. In order to overcome the initial structure problems of the proposed neural networks, the SVR is utilized to determine the number of hidden nodes, the initial parameters of the kernel and the initial weights for the proposed ARRBFNs. Then, the ARLA that can against the outliers is applied to tune the parameters of the kernel and the weights in the proposed ARRBFNs under the initial structure with SVR. From the simulation results of Mackey–Glass time series show that the proposed approach with different SVR can overcome outliers and fast learning speed. Besides, results of simulation are also given to demonstrate the validity of proposed method for the chaotic time series with outliers.

Keywords: Chaotic time series; Outliers; Support vector regression; Annealing robust learning algorithm.

I. INTRODUCTION

The prediction of time series is a very important application including weather forecast, economic plan, behavior-based intelligent mobile robot, etc. With the character of nonlinear system, chaotic time series presents sensitive dependence on initial conditions and unstable activities. Hence, many approaches with neural networks are proposed recently.

Gu and Wang [1] used recursive least square algorithm with singular value decomposition to estimate the parameters of fuzzy model. Zhang et al. [2] used genetic algorithm and particle swarm optimization to model chaotic time series. Harpham and Dawson [3] developed a chaotic time series prediction based on radial basis function networks (RBFNs). RBFNs can be used to approximate the desired outputs without needing a mathematical explanation of how the outputs functionally depend on the inputs. However, these approaches haven't a methodical way to determine the number of hidden nodes, the initial parameters of the kernel and the initial weights of the networks. Besides, the data we obtained sometimes contain outliers. Outliers may occur due to improper measurements or noisy data. When outliers occur there is overfitting problem that appears in the learning of these neural networks approaches [4].

Support vector regression (SVR) approach was proposed by Vapnik [5], by the insensitive loss function can make use of a small subset of the training data, called the support vectors (SVs), to approximate the desired outputs within a tolerance band. The selection of the hyper-parameters for SVR lacks systematic way to determine [6], and could not do with learning mechanism to update the weights and the parameters of kernel. In this paper, in order to overcome the above problem, an annealing robust radial basis function networks (ARRBFNs) based on different SVR is proposed to predict Mackey–Glass time series with outliers. First, an ε or ν SVR is used to determine the number of hidden nodes, the initial parameters of the kernel, and the initial weights of the ARRBFNs. Then the ARLA is applied to tune the parameters of radial basis functions and the synaptic weights. It is expected that the proposed approach has fast convergence speed and the ability facing outliers can predict perfectly.

II. THE PROPOSED ARRBFNS FOR PREDICTION OF CHAOTIC TIME SERIES WITH OUTLIERS

In this paper, the Mackey-Glass time series is used to simulation, and is defined as

$$\dot{x} = \frac{\alpha x(t-\tau)}{1+x^{\gamma}(t-\tau)} - \beta x(t). \qquad (1)$$

The problem is how to utilize the past values of x to predict the future value $x(t+\Delta t)$, as following

$$x(t+\Delta t) = f(x(t), x(t-\Delta t), \cdots, x(t-n\Delta t)), \qquad (2)$$

where Δt is the sampling interval. In general, outliers occur due to various reasons, such as improper

measurements or noisy data. Our purpose is to find a suitable nonlinear identification model

$$\hat{x}(t+\Delta t) = \hat{f}(x(t), x(t-\Delta t), \cdots, x(t-n\Delta t)) \quad (3)$$

of the time series, where $\hat{x}(t+\Delta t)$ is the output of the neural network and $\hat{f}$ is the estimate of f, which when subjected to the same past values of x, produces an output $\hat{x}(t+\Delta t)$ which approximates $x(t+\Delta t)$ as close as possible.

When the radial basis functions are chosen as Gaussian functions, an ARRBFNs can be expressed as the form

$$\hat{x}_l(t+\Delta t) = \sum_{j=0}^{L} \omega_{jl} G_j = \sum_{j=0}^{L} \omega_{jl} \exp\left(-\frac{\left\| \boldsymbol{x} - \boldsymbol{m}_j \right\|^2}{2\sigma_j^2}\right), \quad \text{for } l = 1, 2, \cdots, p, \quad (4)$$

where $\hat{x}_l$ is the lth output of the neural networks, $\boldsymbol{x} = (x(t), x(t-\Delta t), \cdots, x(t-n\Delta t))$ is the input to the neural networks, w_{jl}, $0 \le j \le L$, $1 \le l \le p$, are the synaptic weights, $G_j, 0 \le j \le L$, are the Gaussian functions, $\boldsymbol{m}_j, 0 \le j \le L$, and $\sigma_j, 0 \le j \le L$, are the centers and the widths of G_j, respectively, and L is the number of the Gaussian functions, in which we can find that L also denotes the number of hidden nodes.

When utilizing an ARRBFNs for the identification of time series, the goal is to minimize the index as

$$J_N(h) = \frac{1}{N} \sum_{i=1}^{N} \rho\left[e_i(h); \beta(h)\right], \quad (5)$$

where

$$e_i(h) = x_i(t+\Delta t) - \hat{x}_i(t+\Delta t), \quad (6)$$

h is the epoch number, $e_i(h)$ is the error between the ith desired output and the ith output of the ARRBFNs at epoch h and $\rho(\cdot)$ is a logistic loss function and defined as

$$\rho[e_i; \beta] = \frac{\beta}{2} \ln\left[1 + \frac{(e_i)^2}{\beta}\right], \quad (7)$$

where $\beta(h)$ is a deterministic annealing schedule acting like the cut-off points. Hence the ARRBFNs are proposed to overcome the issues while the time series in equation (1) facing with outliers and have faster learning speed than the traditional RBFNs to attain perfect prediction. In the following section, it will be shown how to use SVR approaches to choose these initial values methodically.

III. INITIAL STRUCTURE OF ARRBFNS BY DIFFERENT SVR APPROACH

An SVR approach is used to approximate an unknown function from a set of (input, output) samples $\{ (\boldsymbol{x}_i, x_i(t+\Delta t)), i = 1, \cdots, N \}$. Suppose that a set of basis functions $\{g_k(\boldsymbol{x}), k = 1, 2, ..., m\}$ is given, there exists a family of functions that can be expressed as a linear expansion of the basis functions. The theme is then be changed into finding the parameters of the following basis linear expansion

$$g(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{k=1}^{m} \theta_k g_k(\boldsymbol{x}) + b \quad , \quad (8)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)$ is a parameter vector to be identified and b is a constant to be found. The derivation of ARRBFNs for initial structure with the different SVR will be derived in the following sections.

1. Initial Structure of the ARRBFNs by the ε-SVR Approach

Vapnik[5] firstly proposed the ε-SVR approach. The solution for the theme is to find $g(\boldsymbol{x}, \boldsymbol{\theta})$ that minimizes

$$R(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} L_\varepsilon(u_i - g(\boldsymbol{x}_i, \boldsymbol{\theta})) \quad , \quad (9)$$

subject to the constraint

$$\left\| \boldsymbol{\theta} \right\|^2 < C \, , \quad (10)$$

where $L_\varepsilon(\cdot)$ is the ε-insensitive loss function defined as

$$L_\varepsilon(e) = \begin{cases} 0 & \text{for } |e| \le \varepsilon \\ |e| - \varepsilon & \text{otherwise} \, , \end{cases} \quad (11)$$

for some previously chosen nonnegative number ε.

By using the Lagrange multiplier method, proposed by Vapnik[5] and Smola et al. [7] and the inner product of basis function $g_k(\boldsymbol{x}_r)$ is replaced via the kernel function

$$K(\boldsymbol{x}_r, \boldsymbol{x}_s) = \sum_{k=1}^{m} g_k(\boldsymbol{x}_r) g_k(\boldsymbol{x}_s) \, . \quad (12)$$

It was shown in Vapnik [5] that the solution of the SVR approach is in the form of the following linear expansion of kernel function

$$g(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{k=1}^{m} (\alpha_k^* - \alpha_k) K(\boldsymbol{x}, \boldsymbol{x}_k) + b \, . \quad (13)$$

This means that the parameter θ_i in equation (8) can be represented as $\sum_{i=1}^{m} (\alpha_k^* - \alpha_k) g(\boldsymbol{x}_i)$. Note that only some of $(\alpha_k^* - \alpha_k)$'s are not zeros and the corresponding vectors $\boldsymbol{x}_k$'s are called support vectors (SVs). The derivation of ARRBFNs for initial structure with the ε-SVR will be derived in next section.

2. Initial Structure of the ARRBFNs by the ν-SVR Approach

As it is hard to select a suitable ε, the ν-support vector machine is proposed by Schölkopf et al. [8], a new parameter ν is introduced and one can control the number of SVs and training errors. By the inner product of basis function $g_k(\boldsymbol{x}_r)$ is replaced via the kernel function, the issue is to find the parameters ν ($0 \le \nu \le 1$) and C to have the solution of the approach is in the form of the following linear expansion of kernel function

$$g(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{k=1}^{m} (\alpha_k^* - \alpha_k) K(\boldsymbol{x}, \boldsymbol{x}_k) + b \,. \qquad (14)$$

In this paper, the Gaussian function is used as the kernel function. Hence, (13) and (14) can be rewritten as

$$g(\boldsymbol{x}, \lambda) = \sum_{k=1}^{\#SV} \lambda_k \exp(-\frac{\left\| \boldsymbol{x} - \boldsymbol{x}_k \right\|^2}{2\sigma_k^2}) + b \quad , \qquad (15)$$

where $\#SV$ is the number of SVs, $\lambda_k = (\alpha_k^* - \alpha_k) \neq 0$ and $\boldsymbol{x}_k$ are SVs. Let $\exp(-\frac{\left\| \boldsymbol{x} - \boldsymbol{x}_0 \right\|^2}{2\sigma_0^2}) = 1$ and $\lambda_0 = b$, (15) can be rewritten as

$$f(x, \lambda) = \sum_{k=0}^{\#SV} \lambda_k \exp(-\frac{\left\| \boldsymbol{x} - \boldsymbol{x}_k \right\|^2}{2\sigma_k^2}). \qquad (16)$$

As mentioned above and comparing (16) with (4), $\#SV$, k, λ_k, $\boldsymbol{x}_k$ and σ_k in (16) can be regarded as the L, j, w_{jl}, $\boldsymbol{m}_j$ and σ_j, in (4), respectively. From the above derivation, we can find that the number of hidden node L, the initial weights w_{jl}, $\boldsymbol{m}_j$, and σ_j, of the ARRBFNs are determined via the ε-SVR and ν-SVR approach.

IV. ARLA FOR THE ARRBFNS

In this paper, the ARLA is applied to train the proposed ARRBFNs. Using the annealing concept in the cost function of robust back-propagation (BP) learning algorithm, can overcome the existing issues in robust BP learning algorithm. A cost function for ARLA is defined here as (5), where $\beta(h)$ is a deterministic annealing schedule acting like the cut-off points to decide that how large errors can be considered as outliers. The proposed ARRBFNs with the ARLA can overcome the problems of initialization and can deal with the time series with outliers. In the ARLA, the properties of the annealing schedule $\beta(h)$ have [4]:

(A) β_{initial}, $\beta(h)$ for the first epoch, has large values;

(B) $\beta(h) \to 0^+$ for $h \to \infty$;

(C) $\beta(h) = k/h$ for any h epoch, where k is a constant.

Based on the gradient-descent kind of learning algorithms, the synaptic weights w_{jl}, the centers $\boldsymbol{m}_j$ and the width σ_j of Gaussian function are updated as

$$\Delta w_{jl} = -\eta \frac{\partial J_N}{\partial w_{jl}} = -\eta \sum_{i=1}^{N} \varphi(e_i; \beta) \frac{\partial e_i}{\partial w_{jl}}, \qquad (17)$$

$$\Delta \boldsymbol{m}_j = -\eta \frac{\partial J_N}{\partial \boldsymbol{m}_j} = -\eta \sum_{i=1}^{N} \varphi(e_i; \beta) \frac{\partial e_i}{\partial \boldsymbol{m}_j}, \qquad (18)$$

$$\Delta \sigma_j = -\eta \frac{\partial J_N}{\partial \sigma_j} = -\eta \sum_{i=1}^{N} \varphi(e_i; \beta) \frac{\partial e_i}{\partial \sigma_j}, \qquad (19)$$

$$\varphi(e_i; \beta) = \frac{\partial \rho(e_i; \beta)}{\partial e_i} = \frac{e_i}{1 + (e_i)^2 / \beta(h)}, \qquad (20)$$

where η is a learning constant.

The learning algorithm of the ARRBFNs for the chaotic time series with outliers is summarized as follows:

Algorithm A :

Step 1: Generate the data from chaotic time series.

Step 2: Using an ε-SVR approach that is depicted by equations (7) through (13), or using an ν-SVR approach that is described by equations (14), to get the initial structure of ARRBFNs.

Step 3: Compute the estimated output and its error by equation (6) for all training data.

Step 4: Decide the values of annealing schedule $\beta(h) = k/h$ for each epoch, where k is set as $2 \cdot max\{|\, e_i \,|_{\text{initial}}\}$.

Step 5: Update w_{jl}, $\boldsymbol{m}_j$ and σ_j of Gaussian function iteratively updated by equations (17) through (20). In this step, the outliers are presented.

Step 6: Compute the robust cost function J_N defined by equation (5).

Step 7: If the termination conditions are not satisfied, then go to Step 3; otherwise stop the learning process.

V. SIMULATION RESULTS

In this section, the root mean square error (RMSE) of the testing data is used to measure the performance of the learned networks. The RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i(t + \Delta t) - \hat{x}_i(t + \Delta t) \right)^2}{N}}, \qquad (21)$$

where $x_i(t + \Delta t)$ is the desired output and $\hat{x}_i(t + \Delta t)$ is the output of the proposed ARRBFNs.

Example:

Mackey–Glass time series is defined as equation (1). The initial values of the system are $x(0) = 1.2$, $\alpha = 0.2$, $\beta = 0.1$, $\gamma = 10$. τ is the time-delay parameter. If $\tau \ge 17$, the time series show the chaotic phenomenon. Compare with Gu and Wang [1] for being convenient, the

proposed networks are chose as

$$\hat{x}(t+6) = \hat{f}(x(t), x(t-6), x(t-12), x(t-18)), t \geq 19 \quad (22)$$

That is, $x(t), x(t-6), x(t-12), x(t-18)$ are selected as the input variables of the ARRBFNs, and $x(t+6)$ as the output variable of the ARRBFNs.

1000 simulation data points are generated from equation (1), the former 500 points with five artificial outliers are selected as the training data points to build the proposed ARRBFNs of the Mackey–Glass time series, and the rest 500 points as the testing data to test the validity of the proposed ARRBFNs. The parameters in ε-SVR are set as C=10, the Gaussian kernel function with $\varepsilon = 0.35$, $\sigma = 0.15$, with the hidden nodes (i.e. the number of SVs) is obtained as 10. Based on the initial structure of the ARRBFNs and the learning constant is 0.05, after 2000 epochs training, the final training output, the error, the prediction output and the corresponding error are shown in Fig. 1 (a) ~ (d), and the final RMSE is 0.0094. Another initial structure is obtained by an ν-SVR approach, the parameters are set as C=1, $\nu = 0.0009$ and $\sigma = 0.15$, with the hidden nodes is obtained as 10. Based on the above initial structure, the learning constant is 0.05, after 2000 epochs training, the results are shown in Fig. 2 (a) ~ (d), and the final RMSE is 0.0096. From the simulation results show that the proposed ARRBFNs can overcome the outliers and attain a good training and prediction.

VI. CONCLUSIONS

In this paper, an ε or ν SVR based the ARRBFNs with ARLA for the prediction of chaotic time series with outliers is developed. We firstly utilize the SVR approaches to determine the number of hidden nodes, the initial parameters of the kernel and the initial weights of the proposed ARRBFNs. Then the ARLA is applied to tunes the parameters of the kernel and the weights of the time series that can against outliers. From the results indicated that the proposed method can be used as a reliable technique for the prediction of chaotic time series with outliers.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Council, Taiwan, R.O.C., under grants NSC96-2628-E-224-007-MY2.

REFERENCES

[1] Gu H, Wang H (2007), Fuzzy prediction of chaotic time series based on singular value decomposition. Applied Mathematics and Computation 185: 1171–1185

[2] Zhang W, Wu ZM, Yang GK (2005), Genetic programming modeling on chaotic time series. Acta Electronica Sinica 33(4):748-751

[3] Harpham C, Dawson CW (2006), The effect of different basis functions on a radial basis function network for time series prediction: A comparative study. Neurocomputing 69:2161–2170

[4] Chuang CC, Su SF, and Hsiao CC (2000), The annealing robust backpropagation (BP) learning algorithm. *IEEE Trans. Neural Networks* 11(5): 1067–1077

[5] Vapnik V (1995), *The Nature of Statistical Learning Theory*, Springer, Berlin

[6] Jeng JT (2006), Hybrid approach of selecting hyper-parameters of support vector machine for regression. *IEEE Trans. on Systems, Man, and Cybernatics* B36:699-709

[7] Smola AJ, Schölkopf B (1998), From regularization operators to support vector kernels. *Neural Inf. Process System* 10:343–349

[8] Schölkopf B, et al. (1997), Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Process* 45(11): 2758–2765

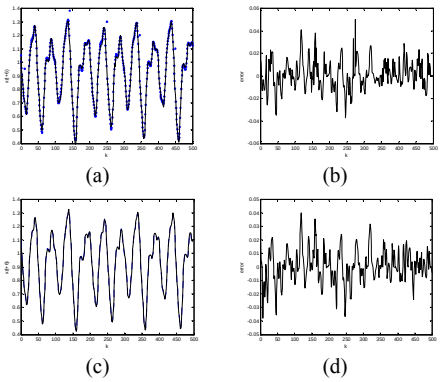

Fig. 1. (a) The final output, (b) the error, (c) the prediction output and (d) the error of prediction for the proposed structure with ε-SVR based under the training data sets contain five artificial outliers

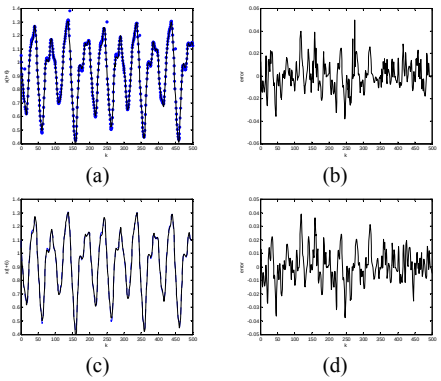

Fig. 2. (a) The final output, (b) the error, (c) the prediction output and (d) the error of prediction for the proposed structure with ν-SVR based under the training data sets contain five artificial outliers