Identification of RBFNs with SVR under censored data

Yu-Yi Fu[1], Yu-Fen Fu[2], Kuo-Lan Su[3] , Chia-Nan Ko[4] and Jhih-Cheng Chang[5]

[1,4,5] Department of Automation Engineering, Nan Kai University of Technology
Tsaotun, Nantou 542, Taiwan
[2]Department of Information Management, China University of Technology
Wunshan, Taipei 116, Taiwan
[3]Department of Electrical Engineering,National Yunlin University of Science and Technology
Douliou, Yunlin 640, Taiwan
(email:[1]t098@nkut.edu.tw;[2]yfen@cute.edu.tw;[3]sukl@yuntech.edu.tw;[4]t105@nkut.edu.tw;[5]cjcb@nkut.edu.tw)

Abstract: In this paper, we proposed RBFNs with SVR to identify a distribution under censored data. Radial basis function networks (RBFNs) with one hidden layer and rapid convergence speed used to identify system generally. Support vector regression (SVR) with optimal quadratic programming to determine the number of hidden nodes and the initial parameters of kernel and the initial weights of RBFNs. By annealing robust learning algorithm to tune the parameters of kernel and the weights and to overcome the error measurement due to data censored. The simulation result of bivariate normal distribution identification under censored data shows the feasibility of proposed method.

Keywords: Radial basis function networks, Support vector regression, censored data, bivariate normal distribution.

1. INTRODUCTION

Interval censored data analysis is important in survival analysis and biomedical statistics. It has lower limit and upper limit, the lower limit replaces the data values when the data values are lower or equal to lower limit, the upper limit replaces the data values when the data values are greater or equal to upper limit. Therefore the data we get are not correct data but censored data. Zheng et al [1] proposed Hybrid Monte Carlo Markov chain and Hamiltonian method to estimate parameters of interval censored data. Cheng and Mordeson [2] investigated the fuzzy estimation of the parameter of underlying probability distribution for total failure time of censored data. In this article we proposed RBFNs with SVR to identify a distribution under the censored data.

Radial basis function networks (RBFNs) with one hidden layer and rapid convergence speed used to identify system and predict generally [3-7]. But the number of hidden nodes, the initial parameters of the kernel and the initial weights of the networks not decided mathematically yet. Besides, the data we obtained sometimes contain the error measurement due to data censored.

Vapnik [8] proposed Support vector regression (SVR) approach in 1995, by the ε 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. That is, the SVR uses the quadratic programming optimization to determine the initial structure of the traditional RBFNs.

In this article, the purpose is to identify a system under censored data in one of the variables. First, an ε SVR is used to determine the number of hidden nodes, the initial parameters of the kernel, and the initial weights of the RBFNs. Then the algorithm is applied to tune the parameters of radial basis functions and the synaptic weights. It is expected that the proposed method has fast convergence speed and the ability coping with censored data can identify perfectly.

2. RBFNs FOR THE IDENTIFICATION OF THE NONLINEAR SYSTEM

Assume that the unknown nonlinear system is expressed by

$$y(t+1) = f(y(t),\cdots,y(t-n),u(t),\cdots,u(t-m)), \quad (1)$$

where $y(t)$ is the output of the system, $u(t)$ is the input of the system, $f(\cdot)$ is the unknown nonlinear function to be estimated by a neural network, and n and m are the structure orders of the system. Our purpose is to find a identification model

$$\hat{y}(t+1) = \hat{f}(y(t),\cdots,y(t-n),u(t),\cdots,u(t-m)) \quad (2)$$

to approximate $y(t+1)$ as close as possible.

A structure of the RBFNs consists of an input layer, a hidden layer of radial basis functions and a linear output layer. When the radial basis functions are chosen as Gaussian functions, an RBFNs can be expressed in the form

$$\hat{y}_j(t+1) = \sum_{i=0}^{L} w_{ij} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{m}_i\|^2}{2\sigma_i^2}\right) \quad (3)$$

where $\hat{y}_j$ is the jth output, $\boldsymbol{x} = (y(t), y(t-1),\cdots,y(t-n),u(t),u(t-1),\cdots,u(t-m))$ is the input to the neural networks, $w_{ij}, 0 \le i \le L, 1 \le j \le p$, are the synaptic weights, $\boldsymbol{m}_i, 0 \le i \le L$, and $\sigma_i, 0 \le i \le L$, are the centers and the widths of Gaussian functions respectively, and L is the number of the Gaussian functions, in which we can find that L also denotes the number of hidden nodes.

3. INITIAL STRUCTURE BY SVR APPROACH

An SVR approach is used to approximate an unknown function from a set of (input, output) samples $\{(\boldsymbol{x}_i, y_i), i = 1, \cdots, N\}$. Suppose that a set of basis functions $\{g(\boldsymbol{x}_k)\}_{k=1}^{N}$ 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

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{k=1}^{N} \theta_k g(\boldsymbol{x}_k) + b, \tag{4}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ is a parameter vector to be identified and b is a constant to be found.

Vapnik [4] firstly proposed the ε-SVR approach. The solution for the theme is to find $f(\boldsymbol{x}, \boldsymbol{\theta})$ that minimizes

$$R(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} L_{\varepsilon}(y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})), \tag{5}$$

subject to the constraint

$$\|\boldsymbol{\theta}\|^2 < C, \tag{6}$$

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

$$L_{\varepsilon}(e) = \begin{cases} 0 & \text{for } |e| \le \varepsilon \\ |e| - \varepsilon & \text{otherwise} \end{cases}, \tag{7}$$

for some previously chosen nonnegative number ε.

By using the Lagrange multiplier method, it was shown in [8] that the minimization of (5) leads to the following dual optimization problem.
Minimize

$$\begin{aligned} Q(\alpha, \alpha^*) = {} & \varepsilon \sum_{r=1}^{N} (\alpha_r + \alpha_r^*) - \sum_{r=1}^{N} y_r (\alpha_r^* - \alpha_r) \\ & + \frac{1}{2} \sum_{r,s=1}^{N} (\alpha_r^* - \alpha_r)(\alpha_s^* - \alpha_s) \langle g(\boldsymbol{x}_r) \cdot g(\boldsymbol{x}_s) \rangle. \end{aligned} \tag{8}$$

subject to the constraint

$$\sum_{r=1}^{N} \alpha_r^* = \sum_{r=1}^{N} \alpha_r, \quad 0 < \alpha_r^*, \alpha_r < C \quad \text{for } r = 1, 2, \cdots, N. \tag{9}$$

It proposed by Vapnik [8] and Smola et al. [9] and the inner product of basis function $\langle g(\boldsymbol{x}_r) \cdot g(\boldsymbol{x}_s) \rangle$ is replaced via the kernel function

$$K(\boldsymbol{x}_r, \boldsymbol{x}_s) = \langle g(\boldsymbol{x}_r) \cdot g(\boldsymbol{x}_s) \rangle, \tag{10}$$

Hence the optimization of (8) is rewritten as
Minimize

$$\begin{aligned} Q(\alpha, \alpha^*) = {} & \varepsilon \sum_{r=1}^{N} (\alpha_r + \alpha_r^*) - \sum_{r=1}^{N} y_r (\alpha_r^* - \alpha_r) \\ & + \frac{1}{2} \sum_{r,s=1}^{N} (\alpha_r^* - \alpha_r)(\alpha_s^* - \alpha_s) K(\boldsymbol{x}_r, \boldsymbol{x}_s). \end{aligned} \tag{11}$$

It was shown in Vapnik [8] that the solution of the SVR approach is in the form of the following linear expansion of kernel function

$$f(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{k=1}^{N} (\alpha_k^* - \alpha_k) K(\boldsymbol{x}, \boldsymbol{x}_k) + b. \tag{12}$$

This means that the parameter θ_k in equation (4) can be represented as $\sum_{k=1}^{m} (\alpha_k^* - \alpha_k) g(\boldsymbol{x}_k)$. 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).

In this paper, the Gaussian function is used as the kernel function. And let $f_s(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}, \lambda) - b$, hence, (9) can be rewritten as

$$f_s(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{k=1}^{\#SV} \lambda_k \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_k\|^2}{2\sigma_k^2}\right), \tag{13}$$

where $\#SV$ is the number of SVs, $\lambda_k = (\alpha_k^* - \alpha_k) \neq 0$ and $\boldsymbol{x}_k$ are SVs.

From equation (13) and (3), $\#SV$, k, λ_k, $\boldsymbol{x}_k$ and σ_k in (13) can be regarded as the $L, i, w_{ij}, \boldsymbol{m}_i$ and σ_i in (3), respectively. From the above derivation, we can find that the number of hidden nodes L, the initial weights w_{ij}, the initial parameters $\boldsymbol{m}_i$ and σ_i, of the RBFNs are determined via the SVR approach.

4. ANNEALING ROBUST LEARNING ALGORITHM FOR UPDATING PARAMETERS

When utilizing the RBFNs for the identification of systems, the goal is to minimize

$$J_N(h) = \frac{1}{N} \sum_{i=1}^{N} \rho[e_k(h); \beta(h)], \tag{14}$$

where

$$e_k(h) = y(k) - \hat{y}_s(\boldsymbol{x}_k), \tag{15}$$

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

$$\rho[e_k; \beta] = \frac{\beta}{2} \ln[1 + \frac{(e_k)^2}{\beta}], \tag{16}$$

where $\beta(h)$ is a deterministic annealing schedule acting like the cut-off points. In this paper, the annealing robust learning algorithm (ARLA) is applied to train the proposed RBFNs. In the ARLA, the properties of the annealing schedule $\beta(h)$ have [10]:

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

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

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

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

$$\Delta w_{ij} = -\eta \frac{\partial J_N}{\partial w_{ij}} = -\eta \sum_{k=1}^{N} \varphi(e_k; \beta) \frac{\partial e_k}{\partial w_{ij}} \tag{17}$$

$$\Delta \boldsymbol{m}_i = -\eta \frac{\partial J_N}{\partial \boldsymbol{m}_i} = -\eta \sum_{k=1}^{N} \varphi(e_k;\beta) \frac{\partial e_k}{\partial \boldsymbol{m}_i} \quad (18)$$

$$\Delta \sigma_i = -\eta \frac{\partial J_N}{\partial \sigma_i} = -\eta \sum_{k=1}^{N} \varphi(e_k;\beta) \frac{\partial e_k}{\partial \sigma_i} \quad (19)$$

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

where η is a learning constant.

5. SIMULATION RESULTS

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}, \quad (21)$$

where y_i is the desired output and $\hat{y}_i$ is the output of the proposed method.

Example:

A bivariate normal distribution is defined as

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\}, \quad (22)$$

where μ_x and μ_y are the mean of x and y, σ_x^2 and σ_y^2 are the variance of x and y, and ρ is the correlation coefficient of x and y. We suppose that x data have censored, and the upper limit is equal to 3, the lower limit is equal to -3. The diagram of exact data and censored data are shown in Fig. 1 and Fig. 2. The proposed networks are chosen as

$$\hat{y}(t+1) = \hat{f}(y(t), y(t-1), x(t), x(t-1), x(t-2)). \quad (23)$$

50 simulation data points are generated from equation (19). The parameters in ε-SVR are set as C=10, the Gaussian kernel function with $\varepsilon = 0.95$, $\sigma = 2$, with the hidden nodes (i.e. the number of SVs) is obtained as 15. Based on the initial structure and the learning constant is 0.05, after 3000 epochs training, the final training output, the error are shown in Fig. 3 and Fig. 4, and the final RMSE is 0.0242. From the simulation results show that the proposed method can overcome the error measurement due to censored data and attain a good identification.

6. CONCLUSIONS

In this paper, an SVR based the RBFNs with ARLA for the identification of bivariate normal distribution is developed. We firstly utilize the SVR approach to determine the number of hidden nodes, the initial parameters of the kernel and the initial weights of the proposed RBFNs. Then the ARLA is applied to tunes the parameters of the kernel and the weights of the bivariate normal distribution. From the result indicated that the proposed method can be used as a reliable technique for the training of bivariate normal distribution.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Council, Taiwan, R.O.C., under grants NSC101-2221-E-252-009.

REFERENCES

[1] Zheng G, Liu J, Zhang G (2010), A Novel Hybrid MCMC Method for Interval-Censored Data, International Conference on Computer Application and System Modeling, Taiyuan, China, Oct 22-24, 2010, pp. V12-66 - V12-69

[2] Cheng SH, Mordeson JN (2008), Fuzzy Confidence Estimation of the Parameter Involving in the Distribution of the Total Time on Test for Censored Data, Fuzzy Information Processing Society, Annual Meeting of the North American, May 19-22, 2008, pp. 1 - 5

[3] Fu YY, Wu CJ, Ko CN, Jeng JT (2011), Radial basis function networks with hybrid learning for system identification with outliers, Applied Soft Computing, 11: 3083–3092

[4] Fu YY, Wu CJ, Ko CN, Jeng JT, Lai LC (2009), ARRBFNs with SVR for prediction of chaotic time series with outliers, Artif Life Robotics, 14:29–33

[5] Liu J, Ling D, Wang S (2012), Ammunition Storage Reliability Forecasting Based on Radial Basis Function Neural Network, International Conference on Digital Object Identifier, Chengdu, China, June 15-18, 2012, pp. 599 – 602

[6] Venkateswarlu RLK, Kumari RV, Jayasri GV (2011), Speech recognition using Radial Basis Function neural network, International Conference on Electronics Computer Technology, Kanyakumari, India, April 8-10, 2011, pp. 441 – 445

[7] Hao Y, Tiantian X, Paszczynski S, Wilamowski BM (2011), Advantages of Radial Basis Function Networks for Dynamic System Design, IEEE Transactions on Industrial Electronics, 58 (12): 5438 – 5450

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

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

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

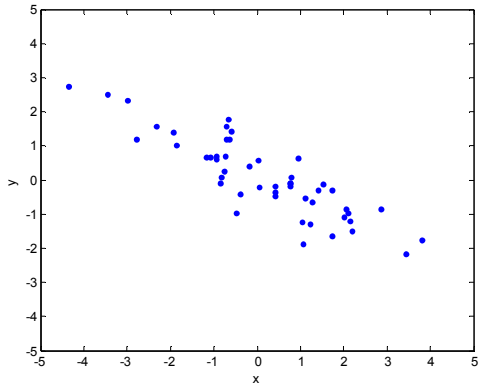

Fig. 1. The diagram of example for exact data.

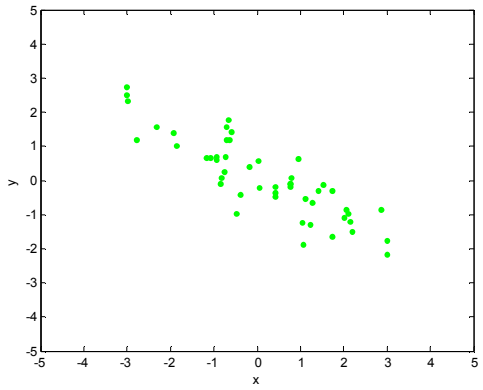

Fig. 2. The diagram of example for censored data.

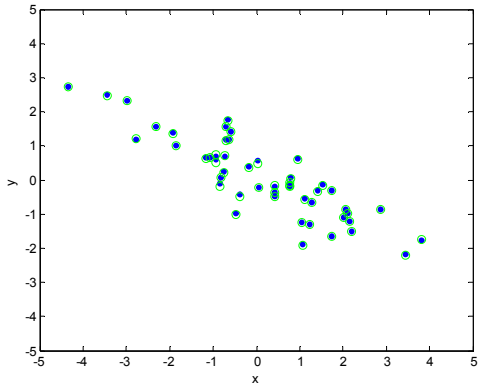

Fig. 3. The diagram of example (dot scatter) and the final result of the identification example (circle scatter) under censored data.

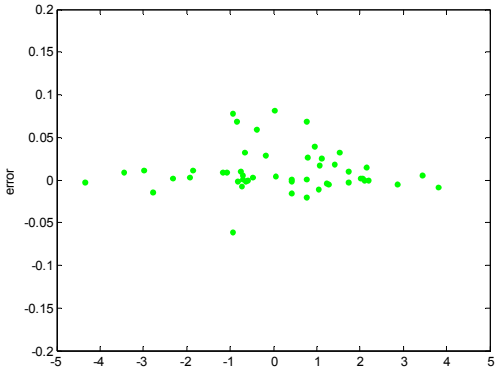

Fig. 4. The error of training under censored data.