Adaptive Support Vector Regression

Bao Rong Chang
Department of Computer Science and Information Engineering
National Taitung University
684 Chunghua Rd., Sec. 1, Taitung, Taiwan 950
brchang@cc.nttu.edu.tw

Abstract- *A novel scheme for training support vector regression (SVR) with self-adaptive mechanism, called adaptive SVR (ASVR), is introduced herein to tune automatically user-defined free parameters, C and ε-tube, optimally in SVR. In the traditional support vector regression, two free parameters, C and ε-tube, are set in the default values, infinite and zero, respectively. However, this default setting is not optimal one for any SVR forecasting applications, and thus it may encounter some big residual errors leading to worst prediction accuracy. In order to best fit SVR model, adaptive support vector regression is applied to tuning free parameters C and ε-tube optimally. In such this way, the generalization capability can be enhanced in SVR model so as to improve prediction accuracy highly.*

Keywords: support vector regression, adaptive support vector regression, generalization capability.

1. Introduction

Well-known methods, such as ARMA [1], artificial network [2], or fuzzy inference system [2], have been widely employed into many forecasting on the scientific or economic applications. However, these methods usually require a lot of observed data for fitting their model to build more accurate approach [3] so that these models probably do not be suitable for the short-term task using only scarce data for modeling. Both grey model (GM) [4] and cumulated 3 points least squared linear polynomial (C3LSP) [5] have introduced acts contrary to the aforementioned just acquiring a few data for modeling without training process, implying a simple and fast short-term task. Notwithstanding it is good to short-term task, the overshooting problem [5] often occurs in grey model and resulted in a big residual error around the region of turning points. In contrast, C3LSP model has encountered the underestimated results. Alternatively, support vector regression (SVR) [6] is very useful to process the short-term task with sparse data given and possibly producing a best-fit model to avoid the overshooting problem. If outliers exist somewhere in the given training data set, the trained SVR (might be deviated one) might not work best-fit in the prediction or estimation. It is interested that fewer literatures have mentioned about the user-defined free parameters C and ε-tube in SVR for years. How to scheme a self-adaptive method for tuning two user-defined free parameters automatically is proposed in this study.

2. Learning Algorithm for Support Vector Regression

The foundations of Support Vector Machines (SVM) have been developed by Vapnik [7], and are gaining popularity due to many attractive features, and promising empirical performance. The formulation embodies the Structural Risk Minimization (SRM) principle, as opposed to the Empirical Risk Minimization (ERM) approach commonly employed within statistical learning methods. SRM minimizes an upper bound on the generalization error, as opposed to ERM that minimizes the error on the training data. It is this difference that equips SVMs with a greater potential to generalize, which is our goal in statistical learning. The SVM can be applied to both classification and regression problems [8].
Support Vector Machines along with neural networks as one of the standard tools for machine learning and data mining [8]. Initially developed for solving classification problems, SV technology can also be successfully applied in regression, i.e. functional approximation, problems. Unlike pattern recognition problems, where the desired outputs are discrete values like Booleans, here there are real-valued functions [9]. We consider approximating functions solved by support vector regression (SVR) as the form of

$$f(x,w)=\sum_{i=1}^{l} w_i\phi(x), \qquad (1)$$

where $\phi(x)$ are denoted by features. In order to introduce all relevant and necessary concept of SV regression in a gradual way, linear regression is considered first.

$$f(x,w) = w^T x + b \tag{2}$$

Furthermore, Vapnik introduced a general type of loss function, namely, error, the linear loss function with ε-insensitivity zone:

$$|y - f(x,w)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(x,w)| \le \varepsilon, \\ |y - f(x,w)| - \varepsilon & \text{otherwise} \end{cases} \tag{3}$$

A new empirical risk is introduced for performing SVM regression.

$$R_{emp}(w,b)_\varepsilon = \frac{1}{l}\sum_{i=1}^{l} |y_i - w^T x_i - b|_\varepsilon \tag{4}$$

According to the learning theory of SVMs, the objective is to minimize the empirical risk and norm-squared of weight vector simultaneously. Thus, estimate a linear regression hyperplane $f(x,w) = w^T x + b$ by minimizing

$$R(w,\xi,\xi^*) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l}\xi_i + \sum_{i=1}^{l}\xi_i^*\right), \tag{5}$$

under constrains

$$y_i - w^T x_i - b \le \varepsilon + \xi_i,\ i = 1,...,l \tag{6}$$

$$w^T x_i + b - y_i \le \varepsilon + \xi_i^*,\ i = 1,...,l \tag{7}$$

$$\xi_i \ge 0,\ i = 1,...,l \tag{8}$$

$$\xi_i^* \ge 0,\ i = 1,...,l \tag{9}$$

where the constant C influences a trade-off between an approximation error and an estimation error decided by the weight vector norm $\|w\|$, and this design parameter is chosen by the user. ξ_i and ξ_i^* are slack variables as the measurement upper bound and lower bound of outputs. This quadratic optimization is equivalence to apply Karush-Kuhn-Tucker (KKT) conditions for regression in which maximizing dual variables Lagrangian $L_d(\alpha,\alpha^*)$:

$$L_d(\alpha,\alpha^*) = -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T x_j - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i, \tag{10}$$

subject to constraints

$$\sum_{i=1}^{l}\alpha_i = \sum_{i=1}^{l}\alpha_i^*, \tag{11}$$

$$0 \le \alpha_i \le C,\quad i = 1,...,l \tag{12}$$

$$0 \le \alpha_i^* \le C,\quad i = 1,...,l \tag{13}$$

After calculating Lagrange multipliers α_i and α_i^*, find an optimal desired weights vector of the regression hyperplane as

$$w_0 = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i \tag{14}$$

and an optimal bias of regression hyberplane as

$$b_0 = \frac{1}{l}\left(\sum_{i=1}^{l}(y_i - x_i^T w_0)\right). \tag{15}$$

In non-linear cases for regression, the kernal function, for typical instances, polynomial, RBF, or sigmoid function, will be adopt to replace the scale product $x_i^T x_j$ with $K(x_i, x_j)$ in Eq. (10).

If the term $\beta_i = (\alpha_i - \alpha_i^*)$ is defined in training data set, the output of SVR can be obtained with new input pattern z_i [9].

$$y = g\beta + b_0 \tag{16}$$

where the vector g is constructed by

$$g = z_i^T x,$$

and matrix x stands for patterns in training data set as well as vector z_i represents new input pattern.

$$x = [x_1 x_2 ... x_l]$$

$$z_i = [z_{i1}, z_{i2}, ..., z_{iN}]^T$$

3. Adaptive Support Vector Regression (ASVR)

3.1 Definition of ASVR Notations

The specific notations that are utilized to form adaptive support vector regression (ASVR) proposed in this study will be defined and described clearly in the following statements. Those are the total absolute differences κ, the coefficient of oscillation μ, the first-degree tail rate ρ, the second-degree tail rate ϑ, the tail weight σ, the q-base q, and the gauge factor v listed in the subsection 3.1. All of them are applied to tuning ε-tube in SVR learning automatically. The formulation of those notations is listed in the subsection 3.2.

3.1.1. Slant rate. A straight line is derived from the most recent specified data points. The slant of this straight line is enlarged by tangent function in order to emphasize the most recent observed data distribution on a time series. This slant rate is denoted as ς on Eq. (20).

3.1.2. Total absolute differences. The total differences κ on Eq. (25) denotes sum of absolute components in the normalized difference sequence Λ_{N-1} on Eq. (24).

3.1.3. Coefficient of oscillation. The coefficient of oscillation μ on Eq. (26) represents the ratio of absolute sum of components in the normalized difference sequence to the total differences κ. It means what percent of amplitude up-and-down (including down-and-up) changes in the difference sequence and interprets the oscillation phenomenon existing over the sampled data set indeed. A very small real value, say 10^{-6}, used as a padding number is placed in both denominator and numerator in order to avoid a case of dividing by zero.

3.1.4. First-degree tail rate. The first-degree tail rate ρ on Eq. (27) stands for the ratio of the last difference to mean difference, and shows the sign of first-degree of the most recent inertia with respect to the current trend. This rate may provide significant information to support the future planning when we want to use the current sampled data set to carry out n-steps-look-ahead prediction.

3.1.5. Second-degree tail rate. The second-degree tail rate ϑ on Eq. (28) expresses the ratio of average of the last two differences to mean difference, and shows the sign of second-degree of the most recent inertia with respect to the current trend. We think of this rate as information of another one used to support the future planning. In other words, both first-degree tail rate ρ and second-degree tail rate ϑ are taken into account as n-steps-look-ahead prediction is proceeding.

3.1.6. Tail weight. The dilate operation, hyperbolic tangent function, on the specific value that is calculated by squaring the percent of coefficient of oscillation on the first-degree tail rate is emphasized herein to form a tail weight σ on Eq. (29) meaning proportionality over the oscillated portion of a sequence. This subtle measure insights the possibility of future outcome, especially applied to single-step-look-ahead forecasting, so that it gives enriched additional information referred by the future planning highly in hope.

3.1.7. Q-base. The q-base q on Eq. (30) is defined as an exponential function with negative power, and the power is assigned either a tail weight scaled by hyperbolic tangent function of φ or a first-degree tail rate based on the dependency of oscillation coefficient μ . This q-base will be used for computing the gauge factor v on Eq. (31), and then this gauge factor will be employed to adjust ε-tube on Eq. (35).

3.1.8. Gauge factor. The gauge factor v on Eq. (31) describes a specific tuning factor used to adapt optimal ε-tube, and ε-tube is one of free parameters should be applied into the training process of support vector regression. In order to speed-up the training process of support vector regression, adaptation of the gauge factor in fact contributing a great help to obtain ε-tube on Eq. (35) is explored herein and designated a way to compute its appropriate value through a formula that is proposed in this paper.

3.2 ASVR Learning Algorithm

A scheme of regularization for the adaptation on free parameter ε-tube in SVR constrained optimization has been introduced herein so that it can reduce the computation burden and converge to the approximately optimal solution faster. Let the original data sequence to be as follows.

$$X_N = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)\} \quad (17)$$

The novel regularization algorithm proposed in this study is proceeding in the following several steps.

Step 1: normalization of data sequence on Eq. (17) as follows.

$$Y_N = \begin{bmatrix} x^{(0)}(1) \\ x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \ y_N = \frac{Y_N}{\max_i |x^{(0)}(i)|} = \begin{bmatrix} \tilde{x}^{(0)}(1) \\ \tilde{x}^{(0)}(2) \\ \tilde{x}^{(0)}(3) \\ \vdots \\ \tilde{x}^{(0)}(n) \end{bmatrix} \quad (18)$$

Step 2: constructing a simple linear regression from the most recent normalized data points, $\tilde{x}^{(0)}(1), \tilde{x}^{(0)}(n), ..., \tilde{x}^{(0)}(n)$ as well as a slant rate ς to the most recent specified data points is also measured.

$$\tilde{x}^{(0)}(k) = \varphi k + \psi \ , \ k = 1,2,...,n \quad (19)$$

$$\varsigma = \tan(|\varphi|) \quad (20)$$

where φ is the slope and ψ is the bias in this line.

Step 3: Eq. (19) turns out to be a normal equation and its solution Θ to this least squared problem is obtained.

$$\tilde{X} = \Omega\Theta \ , \quad (21)$$

$$\Theta = (\Omega^T \Omega)^{-1} \Omega^T \tilde{X} \ , \quad (22)$$

where $\tilde{X} = [\tilde{x}^{(0)}(1), \tilde{x}^{(0)}(2), ..., \tilde{x}^{(0)}(n)]^T$

$$\Omega = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ \vdots & \vdots \\ n & 1 \end{bmatrix}, \text{ and } \Theta = [\varphi, \psi]^T \quad (23)$$

Step 4: taking difference on Eq. (18) led to a difference sequence Δ_{N-1} followed by a normalized difference sequence Λ_{N-1} which is normalized from Δ_{N-1} .

$$\Delta_{N-1} = \begin{bmatrix} x^{(0)}(2) - x^{(0)}(1) \\ x^{(0)}(3) - x^{(0)}(2) \\ x^{(0)}(4) - x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) - x^{(0)}(n-1) \end{bmatrix} = \begin{bmatrix} \delta^{(0)}(2) \\ \delta^{(0)}(3) \\ \delta^{(0)}(4) \\ \vdots \\ \delta^{(0)}(n) \end{bmatrix}, \Lambda_{N-1} = \frac{\Delta_{N-1}}{\max_i |\delta^{(0)}(i)|} = \begin{bmatrix} \tilde{\delta}^{(0)}(2) \\ \tilde{\delta}^{(0)}(3) \\ \tilde{\delta}^{(0)}(4) \\ \vdots \\ \tilde{\delta}^{(0)}(n) \end{bmatrix} \quad (24)$$

Step 5: according to the above difference sequence Δ_{N-1} and normalized difference sequence Λ_{N-1} , total absolute differences κ , coefficient of oscillation μ , first-degree tail rate ρ , second-degree tail rate ϑ , and tail weight σ are calculated.

$$\kappa = \sum_{i=2}^{n} |\tilde{\delta}^{(0)}(i)| \quad (25)$$

$$\mu = \frac{\left| \sum_{i=2}^{n} \tilde{\delta}^{(0)}(i) \right| + e}{\kappa + e} \ , \ e = 10^{-6} \quad (26)$$

$$\rho = \frac{|\tilde{\delta}^{(0)}(n)|}{\frac{\kappa}{n-1}} \quad (27)$$

$$\vartheta = \frac{|\tilde{\delta}^{(0)}(n-1) + \tilde{\delta}^{(0)}(n)|}{2\frac{\kappa}{n-1}} \quad (28)$$

$$\sigma = \tanh(\rho^2 / \mu) \quad (29)$$

Step 6: based on Eq. (25)-(29), two specific factors, q-base q and gauge factor v, are determined.

$$q = \exp\left\{- \tanh(|\varphi|)^{1/2} \sigma^{(1-u(\mu-1))} - u(\mu-1)\cdot\rho^{\vartheta}\right\} \quad (30)$$

$$v = \begin{cases} q^{(1-pulse(\kappa))} & if \quad sign(\varphi) = 1 \\ \left[q\mu^{(1-u(\mu-1))}\vartheta^{2u(\mu-1)}\right]^{(1-pulse(\kappa))} & if \quad sign(\varphi) = -1 \end{cases} \quad (31)$$

where the unit-step function $u(t)$ is written by

$$\mu(t-t_0) = \begin{cases} 0, & if \ t < t_0 \\ 1, & if \ t \geq t_0 \end{cases} \quad (32)$$

pulse function is represented as

$$pulse(t-t_0) = \begin{cases} 0, & if \ t \neq t_0 \\ 1, & if \ t = t_0 \end{cases} \quad (33)$$

and the sign function is defined as

$$sign(h) = \begin{cases} 1, & if \ h \geq 0 \\ -1, & if \ h < 0 \end{cases} \quad (34)$$

Step 7: the absolute difference between maximum and minimum observed value is set. The half width of this absolute difference has tuned by gauge factor v in Eq. (35) and then ε-tube is established as follows.

$$\varepsilon = v \times \frac{|Max(X_N) - Min(X_N)|}{2}, \quad (35)$$

Once the value of v has been determined, the ε in Eq. (35) is set, and then the constrained optimization on Eq. (10)-(13) will start for several iterations to search the optimal w_0 and b_0 on Eq. (14)-(15). In this subsection, the goal is to search the appropriate free parameter ε-tube and C such that a fast convergence to the optimal w_0 and b_0 can be obtained. In support vector regression, an increase of value of parameter C will highly penalizes the big empirical error while an increase of value of ε-tube will reduce the support vectors to loose the bound of empirical error [2]. Therefore, how to deal with a trade-off between C and ε-tube so as to achieve the optimal generalization in SVR is become a very important topic. In this research, the relationship between ε-tube and C, we proposed, can be constructed in the basis of modified Bessel function of second kind with the order n as expressed below. A specific integer number n is obtained from a function $\lceil \ \rceil$ of the coefficient of the oscillation μ as described on Eq. (26).

$$n = \lceil \mu \rceil \quad (36)$$

where the operator $\lceil \mu \rceil$ is represented as a smallest integer bigger than μ.

$$C = K_n(\varepsilon) = (-1)^{n+1}\{\ln(\varepsilon/2)+\gamma\}I_n(\varepsilon) + \frac{1}{2}\sum_{k=0}^{n-1}(-1)^k(n-k-1)!(\varepsilon/2)^{2k-n} + \frac{(-1)^n}{2}\sum_{k=0}^{\infty}\frac{(\varepsilon/2)^{n+2k}}{k!(n+k)!}\{\Phi(k)+\Phi(n+k)\} \quad (37)$$

$$I_n(\varepsilon) = \sum_{k=0}^{\infty}\frac{(\varepsilon/2)^{n+2k}}{k!\Gamma(n+k+1)} \quad (38)$$

where $\gamma = 0.5772156...$ is Euler's constant and $\Phi(p) = 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{p}$, $\Phi(0) = 0$.

The tunable free parameters in SVR can be done automatically and referred it to as adaptive support vector regression (ASVR).

4. Model Simulation

As shown in Fig. 1, a remarkable benchmark, Mackey-Glass chaotic time [11], is used to test the forecasting performance on ASVR. This data sequence is generated by the following Mackey-Glass time-delay differential equation.

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (39)$$

This time series is chaotic, and so there is no clearly defined period. Here we assume $x(0)=1.2$, $\tau=17$, and $x(t)=0$ for $t<0$ as the initial conditions to apply fourth-order Runge-Kutta method to find the numerical solution to the above MG equation. Comparing with grey model GM(1,1) [4], the forecasting of ASVR, as shown in Fig. 2, 3, and 4, can actually avoid the overshooting results because the appropriate tunable parameters, C and ε-tube, are computed directly by Eq. (18)-(38), as shown in Fig. 5 and Fig. 6, to obtain higher accurate predicted values. However, the GM(1,1) has encountered a severe problem, overshooting phenomenon, and caused big residual errors around turning-point region in that chaotic time series.

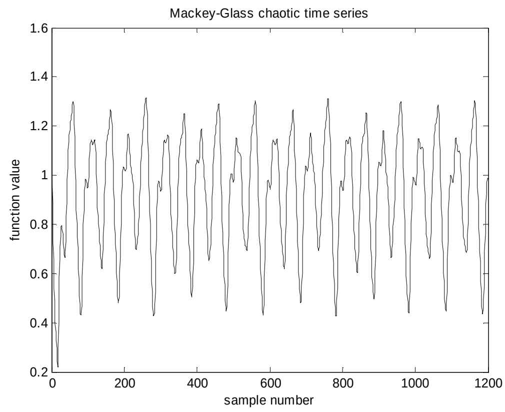

Figure 1. The Mackey-Glass chaotic time series for 1201 sample points as a benchmark of complex time series prediction.

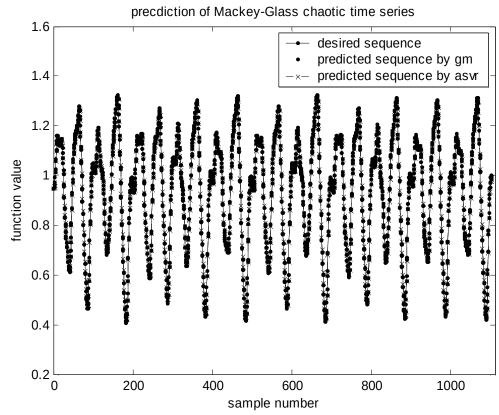

Figure 2. The prediction of Mackey-Glass chaotic time series for 1102 sample points is performed by applying both GM(1,1) indicated by $\bullet$ and ASVR marked by $\times$.

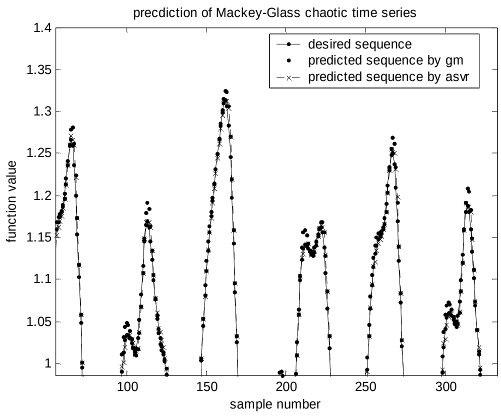

Figure 3. A room-in plot shows that the overshooting phenomenon has occurrence indicated by $\bullet$ from GM(1,1) predicted results around the turning-point region.

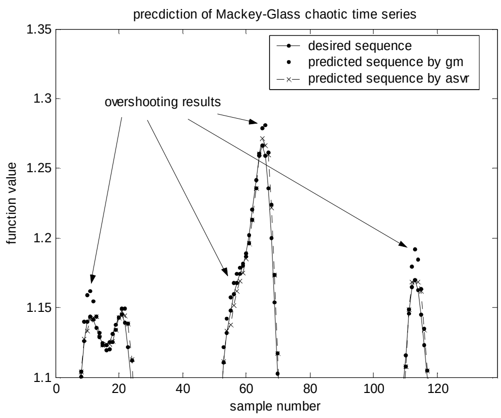

Figure 4. A higher room-in plot has clearly shown the overshooting results from GM(1,1) prediction marked by $\bullet$. In contrast, ASVR has perform very well in the prediction of G-M chaotic time series without any overshooting problem and its predicted results are indicated by $\times$.

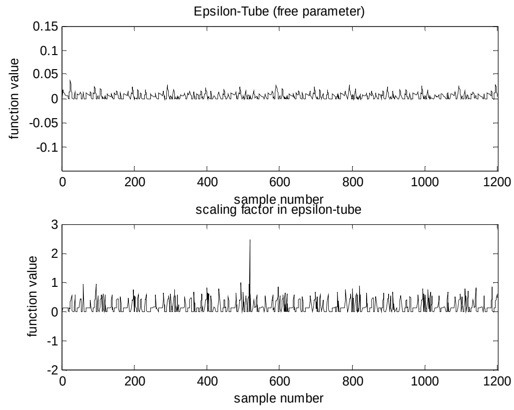

Figure 5. The top plot represents a tuneable epsilon-tube ε used for training SVR. The bottom one shows a scaling factor ν applied to computing ε-tube.

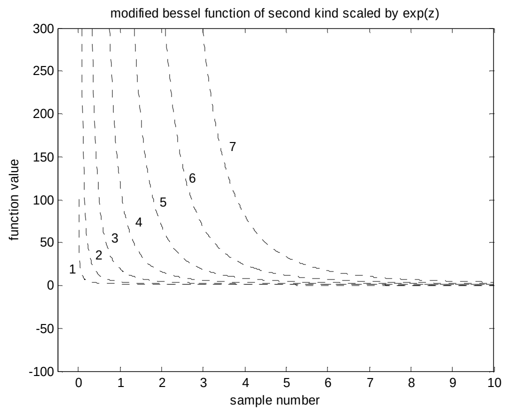

Figure 6. The plotted curves represents modified Bessel function of second kind scaled by exp(z) with different order 1, 2, 3, 4, 5, 6, and 7, indicated aside, respectively.

5. Experimental Results

As shown in Fig. 7 to Fig. 10 or Fig. 11 to Fig. 12, several models are used to compare their performance applied to (1) one-dimensional application: the prediction on international stock price indices, and (2) two-dimensional application: the forecasting on typhoon moving traces. The applied models are follows: grey model GM(1,1| τ) (GM), cumulated 3-point least squared linear polynomial (C3LSP), auto-regression moving-average (ARMA), radial basis function neural network (RBFNN), adaptive support vector regression (ASVR). In these experiments, the most recent four actual values is considered as a set of input data used for modeling to predict the next desired output. As the next desired value is obtained, the first value in the current input data set is discarded and joins the latest observed value to form a new input data set for the use of next prediction. The predictions of international stock price indices for four areas (U.S.A New York Dow Jones, Taiwan

TAIEX, Japan Nikkei Index, and Korea Comp. Index) [10] have been experimented as shown in Fig. 1 to Fig. 4. The accuracy of prediction method is also compared and the summary of this experiment is listed in Table 1, and the goodness of model fitting is tested by Q-test successfully due to averaged p-value (0.4406) greater than level of significance (0.05) [11]. The forecast of typhoon moving trace is a very important issue herein provided that two typhoon moving traces, Nari and Toraji typhoons [12], have been taken for forecasting their future moving position as shown in Fig. 11 and Fig. 12. Table 2 has listed the summary of forecasting accuracy for the comparison between methods, and. the goodness of model fitting is tested by Q-test successfully due to averaged p-value (0.2752) greater than level of significance (0.05).

Table 1. The mean squared error (MSE) between the desired values and the predicted results for international stock price indexes is up to 41 months from Aug. 2000 to Dec. 2003. (unit=10^5) (p-value 0.4406 >0.05 in Q-test of goodness of model fitting)

Methods	N Y- D.J. Industry Index	Taiwan TAIEX Index	Japan Nikkei Index	Korea Composite Index	Average of MSE
GM	4.0577	2.8018	4.7121	0.048139	2.9049
C3LSP	3.4603	3.0014	6.4032	0.048856	3.2284
ARMA	7.4955	5.5694	7.1935	0.085545	5.0860
RBFNN	2.5457	2.6482	4.4123	0.045783	2.4130
ASVR	2.3300	2.0803	3.6498	0.029807	2.0225

Note: method abbreviation
1. GM- GM(1,1| τ) Model
2. C3LSP- Cumulated 3-point Least Squared Linear Prediction
3. ARMA- Autoregressive Moving-Average
4. RBFNN- Radial Basis Function Neural Network
5. ASVR-Adaptive Support Vector Regression

Table 2. The mean squared error (MSE) between the desired values and the predicted results for Nari typhoon moving trace during 6-19, September 2001 and Toraji typhoon moving trace during 28-31, July 2001 (p-value 0.2752>0.05 in Q-test of goodness of model fitting)

Methods	Nari Typhoon Moving Trace	Toraji Typhoon Moving Trace	Average of MSE
GM	0.0648	0.0594	0.0621
C3LSP	0.3698	0.9682	0.6690
ARMA	0.2165	0.1438	0.1802
RBFNN	0.4647	1.9409	1.2028
ASVR	0.0609	0.0407	0.0508

6. Concluding Remarks

The traditional support vector regression is a remarkable model especially applied to the non-periodic short-term forecasting under the condition of scarce data sequence; however, the default setting for free parameters might cause a trained model rather than best-fit one and lead to bad performance due to not sufficient generalization capability. This study introduces a scheme of adaptive support vector regression (ASVR) self-tuning the user-defined free parameters, C and ε -tube, optimally. In such this way, enhancing generalization capability is realized so as to achieve best performance.

7. Acknowledgements

This work is fully supported by the National Science Council, Taiwan, Republic of China, under grant number **NSC 93-2218-E-143-001**.

References

[1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis: Forecasting & Control, New Jersey: Prentice-Hall, 1994.

[2] V. Kecman, Learning and Soft Computing, Massachusetts : MIT Press, 2001.

[3] F. X. Diebold, Elements of Forecasting, Cincinnati: South-Western, 1998.

[4] J. L. Deng, Control Problems of Grey System, System and Control Letter, Vol. 1, No. 5, pp. 288-94, 1982.

[5] B. R. Chang, and S. F. Tsai, "A Grey-Cumulative LMS Hybrid Predictor with Neural Network Based Weighting Mechanism for Forecasting Non-Periodic Short-Term Time Series," IEEE International Conference on Systems, Man, and Cybernetics, Vol. 6, WA2P3, 2002.

[6] ISIS Research Group, Support Vector Machines, http://www.isis.ecs.soton.ac.uk/resources/svminfo/, 2001.

[7] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[8] N. Cristianini, and J. Shawe-Taylor, An Introduction to Support Vector Machines (and other kernel-based learning methods), London: Cambridge University Press, 2000.

[9] S. R. Gunn, Support Vector Machines for Classification and Regression. Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

[10] International Stock Price Index, FIBV FOCUS MONTHLY STATISTICS, 2002.

[11] G. M. Ljung, and G. E. P. Box, "On a Measure of Lack of Fit in Time Series Models," Biometrika, Vol. 65, pp. 67-72, 1978.

[12] Typhoon Moving Trace Inquire, Central Weather Bureau, Taipei, Taiwan, 2001.

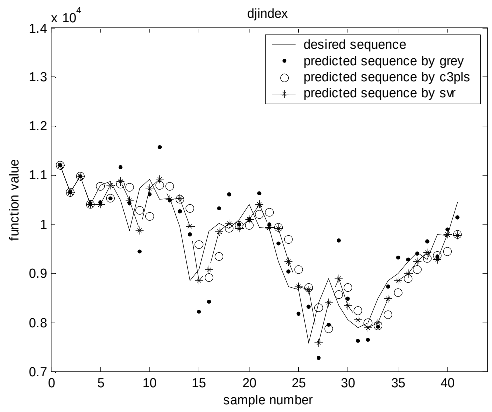

Figure 7. Forecasts of monthly N.Y.D.J. Indus. Index for 41 months from Aug. 2000 to Dec. 2003.

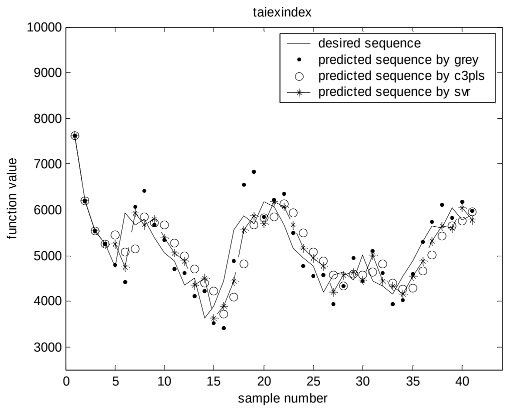

Figure 8. Forecasts of monthly Taiwan TAIEX index for 41 months from Aug. 2000 to Dec. 2003.

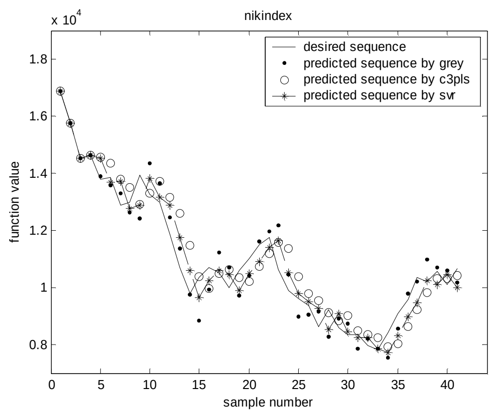

Figure 9. Forecastsof monthly Japan Nikkei Index for 41 months from Aug. 2000 to Dec. 2003.

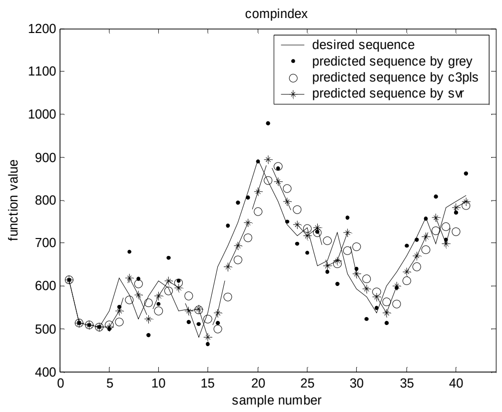

Figure 10. Forecasts of monthly Korea Comp. Index for 41 months from Aug. 2000 to Dec. 2003.

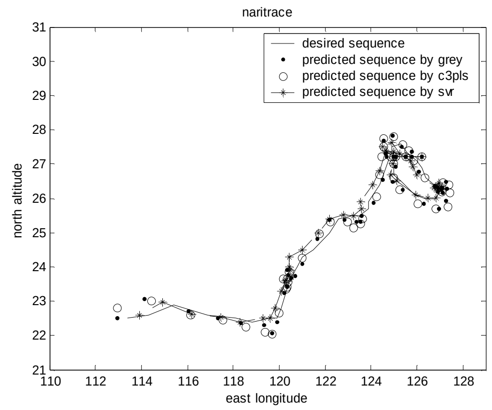

Figure 11. Forecasts of Nari typhoon moving trace during 6-19, September 2001.

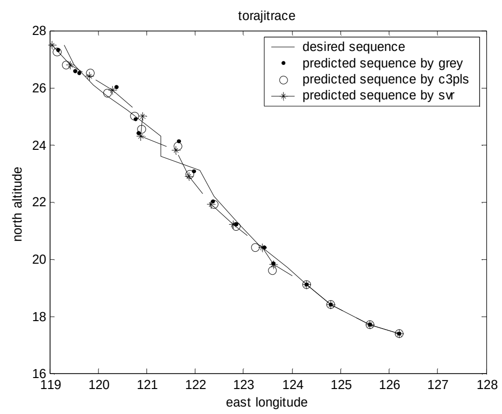

Figure 12. Forecasts of Toraji typhoon moving trace during 28-31, July 2001.